Spatial Information and Boolean Genetic Regulatory Networks.

Matthieu Manceny, Marc Aiguier, Pascale Le Gall, Joan Hérisson, Ivan Junier, François Képès

Spatial Information to Restrict the Dynamics of Genetic Regulatory Networks*

Matthieu Manceny[1,2], Marc Aiguier[1], Pascale Le Gall[1,2], Joan Hérisson[2], Ivan Junier[2], and François Képès[2]

[1] École Centrale Paris, MAS Laboratory, France
[2] Epigenomics Project, University of Évry, France

Abstract. In the course of understanding the functioning of cellular processes, modelling frameworks for biological networks are mandatory in order 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). Formal techniques, most of them based on model checking, 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. Recently, it has been observed in *in vivo* experiments and in genomic and transcriptomic studies, that spatial information are 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. Gene Regulatory Networks, Spatial Information, 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, 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.

* This work is performed within the European project GENNETEC (STREP 34952).

Recent experiments have shown that both in eukaryotes [1] 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 [3, 4]. These have revealed that the genes which are regulated by a given transcription factor and the gene which codes for the transcription factor tend to be located periodically along the DNA [3]. In this way, the genes can be easily co-localized in the three-dimensional space according to a solenoidal structure, even in the presence of several kinds of transcription factors [5]. 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 [6] a decade ago and formalised by Vilar Dheibler [7]. 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 proximity of reactive groups [5, 6, 8].

We propose, in this article, a simple scheme in order 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 [9]. The spatial information is modelled through the notion of privileged interaction which 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 a contradictory information, then the privileged interaction is prefered to the non privileged one.

The paper is structured as follows. Section 2 presents our model of GRN including privileged interactions. In section 3, we are interested in the Boolean dynamics of such GRN. The dynamics is governed by a set of so called 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 the new information on privileged interactions to reduce the number of dynamics to consider. A toy biological example is given in section 5, as some numerical results on artificial GRN. Finally, section 6 gives some concluding remarks.

2 GRN with Privileged Interactions (PGRN)

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 Boolean GRN, that is GRN where gene can only have two *levels of expression* (see section 3). An interaction $(a \to b)$ can

be either an activation or an inhibition, which will imply different behaviours considering the dynamics: 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.

Definition 1 (GRN with privileged interactions). *A* GRN with privileged interactions *(PGRN for short) is a labelled directed graph* $G = (V, E, S, P)$ *where* (V, E, S) *is an* interaction graph *i.e.:*

1. V *is a finite set whose elements are called* variables*,*
2. $E \subseteq V \times V$ *is the set of* interactions*,*
3. $S : E \to \{+, -\}$ *associates to each interaction its* sign *("+" for* activation *and "−" for* inhibition*),*

and $P \subseteq E$ *is the set of* privileged interactions*.*

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):

$$V^-(i) = \{j | j \in V, (j, i) \in E\} \qquad V^+(i) = \{j | j \in V, (i, j) \in E\}$$

Finally, We denote by $P(i)$ 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) *be a PGRN, and let* $i \in V$ *be a gene. We denote by* $A(i)$ *(resp.* $I(i)$*) the set of* activators *(resp.* inhibitors*) of* i*:*

$$A(i) = \{j | j \in V^-(i), S(j, i) = +\} \qquad I(i) = \{j | j \in V^-(i), S(j, i) = -\}$$

In the following, a PGRN 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 (Example of interaction graph). Let us exemplify definition 1 with the toy interaction graph (that is without any information on privileged interactions) from figure 1 where a gene i is inhibited by j_1 and j_2 and activated by k. Section 3 will present the dynamics of such a graph; the influence of privileged interactions among these three interactions is presented in section 4.

3 Boolean Dynamics of PGRN

3.1 Boolean Dynamics and Logical Parameters

In *Boolean dynamics*, genes can attain two levels, called *level of expression*: *effective* denoted by 1, or *ineffective* denoted by 0 (for example, genes can be described as expressed or not expressed at any time). The knowledge of the levels of expression of all the genes define a *Boolean dynamic state*.

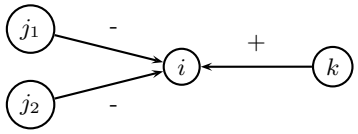

Fig. 1. Example of interaction graph

Definition 3 (Boolean dynamic states). *Let $G = (V, E, S, P)$ be a PGRN, and let $i \in V$ be a gene. We denote*[3] *by $\mathbb{X}(G)$ the set of Boolean dynamic states of G: $\mathbb{X}(G) = \{0,1\}^{|V|}$. For $x = (x_1, ..., x_{|V|}) \in \mathbb{X}(G)$, $x_i \in \{0,1\}$ is the* level of expression *of gene i in x.*

The *dynamics of a PGRN* consists in the evolution of each gene's level of expression 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 4 (Effective predecessors). *Let $G = (V, E, S, P)$ be a PGRN, and let $i \in V$ be a gene. Given a dynamic state $x \in \mathbb{X}(G)$, we denote by $A^*(i,x)$ (resp. $I^*(i,x)$) the set of* effective activators *(resp.* effective inhibitors*) of i; $w^*(i,x)$ denotes the set of* effective predecessors *of i:*

$$A^*(i,x) = \{j | j \in V^-(i), S(j,i) = +, x_j = 1\}$$
$$I^*(i,x) = \{j | j \in V^-(i), S(j,i) = -, x_j = 1\}$$
$$w^*(i,x) = A^*(i,x) \cup I^*(i,x)$$

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

Definition 5 (Logical parameters). *Let (V, E, S, P) be a PGRN. For each gene $i \in V$, we denote by $K_i : 2^{V^-(i)} \to \{0,1\}$ the set of* logical parameters *associated to i.*

Example 2 (Logical parameters). In Fig. 1, gene i has three predecessors. Thus, there is 8 logical parameters K_i to consider: $K_i(\emptyset)$, $K_i(\{j_1\})$, $K_i(\{j_2\})$, $K_i(\{k\})$, $K_i(\{j_1, j_2\})$, $K_i(\{j_1, k\})$, $K_i(\{j_2, k\})$ and $K_i(\{j_1, j_2, k\})$.

For example, the logical parameter $K_i(\{j_2, k\})$ represents i's next level of expression when the dynamic state is such that $x_{j_1} = 0$, $x_{j_2} = 1$ and $x_k = 1$.

Determining the dynamics of a PGRN consists in the attribution of values to the different logical parameters. The number of the possible attributions is huge: given a gene i, there are $2^{|V^-(i)|}$ logical parameters K_i, and each parameter can take 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 we have to consider $2^{2^3} = 256$ possibilities. Nevertheless, the structure of the interaction graph restricts the possible values of logical parameters, which is presented in the next section.

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

3.2 Valid Logical Parameters

Given an interaction graph, the attribution of values to logical parameters must respect some constraints, linked to the structure of the interaction graph and to the type of interaction. Logical parameters which respect the three 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 level of expression of i cannot decrease.

Constraint 1 (Definition). *Let (V, E, S, P) be a PGRN, and let i, j in V be two genes such that $j \in V^-(i)$. Then:*

$$S(j,i) = + \quad \Rightarrow \quad \forall \omega \subseteq V^-(i), K_i(\omega) \leq K_i(\omega \cup \{j\})$$

$$S(j,i) = - \quad \Rightarrow \quad \forall \omega \subseteq V^-(i), K_i(\omega) \geq K_i(\omega \cup \{j\})$$

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 level of expression of i. In other word, at least one of the previous inequality is strict.

Constraint 2 (Observation). *Let (V, E, S, P) be a PGRN, and let i, j in V be two genes such that $j \in V^-(i)$. Then:*

$$S(j,i) = + \quad \Rightarrow \quad \exists \omega \subseteq V^-(i), K_i(\omega) < K_i(\omega \cup \{j\})$$

$$S(j,i) = - \quad \Rightarrow \quad \exists \omega \subseteq V^-(i), K_i(\omega) > K_i(\omega \cup \{j\})$$

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

Constraint 3 (Maximum). *Let (V, E, S, P) be a PGRN, and let i in V be a gene. Then: $K_i(A(i)) = 1$, and $K_i(I(i)) = 0$.*

Example 3 (Valid parameters). Let us consider the interaction graph from Fig. 1. The Maximum constraint imposes that $K_i(\{k\}) = 1$ and $K_i(\{j_1, j_2\}) = 0$. Other relations between parameters are resumed in Fig. 3, 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.

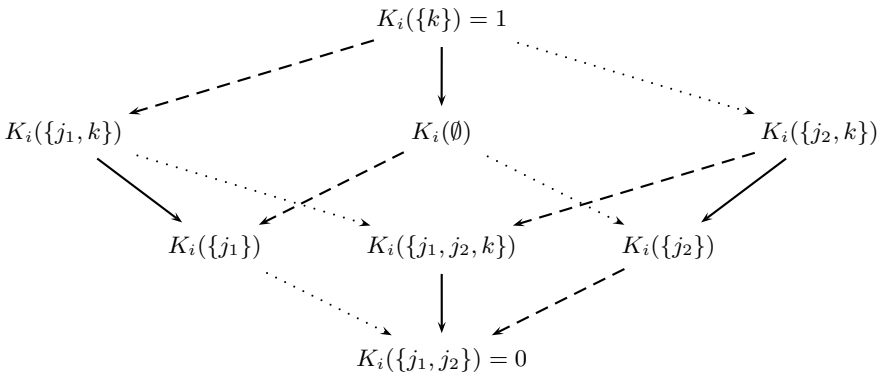

Fig. 2. Relation among logical parameters of the interaction graph from Fig. 1.

4 Toward a reduction of valid dynamics

4.1 Conflicts and Dilemma

Despite the above constraints, possible valid dynamics of PGRN still remains too numerous. These 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 6 (Conflicts and dilemma). *Let $G = (V, E, S, P)$ be an interaction graph, let $i \in V$ be a gene and let $x \in \mathbb{X}(G)$ be a dynamic state. x is a* situation of conflict *for gene i if, and only if, $A^*(i, x) \neq \emptyset$ and $I^*(i, x) \neq \emptyset$. x is a* situation of dilemma *for gene i if, and only if, ($A^*(i, x) \neq \emptyset$ 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 4 (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.

4.2 Constraints Based on Privileged Interactions

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.

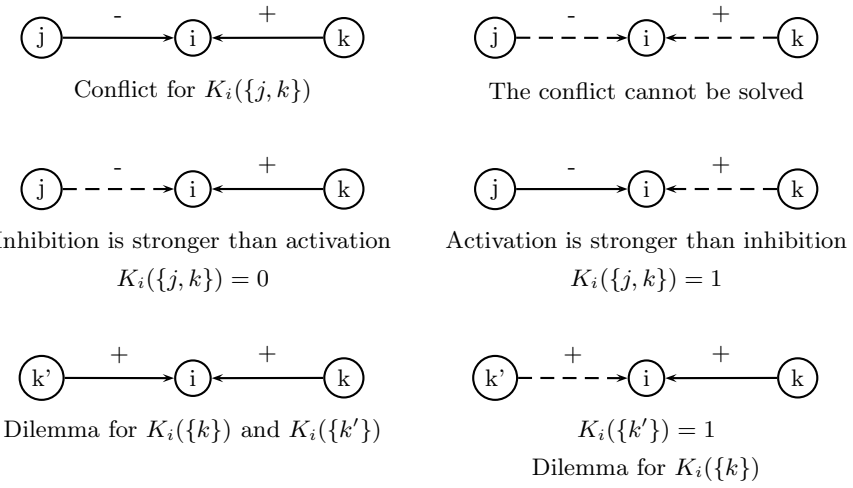

Fig. 3. Solving conflicts and dilemma with privileged interactions

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) are effective, and some privileged inhibitors (resp. activators) of the considered gene are effective, then the expression level is 0 (resp. 1).

Constraint 4 (Direct influence). *Let $G = (V, E, S, P)$ be a PGRN. Let $i \in V$ be a gene and $x \in \mathbb{X}(G)$ be a Boolean dynamic state. Then:*

$$A^*(i,x) \cap P(i) \neq \emptyset \text{ and } I^*(i,x) \cap P(i) = \emptyset \quad \Rightarrow \quad K_i(w^*(i,x)) = 1$$

$$I^*(i,x) \cap P(i) \neq \emptyset \text{ and } A^*(i,x) \cap P(i) = \emptyset \quad \Rightarrow \quad K_i(w^*(i,x)) = 0$$

The second constraint, called *Relative influence*, states that levels of expression 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 5 (Relative influence). *Let (V, E, S, P) be a PGRN. 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 $j \notin P(i)$. Then: $K_i(\omega \cup \{j\}) = K_i(\omega)$.*

Example 5 (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.

4.3 Unique Dynamics

We present here conditions to obtain, given a PGRN, a unique set of parameters leading to a unique dynamics. Obviously, if some genes have no predecessor, we cannot determine their levels of expression, which in fact does 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). *The conflict situations of a PGRN (V, E, S, P) can be solved iff for all $i \in V$, $P(i) = A(i)$ or $P(i) = I(i)$*

Proof. Sufficient. Let x be a situation of conflict for gene i, that is $A^*(i,x) \neq \emptyset$ 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. Dilemma 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). *The dilemma situations of PGRN* (V, E, S, P) *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* $|I(i)| = 1$*).*

Proof. 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.

These two theorems lead to the following necessary and sufficient condition to have *no conflict nor dilemma*.

Theorem 3 (No conflict nor dilemma). *Conflict and dilemma situations of a PGRN* (V, E, S, P) *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. The theorem is a direct consequence of theorems 1 and 2.

5 Some Results

5.1 A Toy Biological Example

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 [10] but a simplified genetic regulatory network has been proposed by Guespin and Kaufman [11], see Fig.4.

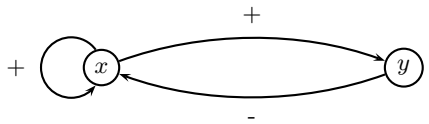

Fig. 4. Interaction graph for the mucus production system in *P. aeruginosa*

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 [11]. The models compatible with this hypothesis are constructed in [12].

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 $K_x(\{x,y\}) = 1$.

5.2 Artificial PGRN

In order to estimate the reduction in number of models allows by the use of privileged interactions, we have generated random PGRN. These artificial PGRN are composed of 10, 25, 50 and 100 genes, with different ratios of privileged interactions: no privileged interactions, one out of ten, one out of five, one out of two and when all interactions are privileged ones. We then compute the number of total dynamics (that is without any constraints) and this number when all the constraints are applied. The results given in Tab. 1 are a mean over 100 tests, the first three tables correspond to situation where each gene has exactly two, three or four predecessors, and the last table to a situation where each gene has a random number of predecessors between one and four.

Obviously, the number of dynamics we have to deal with is huge (at least 10^{12}, see row "*Total*"), and this number is squared when the number of genes doubles, or when the number of predecessors is increased by one. The non privileged constraints allow us to reduce significantly this number (row "0"). With the privileged constraints, the best results are obtained when half of interactions are privileged ones (row "1/2"), but the improvement is clearly observed even with small information, e.g. when only one interaction out of ten is privileged (row "1/10"). In the fourth table, the number of dynamics is divided by 10 for a ten genes network, by 10^5 for 25 genes, and by 10^{18} for 100 genes.

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

Privileged	*Genes*			
	10	25	50	100
0	1024	3.10^7	1.10^{15}	1.10^{30}
1/10	408	2.10^7	7.10^{12}	6.10^{25}
1/5	174	2.10^5	5.10^{11}	3.10^{21}
1/2	22	1171	1.10^6	9.10^{11}
1	17	1493	1.10^6	6.10^{12}
Total	10^{12}	10^{30}	10^{60}	2.10^{120}

Each gene has exactly 2 predecessors

Privileged	*Genes*			
	10	25	50	100
0	3.10^9	7.10^{23}	5.10^{47}	2.10^{95}
1/10	4.10^8	6.10^{20}	1.10^{41}	2.10^{83}
1/5	2.10^7	2.10^{18}	4.10^{35}	6.10^{67}
1/2	4.10^4	1.10^{11}	2.10^{20}	2.10^{38}
1	3.10^5	1.10^{13}	6.10^{26}	6.10^{47}
Total	1.10^{24}	1.10^{60}	2.10^{120}	6.10^{240}

Each gene has exactly 3 predecessors

Privileged	*Genes*			
	10	25	50	100
0	3.10^{20}	2.10^{51}	7.10^{102}	–
1/10	6.10^{18}	2.10^{46}	7.10^{88}	–
1/5	3.10^{16}	9.10^{41}	1.10^{75}	–
1/2	2.10^9	2.10^{21}	3.10^{38}	–
1	1.10^{14}	6.10^{33}	4.10^{61}	–
Total	1.10^{48}	2.10^{120}	6.10^{240}	4.10^{481}

Each gene has exactly 4 predecessors

Privileged	*Genes*			
	10	25	50	100
0	2.10^{12}	2.10^{29}	1.10^{54}	2.10^{101}
1/10	7.10^{11}	5.10^{24}	3.10^{41}	6.10^{83}
1/5	3.10^8	2.10^{21}	1.10^{38}	8.10^{63}
1/2	2.10^4	1.10^{10}	1.10^{18}	7.10^{36}
1	1.10^7	3.10^{13}	1.10^{25}	1.10^{48}
Total	1.10^{33}	1.10^{73}	1.10^{140}	1.10^{265}

Each gene has between 1 and 4 predecessors

Table 1. Number of Dynamics for Artificial PGRN

is captured as privileged interactions, which are a subset of classic interactions. With this notion, we have been able to determine conditions on interactions which lead to a reduction of the number of GRN Boolean dynamics to consider. The different tests we have made on artificial GRN show that even if the number of models to verify still remains huge with spatial information, the reduction is important (the number of valid dynamics is divided by 10^5 for a GRN of 25 genes). We are now interested in validation of this work with real GRN. But, although spatial information seams to be central in order to apprehend the complexity of biological networks, experimental data are rare, and mainly concern large GRN, which are for the moment hardly attainable with this 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.

An extension of this work we are particularly interested in deals with multivalued dynamics. In such framework, levels of expression of genes are not Boolean, but can take a finite number of values. To each level of expression is associated the capacity of a gene to influence a subset of its successors. When a gene i acts on j and k for example, it may be known that the level of expression of i mandatory for an action on j to be effective is higher than the level necessary for the action of i on k. To each interaction is associated a *threshold* the level of expression of the source gene must exceed in order to the interaction to become effective. Thus, given an interaction graph, the number of dynamics to consider

is even higher than in Boolean dynamics, because we have to consider all the different values for thresholds parameters (in fact, Boolean dynamics may be viewed as multivalued dynamics where all thresholds are equal to 1).

In such a context, the spatial information we considered is composed of two aspects. The first one is identical to privileged interaction in Boolean dynamics. The second one is the notion of *cluster* which 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). As privileged interactions lead to constraints on logical parameters, clusters will impose constraints on threshold parameters.

References

1. Jackson, D.A., Hassan, A.B., Errington, R.J., Cook, P.R.: Visualization of focal sites of transcription whitin human nuclei. J. Cell Biol. **164** (2004) 515–526
2. Cabrera, J.E., Jin, D.J.: The distribution of rna polymerase in escherichia coli is dynamic and sensitive to environmental cues. Mol Microbiol **50**(5) (2003) 1493–1505
3. Képès, F.: Periodic transcriptional organization of the e.coli genome. J Mol Biol **340**(5) (2004) 957–964
4. Carpentier, A.S., Torresani, B., Grossmann, A., Henaut, A.: Decoding the nucleoid organisation of bacillus subtilis and escherichia coli through gene expression data. BMC Genomics **6**(1) (2005) 84
5. Képès, F., Vaillant, C.: Transcription-based solenoidal model of chromosomes. Complexus **1**(4) (2003) 171–180
6. Muller-Hill, B.: The function of auxiliary operators. Mol Microbiol **29**(1) (1998 Jul) 13–18
7. Vilar, J.M.G., Leibler, S.: Dna looping and physical constraints on transcription regulation. J Mol Biol **331**(5) (2003) 981–989
8. Sexton, T., Schober, H., Fraser, P., Gasser, S.: Gene regulation through nuclear organization. Nat Struct Mol Biol **14**(11) (2007 Nov 5) 1049–1055
9. Thomas, R.: Logical analysis of systems comprising feedback loops. J. Theor. Biol. **73**(4) (1978) 631–56
10. Govan, J., Deretic, V.: Microbial pathogenesis in cystic fibrosis: mucoid pseudomonas aeruginosa and burkholderia cepacia. Microbiol rev. **60**(3) (1996) 539–74
11. Guespin-Michel, J., Kaufman, M.: Positive feedback circuits and adaptive regulations in bacteria. Acta Biotheor. **49**(4) (2001) 207–218
12. Bernot, G., Comet, J.P., Richard, A., Guespin, J.: Application of formal methods to biological regulatory networks: Extending Thomas' asynchronous logical approach with temporal logic. Journal of Theoretical Biology **229**(3) (2004) 339–347