Improving Snoussi constraints in the Thomas framework for Gene Networks

Guillaume GRATALOUP[1], Jean-Paul COMET[1] and Gilles BERNOT[1]

Laboratoire I3S, 2000 route des lucioles, 06410, Sophia-Antipolis, France

Corresponding author: jean-paul.comet@univ-cotedazur.fr

Keywords Gene regulatory networks, R. Thomas' discrete modeling framework, Snoussi's constraints

Abstract *The modeling of gene networks plays a crucial role for the comprehension and control of gene regulatory networks, and the extraordinarily wide range of its applications reinforces the craze for systems biology. This is a transdisciplinary field where the cross-fertilization of disciplines aims at providing tools for helping the modeling activity. Whatever the modeling framework, the bottleneck of the modeling process remains the identification of parameters. Even in discrete abstract modeling frameworks such as the R. Thomas one, the combinatorics of parameter settings make unrealistic the brute force approach consisting in enumeration of all parameterizations. Snoussi introduced in the late 80', a constraint allowing the discrete model to be consistent with a continuous one. In this article, we show that this constraint is not sufficient and propose an extension to the Snoussi constraints.*

Introduction

The study of biological systems aims to understand complex biological interaction mechanisms involved in a wide range of functions, from cell division to circadian rhythms. The modeling of these systems relies heavily on graphical representations that aim to provide a global, static view of the various elements involved in the interactions. From such a representation, various modeling techniques allow the computer to simulate molecular mechanisms at different scales, providing an overview of a variety of plausible behaviors of the biological processes under study.

One of the classes of biological systems, the Genetic Regulatory Networks (GRNs), aims at capturing the temporal succession of system regulations and is a central aspect of systems biology [1,2]. To represent gene regulations, modelers classically distinguish two types of interactions: activations (when a source gene is on, it tends to activate the target) and inhibitions (when a source gene is on, it tends to inhibit the target). The combinations of such basic blocks (activations and inhibitions) are sufficient to study a wide range of GRNs, and their computer simulation allows the modeler to provide predictions of various non-obvious biological behaviors.

For simulating these GRNs, one needs a so-called "modeling framework" which explains how the different elements of the system evolve according to the regulations they are subject to. Multiple mathematical frameworks exist [3,4] and each of them highlights a particular aspect of GRNs: discrete models highlight the qualitative nature of the regulations [5,6], stochastic models emphasize the non-determinism [7], differential equations often aim to base the simulations on the concentration of chemical species [8], and hybrid models try to mix some qualitative aspects with continuous ones [9,10]. While each of them has its own benefits and limitations, they share a common concern: the dynamic of the model depends on a parameter setting which is mandatory for animating the model. And unfortunately, whatever the modeling framework, the identification of parameters con-

stitutes the main bottleneck when modeling complex biological systems.
In this article, we focus on a hybrid modeling framework, the one introduced by H. Snoussi [9] combining R. Thomas' discrete modeling of GRN [11] with continuous differential equations. To make a long story short, the phase space is sliced into several parts (each corresponding to a region where all genes are constantly regulated) and inside each part, a differential equation system describes the evolution. This modeling framework is named *piecewise linear differential equations*. The strength of this modeling framework is to emphasize the qualitative nature of the regulations (the different parts of the phase space) while preserving a precise notion of time, which is often easily measurable. This cohabitation of discrete aspects with time allows to envision computer-aided techniques for helping biologists reason on these systems as they do in the classical completely discrete R. Thomas modeling framework, but with time information. Unfortunately, this is possible at the cost of introducing new continuous parameters which are to be determined.
The entanglement of discrete and continuous parameters of this model has been explored in the past [9] and led to the expression of constraints on discrete parameters if one wants to make coherent the discrete model and the piecewise differential equation one. Snoussi demonstrated that the more numerous the predecessors of a gene promoting the expression of the target gene, the greater the value of the parameter associated with the gene. This article introduces finer constraints that make it possible to better prune the combinatorics of the parameter settings to be considered.

The methodology section begins by introducing the modeling framework, starting with a discussion on R. Thomas' discretization of Gene Regulatory Networks (GRNs) followed by an exploration of the piecewise linear differential equations approach. Then, in the results section, we examine the links between the parameters in the discrete framework and those in the corresponding differential equations, and we give a theorem showing the constraints that the discrete parameters must satisfy in order to be compatible with a differential model. Using the above theorem in various scenarios, we demonstrate in the discussion section the usefulness of these novel constraints, which are particularly beneficial when dealing with scenarios where the number of predecessors exceeds two. In particular, in the context of an abstract cell cycle model, these constraints effectively divide the number of parameterisations to be considered by a factor of between 8 and 9. Finally, concluding remarks are given in the last section.

Methods

The R. Thomas modeling framework [11,5] used in this article places the modeling activity at a very high level of abstraction. For example, to represent that the product of gene X activates (*resp.* inhibits), directly or indirectly, the production of product of gene Y, one represents this interaction by two nodes X and Y (for abstracting both genes and their products) and a directed edge from X to Y labeled with "$+$" for activation (*resp.* "$-$" for inhibition).
Because our aim is to demonstrate that there exist some *new* constraints on discrete parameters of a R. Thomas model for making it coherent with the underlying piecewise linear differential equation system, we first introduce the R. Thomas modeling framework and, afterwards, the hybrid differential system.

R. Thomas' Discrete modeling framework

The intuition underlying R. Thomas' discrete modeling framework is that when one gene acts on another, the curve of the target gene's product concentration is often a sigmoid function of the source gene's product concentration. For instance, let us consider a GRN in which a gene X activates a gene Y and inhibits a gene Z (Fig.1-Left). The curve showing the production rate of the products of Y and Z versus the concentration of the product of X are both sigmoidal. The inflection points of these sigmoids can be used to place thresholds (Fig.1-Right). Below the threshold, the production rate of the target product is near zero, while above the threshold it is saturated. Thus, interaction can then be represented by a multi-valued switch-like system: the interaction is viewed as inactive or fully active depending on the position relative to the threshold. The intervals defined by the different thresholds are numbered and viewed as *discrete states* of the gene. For example, in Figure 1-Right, X can be in the discrete state 0 if under the first threshold, 1 if it is between the first threshold and the second one, and 2 if above the second threshold.

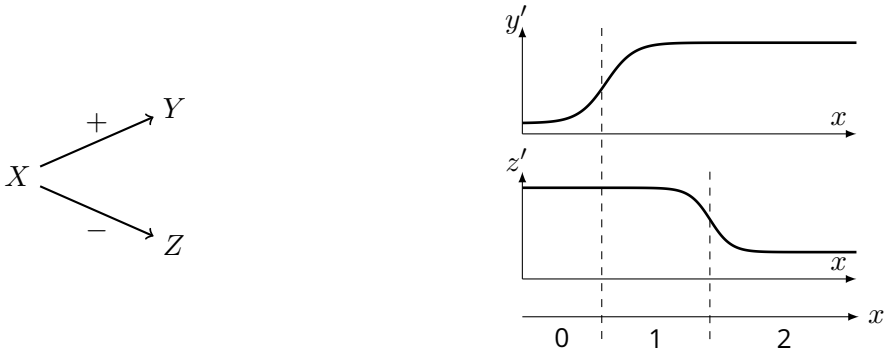
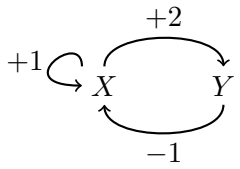


Fig. 1. (Left) Interaction graph: X activates Y and inhibits Z. (Right) Plot of the production rate of Y and Z as a function of the concentration of the product of X. The thresholds define the qualitative states of the system.

In addition, experimental observations can be used to deduce discrete parameters of the R. Thomas' model under construction. To be more precise, each parameter describes the discrete state that a gene of interest is attracted to, under known conditions, those conditions being the discrete state of the genes that are predecessors of the gene of interest. These parameters are denoted $K_{X,\omega}$ where X is the gene of interest and ω is the set of so-called *resources*, that is, either X's activators currently activating X or X's inhibitors not currently inhibiting X. For instance, the knowledge of the graph (Fig. 2a) determines the set of parameters to consider: there is a parameter for each gene and each possible set of resources for the considered gene. Fig. 2b lists 4 parameters for X (because X has 2 predecessors, leading to 4 combinations of resources) and 2 parameters for Y (because Y has a unique predecessor, leading to 2 combinations of resources). When a value is given to each of these parameters (right column of Fig. 2b), we can associate with each possible discrete state of the model (first and second columns of Fig. 2c), the set of resources (third and fourth columns of Fig. 2c) and the state toward which the system is attracted (fifth and sixth columns of Fig. 2c). From the table in Fig. 2c, it is possible to deduce the dynamics of the model. For example, when the system is in the discrete state $(1, 0)$, it is attracted toward $(2, 0)$ (Figure 2c). When the state, toward which the system is attracted, is equal to the current state, the state is *stable* (for example $(2, 1)$). Finally, when the Manhattan distance between the targeted state and the current one is strictly larger than 1 (see

state $(1,1)$, in which X is attracted upwards and Y is attracted downwards), there is two possible evolutions: either X increases or Y decreases.



(a)

K_X	=0
$K_{X,X}$	=2
$K_{X,Y}$	=1
$K_{X,XY}$	=2
K_Y	=0
$K_{Y,X}$	=1

(b)

Discrete state		Resources		Targeted state	
X	Y	X	Y	X	Y
0	0	Y	$\varnothing$	1	0
0	1	$\varnothing$	$\varnothing$	0	0
1	0	X, Y	$\varnothing$	2	0
1	1	X	$\varnothing$	2	0
2	0	X, Y	X	2	1
2	1	X	X	2	1

(c)

Fig. 2. (a) An interaction graph. (b) Possible set of of values for discrete parameters following the R. Thomas' framework. (c): Table describing the behavior of the model in each state

Piecewise linear differential equations

R. Thomas' framework is a beautiful framework for reasoning on succession of states, for example for confronting the built model with known temporal properties [12]. Nevertheless, it is not well suited when modeling a biological system in which time plays a crucial role, such as the cell cycle or the circadian clock [13,14] because it is not dedicated to manipulating the durations passed in each discrete state of a trajectory.

In such a case, it could be interesting to study the system in another framework where time is explicitly handled, such as piecewise linear differential equation systems. The parameters of such a system are difficult to measure in vivo, but Artificial Intelligence learning can help identify such parameters [15]. Nevertheless, there exists a strong connection between the piecewise linear differential equation framework and the discrete R. Thomas' parameters, and these relationships can be helpful for deriving constraints on discrete parameters.

Let us first present the considered differential framework. As for the discrete framework, thresholds define zones. In each of these zones, the evolution of the system is defined by a system of differential equations: The value of the derivative of the concentration of a gene product is the difference between a synthesis term and a degradation rate [9]. The synthesis term is the sum of the contribution to the synthesis rate of each of the predecessors that helps the rate grow (the resources). The equations are as follows: $\frac{d}{dt}x_i = F(x_1, ..., x_n) - \gamma_i \times x_i$ where t is the time, x_i the concentration of the product of gene i, γ_i the degradation rate of product of gene i and F the synthesis rate at the current state. Introducing

- k_i the basal synthesis rate (independent of the regulations) of x_i,
- $\omega_i(x_1, ..., x_n)$ the set of resources of i in the current state $(x_1, ..., x_n)$ where n is the total number of genes,
- k_{ji} the contribution of the regulation of j on the synthesis rate of i (only when j is an effective activator, or an ineffective inhibitor),

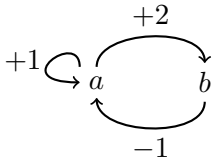
the synthesis rate can be detailed: $F(x_1, ..., x_n) = k_i + \sum_{j \in \omega_i(x_1,...,x_n)} k_{ji}$ is defined as the sum of individual contributions over the set of resources of i (the set of resources $\omega_i(x_1, ..., x_n)$ depends on the concentrations of each product species because it is necessary to know the position of these concentrations with respect to threshold in order to deduce if the regulation is effective). The analytic

solutions to these equations are: $x_i(t) = \frac{F(x_i)}{\gamma_i} - \left(\frac{F(x_i)}{\gamma_i} - x_i^0\right) e^{-\gamma_i \times t}$
with x_i^0 the initial value of x_i which can be identified using the initial conditions.

This expression of the concentration of a gene product over time allows us to link the discrete parameters of the associated R. Thomas' model and the parameters of the differential system: $\frac{F(x_i)}{\gamma_i}$ (which is equal to $\lim_{t\to\infty} x_i(t)$) is the concentration value towards which the system is attracted when i is under the control of the set of resources $\omega_i(x_1, ..., x_k)$. In other words, if we want to have a consistency between the discrete R. Thomas model and the differential one, the value of $\frac{F(x_i)}{\gamma_i}$ has to be in the interval denoted, in the discrete model, by the value of K_{x_i,ω_i}.

Results

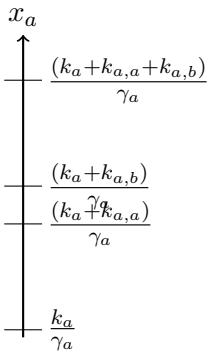
The previous remark implies that for every possible set of resources, a sum of k_{ij}, is linked to a discrete state of the discrete model. For example, in Figure 3, the consistency between the discrete R. Thomas' model and the piecewise differential system is insured only if $\frac{k_a}{\gamma_a}$ is in the interval given by the value of K_a, $\frac{k_a+k_{a,a}}{\gamma_a}$ is in the interval given by the value of $K_{a,a}$ and so on, following the table of Figure 3b.
As a result, the knowledge of differential equation parameters can be used to deduce, by ordering the different sums, the relative order of discrete parameters. Let us consider the example of Figure 3c where we suppose known the values of $\frac{k_a}{\gamma_a}$, $\frac{k_a+k_{a,a}}{\gamma_a}$ and $\frac{k_a+k_{a,b}}{\gamma_a}$. The value of $\frac{(k_a+k_{a,a}+k_{a,b})}{\gamma_a}$ can be deduced. Because $\frac{k_a}{\gamma_a}$ has to be in the interval defined by K_a, $\frac{k_a+k_{a,a}}{\gamma_a}$ (resp. $\frac{k_a+k_{a,b}}{\gamma_a}$, $\frac{k_a+k_{a,a}+k_{a,b}}{\gamma_a}$) has to be in the interval defined by $K_{a,a}$ (resp. $K_{a,b}$, $K_{a,ab}$), we deduce from the schema of Fig. 3c: $K_a \leq K_{a,a} \leq K_{a,b} \leq K_{a,ab}$. This puts the light on an intuitive first constraint: Whenever a set of resources ω_1 is included in another ω_2 then $K_{i,\omega_1} \leq K_{i,\omega_2}$. This constraint has already been described by H. Snoussi [9].



(a)

gene	ω_i	$\lim_{t\to\infty}(x_i(t))$	discrete state
a	$\{\varnothing\}$	$\frac{k_a}{\gamma_a}$	K_a
a	$\{a\}$	$\frac{k_a+k_{a,a}}{\gamma_a}$	$K_{a,a}$
a	$\{b\}$	$\frac{k_a+k_{a,b}}{\gamma_a}$	$K_{a,b}$
a	$\{a,b\}$	$\frac{k_a+k_{a,a}+k_{a,b}}{\gamma_a}$	$K_{a,ab}$
b	$\{\varnothing\}$	$\frac{k_b}{\gamma_a}$	K_b
b	$\{a\}$	$\frac{k_b+k_{b,a}}{\gamma_a}$	$K_{b,a}$

(b)



(c)

Fig. 3. (a) An interaction graph. (b) Relationships between the continuous and the discrete parameters. (c) Display of possible arbitrary values of continuous parameters associated with a.

However, this constraint can be completed by others deduced from the same remarks. Let us consider Figure 4a where the gene a has 3 predecessors b, c, d. Let us consider that the continuous parameters k_a, $k_{a,a}$, $k_{a,b}$ and $k_{a,d}$ have been chosen as in Figure 4b. From these values, one can deduce the sums of continuous parameters that correspond to discrete parameters, as in Figure 4b. If one considers that a can take $n+1$ discrete values and that the discrete parameters for a fulfill the

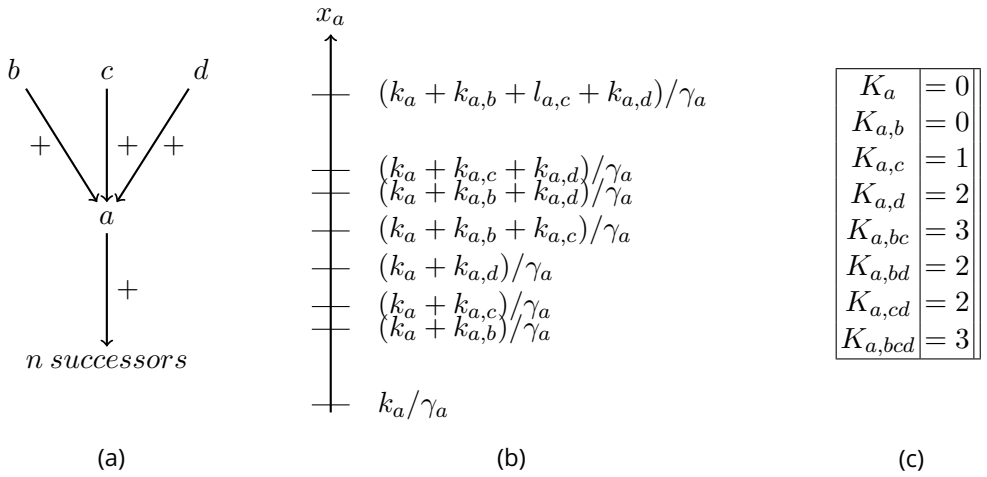


Fig. 4. (a): Interaction graph. (b): Display of possible arbitrary values of continuous parameters of a. (c): Possible set of values for discrete parameters of a respecting Snoussi's constraint.

previous constraints (also known as Snoussi constraints), one can choose the discrete values of the table in Figure 4c. However, $(k_{a,b} + k_{a,c} + k_a)/\gamma_a$ is less than $(k_{a,b} + k_{a,d} + k_a)/\gamma_a$, so it is impossible to place the first in an interval above the interval associated with the second. The values of the discrete parameters must therefore satisfy $K_{a,bc} \leq K_{a,bd}$ contrary to the choice made in the table of Figure 4c. The following theorem is the generalization of this observation:

Theorem 1. *If the discrete model comes from a piecewise linear differential equation model then for all genes i and for all subsets ω_1 and ω_2 of the predecessors of gene i, if there exists an injection ι from a partition P_1 of ω_1 to a partition P_2 of ω_2 such that for all ω element of P_1, one has either $\iota(\omega) = \omega$ or $K_{i,\omega} < K_{i,\iota(\omega)}$, then we have:* $K_{i,\omega_1} \leq K_{i,\omega_2}$

Proof of Theorem 1: Let us consider two subsets of predecessors (ω_1 and ω_2) of a gene i such that there exists an injection ι from a partition P_1 of ω_1 to a partition P_2 of ω_2 such that for all ω element of P_1, one has either $\iota(\omega) = \omega$ or $K_{i,\omega} < K_{i,\iota(\omega)}$. Two cases have to be considered:

1. If for all ω element of P_1, we have $\iota(\omega) = \omega$, we can deduce $\omega_1 \subset \omega_2$. Then,

$$\frac{k_i}{\gamma_i} + \sum_{j \in \omega_1} \frac{k_{i,j}}{\gamma_i} < \frac{k_i}{\gamma_i} + \sum_{j \in \omega_2 \cap \omega_1} \frac{k_{i,j}}{\gamma_i} + \sum_{j \in \omega_2 \setminus \omega_1} \frac{k_{i,j}}{\gamma_i}$$

   which can be expressed in terms of discrete parameters: $K_{i,\omega_1} \leq K_{i,\omega_2}$. This is the Snoussi condition.

2. Let us consider the case where there exists some $\omega \in P_1$ such that $\iota(\omega) \neq \omega$. In such a case, the partition P_1 is split into 2 sets: the set P_1^a which contains only ω such that $\iota(\omega) \neq \omega$ and the set P_1^b which contains only ω such that $\iota(\omega) = \omega$.
   Because for each ω element of P_1^a we have $K_{i,\omega} < K_{i,\iota(\omega)}$, the two focal values associated with the previous discrete parameters are in the same order:

$$\frac{k_i}{\gamma_i} + \sum_{j \in \omega} \frac{k_{i,j}}{\gamma_i} < \frac{k_i}{\gamma_i} + \sum_{j \in \iota(\omega)} \frac{k_{i,j}}{\gamma_i} \quad \Leftrightarrow \quad \sum_{j \in \omega} \frac{k_{i,j}}{\gamma_i} < \sum_{j \in \iota(\omega)} \frac{k_{i,j}}{\gamma_i}$$

Because for each ω' element of P_1^b we have $K_{i,\omega} = K_{i,\iota(\omega)}$, the two focal values associated with the previous discrete parameters are the same:

$$\frac{k_i}{\gamma_i} + \sum_{j \in \omega'} \frac{k_{i,j}}{\gamma_i} = \frac{k_i}{\gamma_i} + \sum_{j \in \iota(\omega')} \frac{k_{i,j}}{\gamma_i} \quad \Leftrightarrow \quad \sum_{j \in \omega'} \frac{k_{i,j}}{\gamma_i} = \sum_{j \in \iota(\omega')} \frac{k_{i,j}}{\gamma_i}$$

Summing the inequalities for all ω of $P_1^a \cup P_1^b$:

$$\sum_{j \in \bigcup_{\omega \in P_1^a} \omega} \frac{k_{i,j}}{\gamma_i} + \sum_{j \in \bigcup_{\omega' \in P_1^b} \omega'} \frac{k_{i,j}}{\gamma_i} \quad < \quad \sum_{j \in \bigcup_{\omega \in P_1^a} \iota(\omega)} \frac{k_{i,j}}{\gamma_i} + \sum_{j \in \bigcup_{\omega' \in P_1^b} \iota(\omega')} \frac{k_{i,j}}{\gamma_i} \quad \leq \quad \sum_{j \in \bigcup_{\omega' \in P_2} \omega'} \frac{k_{i,j}}{\gamma_i}$$

Because $\bigcup_{\omega \in P_1^a \cup P_1^b} \omega = \omega_1$ and $\bigcup_{\omega' \in P_2} \omega' = \omega_2$, one can deduce:

$$\frac{k_i}{\gamma_i} + \sum_{j \in \omega_1} \frac{k_{i,j}}{\gamma_i} \quad < \quad \frac{k_i}{\gamma_i} + \sum_{j \in \omega_2} \frac{k_{i,j}}{\gamma_i}$$

which can be expressed in terms of discrete parameters: $K_{i,\omega_1} \leq K_{i,\omega_2}$.

Because the property is satisfied in both cases, the theorem is proved. □

Discussion

Applying this theorem in practice reduces the number of possible settings to consider. To better understand to what extent, Table 1 shows the number of possible parameter settings for a single gene with different numbers of predecessors and different numbers of discrete levels.

Conditions		Total	Snoussi	Theorem 1	% of Snoussi
#Predecessors	#Levels				
2	2	16	6	6	100%
2	3	81	20	20	100%
2	4	256	50	50	100%
2	5	625	105	105	100%
2	6	1296	196	196	100%
3	2	256	20	20	100%
3	3	6561	168	150	89.3%
3	4	65536	887	707	79.7%
3	5	390625	3490	2518	72.1%
3	6	1679616	11196	7416	66.2%
4	2	65536	168	168	100%
4	3	43046721	7581	3863	51.0%

Tab. 1. Number of possible discrete parameter settings of a gene for given conditions and under different constraints. #Levels denotes the number of discrete states the gene can have, and #Predecessors is the number of its predecessors. The columns *Total*, *Snoussi*, and *Theorem 1* represent the total number of parameter settings without constraints, with Snoussi constraints, and with Theorem 1 constraints, respectively. Finally, the column *% of Snoussi* represents the proportion of parameter settings selected by the constraints of Theorem 1 out of all parameter settings satisfying the Snoussi constraints.

This table of the number of parameter settings that satisfy the conditions of the theorem leads to two observations:

1. When the considered target variable i is binary (2 levels), the newly introduced constraints are not able to eliminate more parameter settings. This is actually trivial, since for the theorem to

lead to a constraint complementary to Snoussi's, there must be an $\omega \in P_1$ such that $K_{i,\omega} < K_{i,\iota(\omega)}$. Since the variable is binary, we have $K_{i,\omega} = 0$ and $K_{i,\iota(\omega)} = 1$. Thus, K_{i,ω_2} is equal to 1, the conclusion is trivial whether K_{i,ω_1} is 0 or 1.

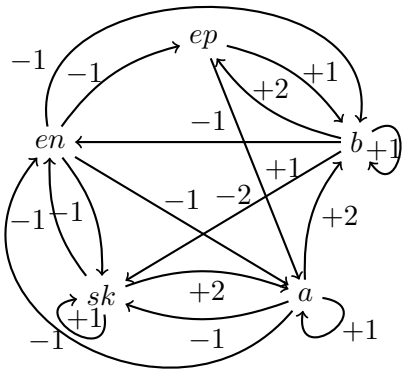
2. Again, when the number of predecessors is 2 (p_1 and p_2), the constraints expressed by the theorem don't seem to reduce the number of possible parameter settings. We can prove this remark:
   — The choice of $\omega_1 = \omega_2$ is not pertinent because the conclusion of the theorem provides the inequality $K_{i,\omega_1} \leq K_{i,\omega_2}$ which is trivial.
   — Let us consider $\omega_1 \neq \omega_2$.
      — If $\omega_1 = \{p_1\}$ and $\omega_2 = \{p_1, p_2\}$, we obtain either a triviality ($K_{i,p_1} < K_{i,p_1p_2}$ implies $K_{i,p_1} \leq K_{i,p_1p_2}$) or Snoussi's constraints ($K_{i,p_1} < K_{i,p_2}$ implies $K_{i,p_1} \leq K_{i,p_1p_2}$).
      — If $\omega_1 = \{p_1, p_2\}$ and $\omega_2 = \{p_1\}$, we obtain a triviality ($K_{i,p_1p_2} < K_{i,p_1}$ implies $K_{i,p_1p_2} \leq K_{i,p_1}$).
      — Other cases are handled by symmetry.

Thus, for the theorem to be more restrictive than the Snoussi constraints, it is necessary to have more than two predecessors and more than two levels.

Finally, we can observe that the greater the number of predecessors and the number of levels, the more restrictive the new constraints of the theorem. For example, if a gene is under the control of three regulators and has four qualitative levels, the new constraints of Theorem 1 eliminate almost half of the parameter settings that satisfy the Snoussi constraints.

More importantly, the gain is more significant when considering a global model rather than a single entity. To evaluate the gain, we consider an abstract model of the cell cycle, inspired by the work of Boyenval [16] who was interested in formalizing the notion of checkpoints. This model aims to highlight the progression through the cell cycle, which is driven by 2 types of genetic entities [17]: complexes of *Cyclins/Cyclin-dependent kinases (Cyc/Cdks)* and their inhibitors, known as *ennemies*. The 5 variables of the graph, (see Fig. 5a) represent these entities: sk is the abstraction of the two complexes CycE/Cdk2 and CycH/Cdk7, known as initiation kinases. a and b respectively represent CycA/Cdk1 and CycB/Cdk1, respectively. en is the abstraction of the main Cyc/Cdks enemies: the anaphase-promoting complex APC/Cdh1, cyclin-kinase inhibitors p21 and p27, and the Wee1 protein. The variable ep is the anaphase-promoting complex APC/Cdc20, which is a Cyc/Cdks enemy involved in mitosis exit and is called exit protein. Regulations between the variables are described in [13]. Note that the cooperation between several entities on their common targets presented in [16] has been introduced to reduce the possible set of parameter settings. Here, we consider the interactions without these cooperations.

As explained above, we can see in the table of Figure 5b that entities with a number of predecessors (or levels) less than or equal to 2 do not provide a constraint more specific than the Snoussi constraint. Nevertheless, even if the gain due to a particular entity is not miraculous, the global gain is important. This is due to the fact that the total number of parameter settings is obtained by multiplying the number of possible parameter settings for each variable. Taking the table of Figure 5b, we can see that the genes ep and en each have 2 levels, therefore Theorem 1 does not provide any supplementary constraints. sk, a and b each have 4 predecessors and 3 levels, therefore they re-



(a) Abstract interaction graph of the cell cycle. See text for the description of each entity.

entities	#predecessors	#levels	Snoussi	Theorem 1	#% of Snoussi
sk	4	3	7581	3863	51.0%
a	4	3	7581	3863	51.0%
b	4	3	7581	3863	51.0%
ep	2	2	6	6	100%
en	3	2	20	20	100%
total	/	60	$52.3 * 10^{12}$	$6.9 * 10^{12}$	13.2%

(b) For each entity of the interaction graph, columns *Snoussi* and *Theorem 1* gives the number of possible parameter settings satisfying either the Snoussi constraint or the constraints of Theorem 1. *% of Snoussi* is the same as in Table 1.

Fig. 5. (a) An influence graph of the cell cycle. (b) Enumeration of possible parameterizations.

duce the number of compatible parameter settings by a factor of around 2. As the total number of possible parameter settings is obtained by multiplying the number of possible parameter settings for each variable, the total number of parameter settings satisfying the constraints of theorem 1 is about 13,2% of the number of parameter settings satisfying the Snoussi constraints ($0.51^3 = 0.132$). In this cell cycle example, applying theorem 1 reduces the total number of parameter setting by a factor of 7.5.

Conclusion

This paper proves that Snoussi's conditions for a R. Thomas' model to be coherent with the underlying piecewise linear differential equation system, are not sufficient and proposes new constraints on discrete parameters. These constraints are necessary to ensure the consistency between both formalisms and appear when both the number of levels and the number of predecessors are greater than 2. Now that modeling support tools are available, the community is interested in models of reasonable size but is constrained by parameter identification. As we show on the cell cycle network, these new constraints can be very useful, as they can drastically reduce the number of parameter settings to consider.

If these constraints are unsatisfied by a possible discrete parameter setting of a discrete model, then, we can deduce that there does not exist any underlying piecewise differential equation systems consistent with the given parameter setting of the discrete model. This is particularly useful when the modeling process starts with the construction of a discrete model and then transforms the discrete model into a continuous one. In such a configuration, a rapid test for knowing whether the introduced constraints are satisfied or not can prune a lot of Snoussi-compliant parameter settings, which cannot be transformed into a continuous model.

The effects of imposing extended constraints on realistic biological models vary widely, and the integration of these constraints always improves our understanding of the biological systems under study, such as the cell cycle model with its intricate checkpoints [17] or abstract regulatory models of metabolism [16]. This can manifest itself in several key ways: *(i)* Enhanced Simplicity and Clarity: By simplifying the model and reducing the degrees of freedom, the interactions between system com-

ponents become more transparent and comprehensible. *(ii)* Sharper Component Roles: With fewer parameterizations, the roles of individual components are more sharply defined, aiding in deciphering their contributions to the system's behavior. *(iii)* Reduced Overfitting Risk: Imposing constraints mitigates the risk of overfitting parameters to training data, as they must adhere to the predefined constraints, enhancing the model's generalizability. *(iv)* Highlighting Fundamental Mechanisms: The imposed constraints guide the model towards fewer potential dynamics, thereby spotlighting underlying fundamental mechanisms governing the system. These factors collectively contribute to a more profound understanding of the biological system under investigation.

References

[1] Wolkenhauer O. Systems biology: the reincarnation of systems theory applied in biology? Brief Bioinform. 2001;2(3):258-70.

[2] Kitano H. Computational systems biology. Nature. 2002;420(6912):206-10.

[3] deJong H. Modeling and simulation of genetic regulatory systems: a literature review. J Comput Biol. 2002;9(1):67-103.

[4] Schlitt T, Brazma A. Current approaches to gene regulatory network modelling. BMC Bioinformatics. 2007 September;8(6):S9.

[5] Thomas R, d'Ari R. Biological Feedback. CRC Press; 1990.

[6] Glass L, Kauffman SA. The logical analysis of continuous, nonlinear biochemical control networks. JTB. 1973;39:103-29.

[7] Bortolussi L, Policriti A. Hybrid approximation of stochastic process algebras for systems biology. IFAC Proceedings Volumes. 2008;41(2):12599-606. 17th IFAC World Congress.

[8] Polynikis A, Hogan SJ, di Bernardo M. Comparing different ODE modelling approaches for gene regulatory networks. J Theor Biol. 2009;261(4):511-30.

[9] Snoussi EH. Qualitative dynamics of piecewise-linear differential equations: a discrete mapping approach. Dynamics and stability of Systems. 1989;4(3-4):565-83.

[10] Glass L, Edwards R. Hybrid models of genetic networks: Mathematical challenges and biological relevance. J Theor Biol. 2018;458:111-8.

[11] Thomas R. Boolean formalization of genetic control circuits. Journal of Theoretical Biology. 1973;42:563-85.

[12] Bernot G, Comet JP, Richard A, Guespin J. Application of Formal Methods to Biological Regulatory Networks: Extending Thomas' Asynchronous Logical Approach with Temporal Logic. Journal of Theoretical Biology. 2004;229(3):339-47.

[13] Behaegel J, Comet JP, Bernot G, Cornillon E, Delaunay F. A hybrid model of cell cycle in mammals. J Bioinformatics and Comput Biol. 2016;14(1):1640001 [17 pp.].

[14] Comet JP, Bernot G, Das A, Diener F, C M, Cessieux A. Simplified models for the mammalian circadian clock. In: Proc. of the Evry Spring school on Modelling complex biological systems in the context of genomics. EDP Sciences; 2012. p. 85-106.

[15] Sun H. Identifying and Analyzing Long-term Dynamical Behaviors of Gene Regulatory Networks with Hybrid Modeling [PhD thesis]. PhD, Ecole centrale de Nantes; 2023.

[16] Boyenval D. Modélisation formelle de comportements cycliques biologiques avec points de contrôle : la régulation du cycle cellulaire [PhD thesis]. Thèse de doctorat, Informatique, Université Côte d'Azur, France; 2022.

[17] Boyenval D, Bernot G, Collavizza H, Comet JP. What is a cell cycle checkpoint? The TotemBioNet answer. In: Proceedings of the 18th International Conference on Computational Methods in Systems Biology (CMSB). vol. 12314 of LNCS. online; 2020. p. 362-72.