Journal of Bioinformatics and Computational Biology

A HYBRID MODEL OF CELL CYCLE IN MAMMALS

Jonathan BEHAEGEL, Jean-Paul COMET, Gilles BERNOT, Emilien CORNILLON
Université Nice-Sophia Antipolis, I3S-UMR CNRS 7271, CS 40121
06903 Sophia Antipolis Cedex, France
behaegel@i3s.unice.fr, {comet,bernot}@unice.fr, ecornillon@i3s.unice.fr

Franck DELAUNAY
Université Nice Sophia Antipolis, CNRS UMR7277, INSERM U1091,
Institut de Biologie Valrose, 06108 Nice, France,
franck.delaunay@unice.fr

Received (Day Month Year)
Revised (Day Month Year)
Accepted (Day Month Year)

Time plays an essential role in many biological systems, especially in cell cycle. Many models of biological systems rely on differential equations, but parameter identification is an obstacle to use differential frameworks. In this paper, we present a new hybrid modeling framework that extends René Thomas' discrete modeling. The core idea is to associate with each qualitative state "celerities" allowing us to compute the time spent in each state. This hybrid framework is illustrated by building a 5-variable model of the mammalian cell cycle. Its parameters are determined by applying formal methods on the underlying discrete model and by constraining parameters using timing observations on the cell cycle. This first hybrid model presents the most important known behaviors of the cell cycle, including quiescent phase and endoreplication.

Keywords: Discrete Modeling; Hybrid Modeling; Mammalian cell cycle.

1. Introduction

Regulatory networks are models based on graphs which we use to obtain a simpler view of gene regulation. Gene regulation is then the process of turning genes on and off and is made possible by a network of interactions using regulatory proteins. Gene regulation guarantees that appropriate genes are expressed at proper times specially during early development where cells begin their specific functions. Gene regulation also helps an organism to respond to its environment.

The various existing modeling frameworks differ by the aspects they highlight. Stochastic models emphasize non-determinism, differential models represent a system with a lot of details (transcription, traduction, transports ...) and give precise trajectories in terms of concentrations, qualitative models focus on the major features that explain the observations (only main causalities are taken into account), and hybrid models link qualitative aspects with continuous variables such as time.

For gene networks, the qualitative framework due to René Thomas[20,21] has become a standard because it highlights the qualitative nature of gene regulations and powerful software platforms have been designed that help the biologists designing and analysing models. Unfortunately, the qualitative nature of discrete models leads to completely abstract time scale: When chronometrical time plays a crucial role in the biological system, R. Thomas' discrete method is not sufficient.

Whatever the modeling framework, the main difficulty relies on the identification of *good* parameter values. For differential, probabilistic, discrete or hybrid models, parameters pilot the dynamics, and finding accurate values remains a difficult step. For discrete modeling frameworks like the one of R. Thomas, *formal methods* assist the modeler to automatically select or constrain the parameters in order to get accurate dynamics[2,7]. When the models have to handle durations, they can be described in terms of differential systems or in terms of hybrid models. Whereas parameter identification in differential systems is a hard problem even when the nature of differential systems is well known, hybrid modeling frameworks can still be assisted by computer-aided methods to help the modeler to set up the parameters.

In the context of gene networks, several hybrid frameworks have been designed in order to take into account timing informations[4,5,6]. In this article we introduce a new hybrid modeling framework where the discrete kinetic parameters of Thomas' approach are extended into "celerities" (real number values), allowing us to deduce the time spent in each discrete state. To illustrate the ability of this hybrid modeling framework to represent timing information, we model the well studied *cell cycle*, that gives several daughter cells starting from a parent cell. In this paper, we focus on mitosis which produces two daughter cells.

The behavior of the cell cycle in mammals can be summed up as follows. Cell cycle displays four phases which are always linked in the same order: G1, S, G2 and M. Each of these phases is controled by a protein complex made of a Cyclin and a Cdk (Cyclin dependant kinase). For example, Cyclin A/Cdk2 governs the S phase of cell cycle. Checkpoints exist for G1/S and G2/M transitions. They allow the cell to control that DNA is not damaged before continuing cell cycle and the second checkpoint additionally controls that DNA is properly replicated before starting mitosis (M phase). At the end of mitosis, that is after cell division, progeny cell can stop cell cycle in order to remain in an idle state G0 called *quiescent phase*. This phase terminates when the cell starts a new cell cycle. In another particular case of the cell cycle, the endoreplication allows the cell to terminate earlier the classical behavior by stopping the cell cycle at the first step of mitosis (no cellular division): Cell starts mitosis (prophase step) but it is stopped just before cell division. The cell directly goes towards the middle or end of the G1 phase without leading to the development of two daughter cells. Therefore the parent cell grows and its genome doubles: This phenomenon leads to polyploidy.

In this article, we use biological knowledge on the cell cycle in order to constrain the parameters of our hybrid model. Because the hybrid model relies on a discrete

one, we determine parameters using both formal methods for discrete parameter identification and measurements of time spent in the different phases.

The paper is organized as follows. We first define the hybrid modeling framework based on René Thomas' discrete one (Section 2). Then Section 3 is devoted to the interaction graph of the cell cycle. Section 4 focuses on the identification of parameters of both discrete and hybrid models. Section 5 sketches the results obtained by simulations. Finally, we discuss the limits of our hybrid model of the cell cycle in Section 6.

2. Hybrid Modeling Framework

2.1. *R. Thomas' Discrete Framework*

In 1973, René Thomas designed a discrete framework well suited for modeling dynamics of gene networks[20,21]. Quantitative concentrations of gene products are abstracted into qualitative levels. This abstraction is an acceptable simplification because real concentrations are not accurately measurable *in vivo* and the thresholds between qualitative levels are chosen in a clever way.

The gene regulations are classically represented by an interaction graph where vertices abstract gene products and arrows the regulations between them. When a gene regulates several targets, there is no reason that the regulations take place exactly at the same concentration. In such a case, we must consider more than two (on/off) abstract levels to represent the set of targets regulated by the gene. These genes are said *multivalued*. The dynamics of a model is driven by kinetic parameters, that give the qualitative variations for each gene product. From these parameters, a state transition graph showing the behaviors of the system can be easily built.

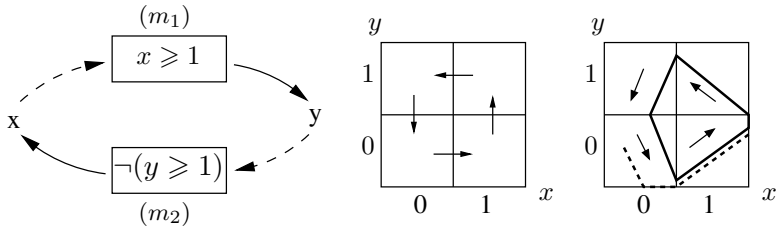


Fig. 1. *(Left)* Graphical representation of a gene regulatory graph with multiplexes. Dashed lines represent participation of variables in multiplexes (not present in Definition 1). *(Center)* State graph obtained with parameters $K_{x,\{\}} = 0$, $K_{x,m_2} = 1$, $K_{y,\{\}} = 0$, $K_{y,m_1} = 1$. *(Right)* Hybrid trajectories obtained by the following celerities: $C_{x,\{\},0} = -1.5$, $C_{x,\{\},1} = -4$, $C_{x,\{m_2\},0} = 1.5$, $C_{x,\{m_2\},1} = 4$, $C_{y,\{\},0} = -3.33$, $C_{y,\{\},1} = -3$, $C_{y,\{m_1\},0} = 2.4$ and $C_{y,\{m_1\},1} = 3$.

Here we complete the first formalization[2,3] of this approach using multiplexes, that specify, *via* a logical formula, the cooperation or concurrency between two or more regulators of the same target (see Fig. 1, left). Vertices are called *variables*

because these latter can gather genes which are co-expressed.

Definition 1. A *gene regulatory network with multiplexes* (GRN for short) is a tuple $N = (V, M, E, \mathcal{K})$ satisfying the following conditions:

- V and M are disjoint sets, whose elements are called *variables* and *multiplexes* respectively.
- $G = (V \cup M, E)$ is a labeled directed graph such that:
  (a) edges of E start from a multiplex and end to a variable.
  (b) every variable v of V is labeled by a positive integer b_v called the *bound* of v.
  (c) every multiplex m of M is labeled by a formula φ_m belonging to the language $\mathcal{L}$ inductively defined by:
    i. If v belongs to V and $s \in \mathbb{N}$, then $v \geqslant s$ is an atom of $\mathcal{L}$.
    ii. If φ and ψ belong to $\mathcal{L}$ then $\neg\varphi$, $(\varphi \wedge \psi)$ and $(\varphi \vee \psi)$ also belong to $\mathcal{L}$.
- $\mathcal{K} = \{K_{v,\omega}\}$ is a family of integers indexed by $v \in V$ and $\omega \subset N^-(v)$, where $N^-(v)$ is the set of predecessors of v in G (that is, the set of multiplexes m such that $m \to v$ is an edge of E). Each $K_{v,\omega}$ must satisfy $0 \leqslant K_{v,\omega} \leqslant b_v$.

Definition 2. (States, satisfaction relation and resources). Let N be a GRN and V be its set of variables. A *discrete state* of N is a function $\eta : V \to \mathbb{N}$ such that $\eta(v) \leqslant b_v$ for all $v \in V$. The satisfaction relation $\models_N$ between a state η of N and a formula φ of $\mathcal{L}$ is inductively defined by:

- If φ is an atom of the form $v \geqslant s$, then $\eta \models_N \varphi$ if and only if $\eta(v) \geqslant s$.
- If $\varphi \equiv \neg\psi$ then $\eta \models_N \varphi$ if and only if $\eta \not\models_N \psi$.
- If $\varphi \equiv \psi_1 \wedge \psi_2$ then $\eta \models_N \varphi$ if and only if $\eta \models_N \psi_1$ and $\eta \models_N \psi_2$; and we proceed similarly for the other connectives.

For $v \in V$, a multiplex $m \in N^-(v)$ is a *resource* of v at state η if $\eta \models_N \varphi_m$. The *set of resources* of v at state η is defined by $\rho(\eta, v) = \{m \in N^-(v) \mid \eta \models_N \varphi_m\}$.

According to Definition 2, a resource is a multiplex whose formula is satisfied at the current state. The parameter that gives the variation of variable v is the one associated with the set of resources of v at the current state. We call *focal point* the state whose coordinates are given by the parameters associated with each variable.

Definition 3. (State Graph). Let $N = (V, M, E, \mathcal{K})$ be a GRN. The *state graph* of N is the directed graph $\mathcal{S}$ defined as follows: The set of vertices is the set of states of N, and there exists an arc (called transition) $\eta \to \eta'$ if one of the following conditions is satisfied:

- for all variables $v \in V$ we have $\eta(v) = K_{v,\rho(\eta,v)}$, and then $\eta' = \eta$.
- there exists $v \in V$ such that $\eta(v) \neq K_{v,\rho(\eta,v)}$, and
  $\eta'(v) = \begin{cases} \eta(v) + 1 \text{ if } \eta(v) < K_{v,\rho(\eta,v)} \\ \eta(v) - 1 \text{ if } \eta(v) > K_{v,\rho(\eta,v)} \end{cases}$ and for all variables $u \neq v, \eta'(u) = \eta(u)$.

The first item of Definition 3 expresses a stable state: The focal point coincides with the current state. The second item expresses when it is possible to go from a state to another one: If the value of the variable v is lower (resp. greater) than the parameter value associated with v, then this variable can increase (resp. decrease) by one unit. Note that at each step, only *one* variable can evolve. The construction of the state graph is illustrated on the toy example in Fig. 1 (center).

The combinatorics of acceptable parameter values is exponential (because the number of parameters associated with a variable depends on the number of possible resources). In order to decrease this combinatorics, we admit the Snoussi's condition[18]: $\forall v \in V, \forall \omega, \omega' \subset N^-(v),$ if $\omega' \subseteq \omega,$ then $K_{v,\omega'} \leqslant K_{v,\omega}$.

2.2. *Hybrid modeling based on Thomas' modeling*

Definition 4 introduces the *hybrid gene regulatory networks*, a formalism using Thomas' discrete states, and adding continuous variables allowing the handling of time. The *celerities* can be viewed as abstract speeds.

Definition 4. A hybrid gene regulatory network (HGRN for short) is a tuple $N = (V, M, E, \mathcal{C})$ where V, M, E satisfy the two first items of Definition 1 and where $\mathcal{C} = \{C_{v,\omega,n}\}$ is a family of real numbers indexed by (v, ω, n) triplets where $v \in V$, $\omega \subset N^-(v)$ and n is an integer such as $0 \leqslant n \leqslant b_v$.

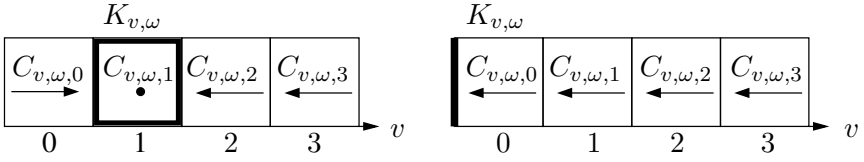


Fig. 2. Relationship between Thomas' parameters and celerities. Here, each discrete state is supposed to have the same set of resources ω for a given variable v. (Left) If $K_{v,\omega} = 1$ then below this discrete level, celerities are positive, above it, they are negative (see arrows). In the state where $v = 1$, the celerity is null. (Right) $K_{v,\omega}$ is supposed to be equal to 0 and the variable v is attracted towards 0: The celerity in the discrete state where $v = 0$, is negative or null.

There is a strong connexion between the discrete Thomas formalism and our hybrid formalism: The sign of the celerities can be deduced from the discrete parameters. The discrete parameters describe the focal point towards which the system is attracted, and, because the celerities express the directions of trajectories, their signs are constrained by the focal point. In Fig. 2, we represent a set of states that differ only by the value of a variable v and we assume that they share the same set of resources ω for v. Let $n = 0 \cdots 3$ be the states of v. When $K_{v,\omega}$ differs from 0 or b_v, three cases have to be considered (Fig. 2-left): If n is below $K_{v,\omega}$, we have $C_{v,\omega,n} > 0$, if n is above $K_{v,\omega}$, we have $C_{v,\omega,n} < 0$ and if n is equal to $K_{v,\omega}$, we have $C_{v,\omega,n} = 0$ because v has already reached the value towards which the variable

is attracted. If the parameter $K_{v,\omega}$ equals 0 or b_v (Fig. 2-right), two choices can be made: One can consider as previously that, when the value of v is equal to the parameter, the variable no longer evolves ($C_{v,\omega,0} = 0$ or $C_{v,\omega,b_v} = 0$), or one can consider that the variable is attracted towards the external boundary and in this case, for all n, the signs of celerities $C_{v,\omega,n}$ are constant.

Definition 5. Let us consider a HGRN $N = (V, M, E, \mathcal{C})$. A hybrid state of N is a couple $h = (\eta, \pi)$ where :

- η is a function of V in $\mathbb{N}$ such as $\forall v \in V$, $0 \leqslant \eta(v) \leqslant b_v$; η is called the discrete state of h,
- π is a function of V in the real interval [0, 1]. π is called the fractional part of h.

Whereas η represents the qualitative part of the current state, the fractional part reveals the exact position $(\pi(v))_{v\in V}$ inside the discrete state. For each variable $v \in V$, the set of hybrid coordinates of v $\{(\eta(v), \pi(v)) \mid \pi(v) \in [0, 1], \eta(v) = n\}$ can be interpreted as the real interval $[n, n + 1]$.

Definition 6. Let us consider a HGRN $N = (V, M, E, \mathcal{C})$ and a hybrid state $h = (\eta, \pi)$. The *touch delay* of v noted $\delta_h(v)$ is the time allowing v to reach the border of the current discrete state. For each $v \in V$, δ_h is the function of V in $\mathbb{R}^+$ defined by:

- if $C_{v,\rho(\eta,v),\eta(v)} = 0$, then $\delta_h(v) = +\infty$,
- if $C_{v,\rho(\eta,v),\eta(v)} > 0$, then $\delta_h(v) = \frac{1-\pi(v)}{C_{v,\rho(\eta,v),\eta(v)}}$,
- if $C_{v,\rho(\eta,v),\eta(v)} < 0$, then $\delta_h(v) = \frac{\pi(v)}{\left|C_{v,\rho(\eta,v),\eta(v)}\right|}$.

Definition 6 gives the times necessary for the variables to reach a border of the current discrete state. When experimental data allow us to derive the time spent in a state, the touch delay leads to constraints on celerities.

We do not formalize here the whole dynamics associated with such a hybrid model because the behavior is intuitive: Starting from a hybrid state $h = (\eta, \pi)$, the celerities define a linear evolution inside the discrete state until touching a border. At the border, several situations can appear. Let us consider that the trajectory reaches a border separating two discrete states that differ by the value of variable v.

(1) If the second discrete state can accept trajectories from the border (celerities of v in both neighbor states have the same sign), then the trajectory goes into the second discrete state and can be extended using celerities of the second one.
(2) If the second discrete state prevents trajectories to enter (celerities of v in both neighbor states have opposite sign), then the trajectory cannot enter the second discrete state and then slides along the border according to celerities of the other variables.

If the reached border is an external one ($\eta(v) = 0$ and $C_{v,\rho(\eta,v),0} < 0$ or symmetrically $v = b_v$ and $C_{v,\rho(\eta,v),b_v} > 0$), the trajectory cannot cross the border and it slides as in the previous case.

Fig. 1 (right) is a representation of the hybrid trajectories of the toy example. Directional vectors represent the relative celerities in each discrete state.

3. Molecular aspects of the Cell cycle

3.1. *Sketch of the cell cycle functioning*

Cell proliferation realized by the cell cycle is essential for individual survival. It ensures tissues renewal or cell growth. Mitosis is a biological phenomenon giving rise to two daughter cells starting from a unique cell. The offspring have the same characteristics than parent cell. The cell cycle is classically divided into 4 phases called respectively G1, S, G2 and M, which come one after another, leading to cell division, see Fig. 3 (left). Between G1 and S and between G2 and M, there exist two checkpoints: If conditions are not satisfied, the cell cycle stops and there is no cell division.

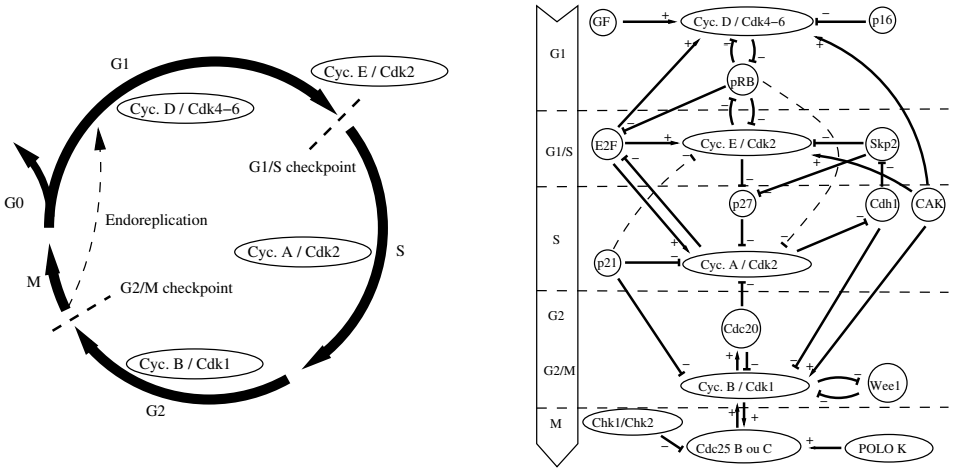


Fig. 3. *(Left)* Cell cycle phases. *(Right)* Molecular aspects of cell cycle.

From a molecular point of view, the cell cycle is mainly controled by complexes made of a cyclin and a kinase of the family Cdk[17], see Fig. 3 (right):

(1) The cell cycle starts when entering into G1 phase depending on growth factors present in the cell. During this phase, the cell prepares replication and increases its size. The growth factors stimulate the Cyclin D/Cdk4-6 complex, ensuring the activation of Cyclin E/Cdk2. This last complex allows the cell cycle to pass G1/S checkpoint[12] (verification of the non-damage of DNA).
(2) During S phase, Cyclin A/Cdk2 is indirectly stimulated by Cyclin E/Cdk2. Then the cell continues to grow and its centrosome is replicated.

(3) During G2 phase, cell prepares its division and achieves replication: The Cyclin A/Cdk2 activates Cyclin B/Cdk1 so that it passes through the G2/M checkpoint (verifications of non-damage of DNA and ending of replication).
(4) The M phase represents the steps of mitosis to make two daughter cells. In particular, a surveillance mechanism exists which recognizes the kinetochores attached on mitotic spindles that activates APC protein: The latter activates proteins triggering the cleavage of sister chromatids (metaphase/anaphase transition).

All these complexes are synthesized and degraded successively during a cell cycle, but several complexes can be expressed simultaneously. These complexes are also stimulated by kinases (Cdc25 or CAK complex) and they are inhibited by CKI (Cdk Inhibitors like p21 or p27) or by kinases like Wee1. Even if the molecular interaction map of the mammalian cell cycle is well known and rather large[14], here we focus on essential components that regulate cell cycle.

In addition to the classical cell cycle, one observes a quiescent phase called G0 that is the state when cell neither divides nor prepares division. Cells can enter G0 at the end of cell division, and then they do not proliferate. They can remain in the quiescent phase from a few hours to several years[19]. Then cells can re-enter into a new cell cycle starting in G1 phase.

Endoreplication is also linked to cell cycle[10]: Cells can start mitosis without finishing it. These cells duplicate their DNA (a cell with more than 2 copies of its chromosomes is said *polyploid*) and their cytosols grow. In mammals, megakaryocytes making platelets in blood have 128 copies of their chromosomes.

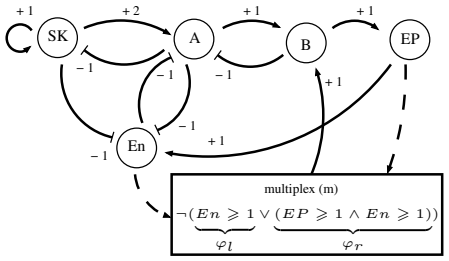
3.2. *A simplified 5-variable Gene Network*

Our interaction graph modeling molecular aspects of cell cycle is inspired by a John Tyson's model[23], which has been designed to represent mammalian or yeast cell cycle. We add supplementary biological phenomena in order to get an interaction graph specific for mammals, in which both checkpoints and surveillance mechanism of mitosis are taken into account. In Fig. 4, the simple multiplexes (with a unique input and a unique output) are replaced by arrows labeled by a sign and a threshold. For example:

- The arrow from SK to A is labeled by $+2$. It codes for the multiplex with formula $(SK \geqslant 2)$ pointing towards A,
- The arrow from SK to En is labeled by -1. It codes for the multiplex with formula $\neg(SK \geqslant 1)$ pointing towards En.

Our model differs from Tyson's model by the three following aspects:

- First, we include the CAK complex (Cyclin H/Cdk7) in SK: CAK is an activator of Cyclin E/Cdk2[15] whose activation is necessary to pass the G1/S checkpoint. Because CAK and Cyclin E/Cdk2 are both in SK, an autoregulation of SK is



Entity	Proteins and complexes represented in the variable
SK	Cyclin E/Cdk2, Cyclin H/Cdk7
A	Cyclin A/Cdk1
B	Cyclin B/Cdk1
En	APC^{G_1}, CKI (p21, p27), Wee1
EP	APC^{M}, Phosphatases

Fig. 4. A simplified 5-variable gene network of the mammalian cell cycle. APC^{M} is a complex involved in surveillance mechanism in order to enter into anaphase (third step of mitosis). The combined effects of the APC^{G_1}, p27 and Wee1 inactivate Cdk1, Cdk2, Cdk4 and Cdk6.

added. Moreover, because Cyclin E activates Cyclin A (involved in A), SK activates A. Knowing that CAK activates first Cyclin E/Cdk2 then Cyclin A/Cdk2, the level of the autoregulation of SK is lower than the level of the activation of SK on A. We chose accordingly the thresholds: The autoregulation of SK takes place at level 1 and the regulation of SK on A at level 2. Thus, SK is multivalued.

- Secondly, in the Tyson's model, Cyclin A and Cyclin B are both represented by a unique variable. We split these cyclin complexes because they act in different phases whose durations are different.
- The last modification is the addition of the multiplex m: It describes, *via* a logical formula, the conditions under which B is inhibited by EP and En. It is well known that these variables, when simultaneously present, inhibit B[22] (part φ_r of the formula). However, other biological knowledge points out that En inhibitors are sufficient[13] (part φ_l of the formula). The formula of m can then be simplified: An equivalent formula is $\neg(En \geqslant 1)$. Thus EP could be removed from the interaction graph. Nevertheless, we decide to keep EP because the APC^{M} protein acts during the surveillance mechanism of mitosis[9] which we want to represent.

4. Parameter Identification

4.1. *Identifying Discrete Parameters from traces*

SK is three-valued and the other variables are boolean. The number of states is $2^4 \times 3 = 48$. In the sequel, we note "`abcde`" the state where $SK = a$, $EP = b$, $A = c$, $B = d$ and $En = e$. For example, the state $(SK{=}0, EP{=}0, A{=}1, B{=}0, En{=}0)$ is denoted 00100. Thomas' dynamics is controled by 26 parameters (see Table 1).

We constrain the set of parameters using a "genetically modified" Hoare logic[1]. Classical Hoare logic for imperative programs has been introduced to prove correctness of programs. Hoare introduced the notation $\{P\}p\{Q\}$ to mean "If the assertion P (precondition) is satisfied before performing the program p and if the program terminates, then the assertion Q (postcondition) will be satisfied afterwards." This constitutes *de facto* a specification of the program under the form of a triple, called

Table 1. Constraints on discrete parameters. By abuse of notation, and because simple multiplexes (with a unique input and a unique output) are replaced by a labeled arrow, resources can be either multiplexes or variables.

Parameter		**Parameter**		**Parameter**		**Parameter**	
$K_{SK,\{\}}$	$=0$	$K_{B,\{\}}$	$=0$	$K_{A,\{\}}$	$=0$	$K_{En,\{\}}$	$=0$
$K_{SK,\{A\}}$	>0	$K_{B,\{A\}}$	-	$K_{A,\{En\}}$	$=0$	$K_{En,\{A\}}$	$=0$
$K_{SK,\{SK\}}$	$=0$	$K_{B,\{m\}}$	-	$K_{A,\{B\}}$	-	$K_{En,\{SK\}}$	-
$K_{SK,\{A,SK\}}$	$=2$	$K_{B,\{A,m\}}$	$=1$	$K_{A,\{SK\}}$	-	$K_{En,\{EP\}}$	-
				$K_{A,\{En,B\}}$	-	$K_{En,\{A,SK\}}$	-
		$K_{EP,\{\}}$	$=0$	$K_{A,\{En,SK\}}$	-	$K_{En,\{A,EP\}}$	-
		$K_{EP,\{B\}}$	$=1$	$K_{A,\{B,SK\}}$	-	$K_{En,\{SK,EP\}}$	-
				$K_{A,\{En,B,SK\}}$	$=1$	$K_{En,\{A,SK,EP\}}$	$=1$

Hoare triple.

Here, we view an experimental trace of the gene network as a program. The elementary instructions of this "program" are the observed transitions in the state graph of the gene network: An assignment of the form $x := x + 1$ or $x := x - 1$ corresponds to an actual observation at this time of the experiment, where the gene x has increased (which we note $x+$ for simplicity), or respectively decreased (which we note $x-$). Such an observation tells us that there is a transition somewhere in the state graph where gene x has changed its abstract expression level.

The first constraints are deduced from the classical behavior of the cell cycle: Its cyclic behavior starts from 00001, goes through a sequence of transitions and goes back to 00001. Both precondition and postcondition are then the constraints defining the state 00001. We have to determine the program corresponding to the sequence of transitions. Although the transitions between phases are known, the precise order of transitions inside a same phase is not. This leads us to consider twelve hypothetical sequences of transitions for the cell cycle. For lack of space, we consider only one of them leading to the following Hoare triple:

$$\left\{\begin{array}{l} SK = 0 \\ EP = 0 \\ A = 0 \\ B = 0 \\ En = 1 \end{array}\right\} SK+; SK+; En-; A+; SK-; SK-; B+; A-; EP+; En+; B-; EP- \left\{\begin{array}{l} SK = 0 \\ EP = 0 \\ A = 0 \\ B = 0 \\ En = 1 \end{array}\right\}$$

Using the genetically modified Hoare logic[1], we obtained 13 exact values of parameters, and one inequality ($K_{SK,\{A\}} > 0$), see Table 1. Among the $3^4 \times 2^{22} = 339\,738\,624$ possible parameter valuations, only $2^{13} = 8192$ valuations satisfy the constraints.

Other constraints can be deduced from experimental timing properties, which cannot be used in the purely discrete modeling framework of Thomas. So, we switch to our enriched hybrid modeling framework to take benefit of time measures.

4.2. *Determining Celerities of Hybrid models from temporal traces*

According to Definition 4, there are 56 celerities. Among them, 6 celerities are useless because the resources and the state are incompatible: For instance $C_{SK,\{SK\},0}$ applies to states where $\eta(SK) = 0$ and SK is a resource of itself. However, when $\eta(SK) = 0$, SK cannot be a resource of itself. Moreover, the sign of many celerities is known from biological observations shown in Table 2 (the arrows give the sign of associated celerities).

Table 2. Biological observations of the 5 variables during cell cycle. The first row represents the sequence of discrete states (see Section 4.1) crossed by the classical behavior of the cell cycle.

	00001	**10001**	**20001**	**20000**	**20100**	**10100**	**00100**	**00110**	**00010**	**01010**	**01011**	**01001**
	G1			**S**				**G2**		**M**		
SK	↗	↗	Max	Max	↘	↘	Min	Min	↗	↗	↗	↗
EP	Min	Min	Min	Min	Min	Min	Min	↗	↗	Max	Max	↘
A	Min	Min	Min	↗	↗	Max	Max	↘	↘	↘	Min	Min
B	Min	Min	Min	↗	↗	↗	↗	↗	Max	Max	↘	Min
En	Max	↘	↘	Min	Min	Min	Min	Min	↗	↗	Max	Max

In order to constrain celerities, we consider that the cell cycle phase durations are 10, 8, 4 and 0.5 hours for phases G1, S, G2 and M[11]. The time spent in each of qualitative states is approximated assuming that the duration of each phase is uniformly distributed in each of its states. This hypothesis will have consequences on the constraints on celerities. Adopting a reasoning way similar to the one of Section 4.1 (using Hoare logic on discrete models), we determine precisely 5 accurate points along the cell cycle: These 5 hybrid states allow us to deduce by Definition 6 relationships between time spent in some qualitative states and associated celerities.

In the sequel, we suppose that boolean variables are attracted towards the external boundaries because this hypothesis does not change the reachability properties. Assuming this and according to Section 2.1, the discrete parameters allow the determination of the sign of the celerities (and conversely): If $K_{v,\omega} = 1$, then $C_{v,\omega,n} > 0$; If $K_{v,\omega} = 0$, then $C_{v,\omega,n} < 0$. Thus, we get the sign of 6 celerities leading to Table 3, which contains 38 constraints on celerities.

4.3. *Additional Biological Observations*

To determine parameters controling behaviors outside of the classical cell cycle behavior, we have to take into account alternative observed trajectories, in particular about the quiescent phase and the endoreplication phenomenon.

(1) We suppose $K_{B,\{A\}} = 0$ because EP and En inhibitors of B outweigh the activator A (A is ressource of B and m is not). Moreover, maintaining mitotic inhibitors involved in EP prevents the continuation of the cell cycle. Thus, it

Table 3. Constraints on celerities.

En	$C_{En,\{\},0} < 0$	**EP**	$C_{EP,\{\},0} < -\frac{3}{\Delta t_{G1}}$
	$C_{En,\{\},1} < 0$		$C_{EP,\{\},1} = -\frac{3}{\Delta t_M}$
	$C_{En,\{A\},0} < -\frac{4}{\Delta t_S}$		$C_{EP,\{B\},0} = \frac{1}{\Delta t_{G2}}$
	$C_{En,\{A\},1} = -\frac{3}{2*\Delta t_{G1}}$		$C_{EP,\{B\},1} > \frac{3}{\Delta t_M}$
	$C_{En,\{SK\},0} < 0$	**B**	$C_{B,\{\},0} < -\frac{3}{\Delta t_M}$
	$C_{En,\{SK\},1} < 0$		$C_{B,\{\},1} = -\frac{3}{\Delta t_M}$
	$C_{En,\{A,SK\},0} < \frac{2}{\Delta t_{G2}}$		$C_{B,\{m\},0} < \frac{4}{\Delta t_S}$
	$C_{En,\{A,SK\},1} > 0$		$C_{B,\{m\},1} > \frac{2}{\Delta t_{G2}} - C_{B,\{A,m\},1}$
	$C_{En,\{A,EP,SK\},0} = \frac{6-3*\Delta t_{G2}*C_{En,\{A,SK\},0}}{2*\Delta t_M}$		$C_{B,\{A,m\},0} = \frac{4-\Delta t_S*C_{B,\{m\},0}}{3*\Delta t_S}$
	$C_{En,\{A,EP,SK\},1} > \frac{3}{\Delta t_M}$		$C_{B,\{A,m\},1} < \frac{2}{\Delta t_{G2}}$
A	$C_{A,\{\},0} < -\frac{6-(3*\Delta t_{G2}+2*\Delta t_M)*\lvert C_{A,\{En\},0}\rvert}{2*\Delta t_M}$	**SK**	$C_{SK,\{\},0} < -\frac{4}{\Delta t_S}$
	$C_{A,\{\},1} < 0$		$C_{SK,\{A\},0} = \frac{6}{2*\Delta t_{G1}+6*\Delta t_M+3*\Delta t_{G2}}$
	$C_{A,\{B\},0} < 0$		$C_{SK,\{SK\},1} = -\frac{4}{\Delta t_S}$
	$C_{A,\{B\},1} < 0$		$C_{SK,\{SK\},2} = -\frac{4}{\Delta t_S}$
	$0 > C_{A,\{En\},0} > -\frac{2}{\Delta t_{G2}}$		$C_{SK,\{A,SK\},1} = \frac{3}{\Delta t_{G1}}$
	$C_{A,\{En\},1} = -\frac{2}{\Delta t_{G2}}$		$C_{SK,\{A,SK\},2} > \frac{3}{\Delta t_{G1}}$
	$C_{A,\{B,En\},0} > 0$		
	$C_{A,\{B,En\},1} > \frac{4}{\Delta t_S} - C_{A,\{B,En,SK\},1} > 0$		
	$C_{A,\{B,SK\},0} < 0$		
	$C_{A,\{B,SK\},1} < 0$		
	$C_{A,\{B,En,SK\},0} = \frac{4}{\Delta t_S}$		
	$0 < C_{A,\{B,En,SK\},1} < \frac{4}{\Delta t_S}$		

is assumed that the presence of EP keeps the activity of En: $K_{En,\{SK,EP\}} = K_{En,\{A,EP\}} = K_{En,\{EP\}} = 1$ (using also Snoussi's condition).

(2) The endoreplication begins in mitosis and ends in G1 without going through cell division. Because endoreplication skips the first step of G1 and the duration of G1 is shortened[10], we deduce that the trajectory going from mitosis (skipping cell division) to the first step of G1 is forbidden, leading to the constraint $C_{A,\{SK,En\},0} < 0$.

(3) The quiescent phase is reachable from the end of mitosis. Because we can stay in G0 for a long time, the model has to present a cyclic behavior inside the G0 phase. Among the two possible cyclic sequences of transitions inside G0 states, only one is compatible with previous constraints.

These remarks allowed us to find the set of useful celerities and the 26 discrete parameters of our model (see Table 4). Note that the discrete parameters do not make all regulations of Figure 4 functional because the discretisation is too rough to keep each regulation. At the opposite, each regulation is visible in terms of celerities.

Table 4. Discrete parameters.

Parameter		**Parameter**		**Parameter**		**Parameter**	
$K_{SK,\{\}}$	$=0$	$K_{B,\{\}}$	$=0$	$K_{A,\{\}}$	$=0$	$K_{En,\{\}}$	$=0$
$K_{SK,\{A\}}$	$=2$	$K_{B,\{A\}}$	$=0$	$K_{A,\{En\}}$	$=0$	$K_{En,\{A\}}$	$=0$
$K_{SK,\{SK\}}$	$=0$	$K_{B,\{m\}}$	$=1$	$K_{A,\{B\}}$	$=0$	$K_{En,\{SK\}}$	$=0$
$K_{SK,\{A,SK\}}$	$=2$	$K_{B,\{A,m\}}$	$=1$	$K_{A,\{SK\}}$	$=0$	$K_{En,\{EP\}}$	$=1$
				$K_{A,\{En,B\}}$	$=1$	$K_{En,\{A,SK\}}$	$=1$
		$K_{EP,\{\}}$	$=0$	$K_{A,\{En,SK\}}$	$=0$	$K_{En,\{A,EP\}}$	$=1$
		$K_{EP,\{B\}}$	$=1$	$K_{A,\{B,SK\}}$	$=0$	$K_{En,\{SK,EP\}}$	$=1$
				$K_{A,\{En,B,SK\}}$	$=1$	$K_{En,\{A,SK,EP\}}$	$=1$

5. Simulations

Several simulations were performed with parameters (celerities and initial states) picked to satisfy constraints of Section 4.2:

- Equality constraints give exact values of celerities.
- Inequality constraints lead to choices: For each constraint of the form $C < 0$ (resp. $C > 0$), we choose $C = -1$ (resp. $C = 1$). For each constraint of the form $0 < C < k$ (resp. $0 < k < C$) where k is a constant, we choose $C = 0.5 \times k$ (resp. $C = 1.5 \times k$). For negative celerities, we choose symmetrically their exact values by multiplying the constant of inequalities by 0.5 or 1.5.
- The chosen initial state is one of the hybrid states determined in Section 4.2.

Fig. 5 shows the traces for such a simulation: After 22.5 hours (simulated time), the simulation goes back exactly to the initial hybrid state from which the same behavior can be repeated infinitely. This cyclic behavior corresponds to the classical cell cycle trajectory. Moreover, it is a limit cycle because, when starting from a neighborhood of this cyclic trace, the limit cycle is reached again after a sufficient time.
Other simulations showed that endoreplication and remaining in the quiescent phase are also possible in the model (results not shown for lack of space).

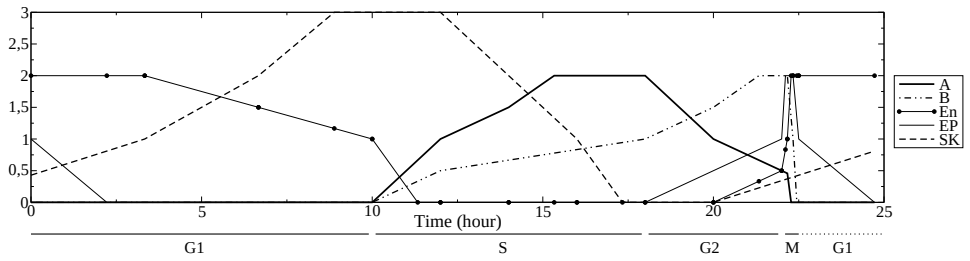


Fig. 5. A simulation representing the classical cell cycle behavior.

6. Conclusion

We developed in this article a hybrid formalism which is well suited for modeling time dependant biological phenomena such as the mammalian cell cycle. Indeed, we built a model which exhibits the limit cycle representing the classical cell cycle, the endoreplication and the quiescent phase. This has been made possible because of the wise choice of variables, in particular with the split of the cyclin complexes. The parameter choice was aided by the determination of constraints, but exact values rely on hypotheses (constants for exact values of celerities, equidistribution of time on all qualitative states within a phase). This model constitutes, to our knowledge, the first hybrid model of the cell cycle, which provides the proof of concept of our computer-aided method for parameter identification.

Our simulations did not allow us to observe traces remaining a long time in the quiescent phase, this constitutes the main limitation of our model. We have now to accurately tune the parameter values in order to get a model presenting such a trace but we didn't find any parameter valuations satisfying this specification. If no valuation of parameters is compatible with a long stay in the quiescent phase, to improve our hybrid model, we could also make the hypothesis of a regulator which would modify the celerities during this phase.

Moreover, the model presented here is based on only one of the 12 possible sequences of transitions of the cell cycle (remember that the exact order of the sequence of transitions is not known). The next time-consuming task is now to take into account the other sequences which would lead to other possible parameter constraints and then to other hybrid traces. More generally, we have to handle a large set of possible sequences of transitions. Hopefully, this difficulty can be overstepped because our hybrid modeling framework can be assisted by computer-aided methods to determine the set of parameters. Thus, our ongoing work is also to develop an automation of the building of constraints on celerities and to study the predictability of our hybrid model.

Finally, a long term objective is to make a coupling of our cell cycle model with a model of the circadian clock[16], in order to apprehend the interactions between these cycles whose functioning plays an important role in oncology[8].

Acknowledgements

We are grateful to B. Miraglio and the different reviewers for their comments to improve the paper. This work is partially supported by the French National Agency for Research (ANR-14-CE09-0011) and ANR "Investments for the Future" LABEX SIGNALIFE program (ANR-11-LABX-0028-01).

References

1. Bernot G, Comet JP, Khalis Z, Richard A, Roux O, A Genetically Modified Hoare Logic, *ArXiv* **abs/1506.05887**, 2015.

2. 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* **229**(3):339–347, 2004.
3. Bernot G, Comet JP, Snoussi EH, *Logical Modeling of Biological Systems*. chap. Formal methods applied to gene network modelling, preprint of 36 pages, 2014.
4. Bernot G, Sriram K, Képès F, Discrete delay model for the mammalian circadian clock, *ComPlexUs* **3**:185–199, 2006.
5. Comet JP, Bernot G, Das A, Diener F, Massot C, Cessieux A, Simplified models for the mammalian circadian clock, *Procedia Computer Science* **11**:127–138, 2012.
6. Comet JP, Fromentin J, Bernot G, Roux O, A formal model for gene regulatory networks with time delays, *proceedings of CSBio'2010*, CCIS, Vol. 115, Bangkok (Thailand), pp. 1–13, 2010.
7. Corblin F, Fanchon E, Trilling L, Applications of a formal approach to decipher discrete genetic networks, *BMC bioinformatics* **11**(1):385, 2010.
8. De Maria E, Fages F, Soliman S, On coupling models using model-checking: Effects of irinotecan injections on the mammalian cell cycle, *CMSB*, pp. 142–157, 2009.
9. DePamphilis ML, de Renty CM, Ullah Z, Lee CY, The octet: eight protein kinases that control mammalian DNA replication, *Frontiers in physiology* **3**:368, 2012.
10. Edgar BA, Orr-Weaver TL, Endoreplication cell cycles: more for less, *Cell* **105**(3):297–306, 2001.
11. Feillet C, *et al.*, Phase locking and multiple oscillating attractors for the coupled mammalian clock and cell cycle, *PNAS* pp. 9828–9833, 2014.
12. Gérard C, Goldbeter A, A skeleton model for the network of cyclin-dependent kinases driving the mammalian cell cycle, *Interface Focus* **1**:24–35, 2010.
13. Kellogg DR, Wee1-dependent mechanisms required for coordination of cell growth and cell division, *Journal of Cell Science* **116**(24):4883–4890, 2003.
14. Kohn KW, Molecular interaction map of the mammalian cell cycle control and DNA repair systems, *Molecular biology of the cell* **10**(8):2703–2734, 1999.
15. Larochelle S, *et al.*, Requirements for Cdk7 in the assembly of Cdk1/cyclin B and activation of Cdk2 revealed by chemical genetics in human cells, *Mol cell* **25**(6):839–850, 2007.
16. Muter J, *et al.*, The clock protein period 2 synchronizes mitotic expansion and decidual transformation of human endometrial stromal cells, *The FASEB J* pp. 13–14, 2015.
17. Satyanarayana A, Kaldis P, Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms, *Oncogene* **28**(33):2925–2939, 2009.
18. Snoussi EH, Qualitative dynamics of piecewise-linear differential equations: a discrete mapping approach, *Dynamics and stability of Systems* pp. 180–207, 1989.
19. Spencer S, *et al.*, The proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit, *Cell* **155**(2):369–383, 2013.
20. Thomas R, Boolean formalization of genetic control circuits, *Journal of theoretical biology* **42**(3):563–585, 1973.
21. Thomas R, Kaufman M, Multistationarity, the basis of cell differentiation and memory. II., *Chaos: An Interdisciplinary J of Nonlinear Science* **11**(1):180–195, 2001.
22. Trunnell NB, Poon AC, Kim SY, Ferrell JE, Ultrasensitivity in the regulation of Cdc25C by Cdk1, *Molecular cell* **41**(3):263–274, 2011.
23. Tyson JJ, Novak B, Temporal organization of the cell cycle, *Current Biology* **18**(17):R759–R768, 2008.