



HAL
open science

Modélisation formelle de comportements cycliques biologiques avec points de contrôle: la régulation du cycle cellulaire

Déborah Boyenval

► **To cite this version:**

Déborah Boyenval. Modélisation formelle de comportements cycliques biologiques avec points de contrôle: la régulation du cycle cellulaire. Bio-informatique [q-bio.QM]. Université Côte d'Azur, 2022. Français. NNT: 2022COAZ4067 . tel-04032514

HAL Id: tel-04032514

<https://theses.hal.science/tel-04032514>

Submitted on 16 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

Modélisation formelle de comportements cycliques biologiques avec points de contrôle: la régulation du cycle cellulaire

Déborah BOYENVAL

Laboratoire d'Informatique Signaux et Systèmes de Sophia Antipolis (I3S)
UMR7271 Université Côte d'Azur CNRS

**Présentée en vue de l'obtention
du grade de docteur en Informatique
de l'Université Côte d'Azur**

Dirigée par : Gilles BERNOT
Co-dirigée par : Franck DELAUNAY
Co-encadrée par : Jean-Paul COMET
Thèse soutenue le 15 décembre 2022

Devant le jury, composé de :

François FAGES, Directeur de Recherche,
Inria Saclay
Olivier ROUX, Professeur des universités,
École Centrale de Nantes
Laurence CALZONE, Ingénieure de re-
cherche, Institut Curie
Hélène COLLAVIZZA, Maître de confé-
rence, Université Côte d'Azur
Clémence FRIOUX, Chargée de recherche,
Inria Bordeaux
Gilles BERNOT, Professeur, Université Côte
d'Azur
Jean-Paul COMET, Professeur, Université
Côte d'Azur
Franck DELAUNAY, Professeur, Université
Côte d'Azur

**MODÉLISATION FORMELLE DE COMPORTEMENTS CYCLIQUES
BIOLOGIQUES AVEC POINTS DE CONTRÔLE: LA RÉGULATION
DU CYCLE CELLULAIRE**

*Formal modelling of cyclic biological behaviours with checkpoints:
the cell cycle regulation*

Déborah BOYENVAL



Jury :

Rapporteurs

François FAGES, Directeur de Recherche, Inria Saclay
Olivier ROUX, Professeur des universités, École Centrale de Nantes

Examineurs

Laurence CALZONE, Ingénieure de recherche, Institut Curie
Hélène COLLAVIZZA, Maître de conférence, Université Côte d'Azur
Clémence FRIoux, Chargée de recherche, Inria Bordeaux
Gilles BERNOT, Professeur, Université Côte d'Azur
Jean-Paul COMET, Professeur, Université Côte d'Azur
Franck DELAUNAY, Professeur, Université Côte d'Azur

Déborah BOYENVAL

Modélisation formelle de comportements cycliques biologiques avec points de contrôle: la régulation du cycle cellulaire

xiii+163 p.

Résumé

Le cycle cellulaire est une séquence d'évènements *via* lesquels une cellule se reproduit tout en préservant l'intégrité de son génome. Cette propriété de préservation est assurée par un ensemble de mécanismes de régulation appelés *checkpoints*. Leur capacité à arrêter le cycle cellulaire en réponse à des lésions de l'ADN motive de nombreuses recherches dans le secteur biomédical. Cette thèse défend l'idée que ce phénomène d'arrêt, observé dans un contexte de perturbation, est un corollaire de propriétés plus fondamentales telles que la séparation temporelle des phases et l'irréversibilité des transitions de phases. Une phase décrit dans ce contexte un aspect particulier d'une cellule en prolifération, par exemple la phase de réplication du génome (phase S) ou celle de division de la cellule (phase M). Les propriétés fondamentales des checkpoints n'ont jamais été formalisées essentiellement parce que la modélisation des phases n'avait elle-même jamais été abordée jusque-là. La thèse présente un modèle discret de la régulation du cycle cellulaire et propose une méthodologie de formalisation d'une phase. Elle est initialement décrite par un multiensemble fini d'évènements discrets qui ont lieu à partir d'un état initial. En choisissant le cadre de modélisation de René Thomas, on s'autorise à appliquer la logique de Hoare génétiquement modifiée pour calculer l'ensemble des séquences d'évènements, exhibées par le multiensemble initial, qui sont compatibles avec un certain ensemble de paramétrisations d'un graphe de régulation. L'ensemble des séquences d'évènements qui caractérise une phase isole temporellement cette phase vis-à-vis des autres s'il existe une paramétrisation telle que tout évènement qui peut clore une phase est requis avant tout évènement qui peut initier la phase suivante. Ce problème a été spécifié à l'aide d'un prédicat Prolog. La propriété d'irréversibilité des transitions de phases est quant à elle formalisée à l'aide d'une logique temporelle et le model-checking identifie les paramétrisations d'un modèle du cycle cellulaire qui satisfont cette propriété. Nous montrons que cette méthodologie de formalisation et de vérification des propriétés fondamentales des checkpoints permet de cibler les contradictions d'un graphe de régulation en s'affranchissant de certains choix arbitraires que l'on retrouve dans la plupart des études de modélisation du cycle cellulaire.

Mots-clés : Modélisation du cycle cellulaire, formalisation des phases, formalisation des checkpoints intrinsèques, René Thomas, logique de Hoare, model-checking, Prolog, TotemBioNet.

Abstract

The cell cycle is a series of events by which a cell reproduces itself while maintaining the integrity of its genome. This property is ensured by a set of regulatory mechanisms called *checkpoints*. Their ability to stop the cell cycle in response to DNA damage has motivated numerous studies in the field of biomedicine. This thesis argues that the phenomenon of cell cycle arrest, which is observed in the context of perturbation, is a consequence of more fundamental properties such as the temporal separation of phases and the irreversibility of phase transitions. In this context, a phase describes a particular aspect of a proliferating cell, for example the phase of genome replication (S-phase) or that of cell division (M-phase). The fundamental properties of checkpoints have never been formalized, mainly due to the fact that phase modelling itself has never been tackled before. The thesis presents a discrete model of the regulation of the cell cycle and proposes a methodology for formalizing a phase. The latter is firstly characterized by a finite multi-set of discrete events which occur from an initial state. By opting for René Thomas' modelling framework, the genetically modified Hoare logic can be applied to compute the set of sequences of events, exhibited by the initial multiset, that are consistent with a certain set of parameterizations of a regulation graph. The set of event sequences that characterizes a phase temporally insulates it from the others if there is a parameterization such that any event that can complete a phase is required before any event that can trigger the next phase. This problem has been specified using a Prolog predicate. The irreversibility property of phase transitions is formalized using a temporal logic and model-checking identifies the parameterizations of a cell cycle model that satisfy this property. We show that this method for formalizing and verifying the fundamental properties of checkpoints allows us to target the inconsistencies of a regulatory graph while getting rid of certain arbitrary choices found in most cell cycle modeling studies.

Keywords : Cell cycle modelling, phase formalization, intrinsic checkpoint formalization, René Thomas, Hoare logic, model-checking, Prolog, TotemBioNet.

Remerciements

Mes remerciements vont en priorité à mes directeurs de thèse : Gilles, Jean-Paul et Franck. Vous avez toujours une longueur d'avance mais j'ai pu réaliser une thèse qui me reflète, qui m'a permis de définir des points de repère intellectuels et axiologiques et qui représente désormais un socle intellectuel et logique indéniable pour moi.

Gilles, tu m'as conseillé ce livre qui a tant résonné, *le traité du zen et de l'entretien des motocyclettes*, au moment où j'avais besoin de trouver des repères face à l'absurdité parfois de la science elle-même. J'ai rencontré *Robert Pirsig* grâce à toi. L'aboutissement de la lecture de ce livre a coïncidé avec la fin de ma thèse, joli symbole (petit clin d'oeil à Valériane). Une petite parenthèse, puisque l'on aborde une aventure littéraire marquante, Cyrille je te suis aussi reconnaissante de m'avoir conseillé le livre *Genesis* à la suite de ton excellente conférence grand public à propos du transhumanisme. Plus généralement, ta richesse intérieure m'a éblouie. Gilles, Jean-Paul, Franck, je vous suis reconnaissante de m'avoir transmis une partie de votre expérience scientifique et humaine et je vous souhaite de collaborer encore par la suite si le coeur vous en dit. Gilles, je suis obligée de faire cette petite boutade : ton radotage a fini par payer, je pense pouvoir faire une liste quasi exhaustive des conseils de raisonnement ou de rédaction ainsi que des anecdotes que tu m'as si souvent répétées.

Je remercie chaleureusement tous les membres de mon jury d'avoir accepté de m'accompagner dans cette étape d'aboutissement de thèse. François, Olivier, je suis honorée que vous ayez accepté de rapporter ma thèse. Vos conclusions m'ont été très précieuses. François, cette phrase te dira quelque chose mais elle est applicable à tous les membres de mon jury, je n'ai pas encore cité Laurence, Clémence et Hélène : je vous suis reconnaissante pour toutes vos critiques essentielles à la robustesse de l'argumentation de ma thèse, surtout avant publication de mon article de thèse. Hélène, tu as été un soutien très efficace en me poussant à valoriser mon travail à des moments clés, je te remercie pour ton suivi et tes conseils.

En vrac, je remercie Oussama d'avoir été un collègue de bureau drôle et humainement précieux. Dino, merci pour tous ces moments caféinés et d'avoir tenté si souvent de me convertir à la course à pied, soyons sérieux ça n'arrivera pas ! Je remercie Valériane de m'avoir transmis avec générosité des astuces oratoires, tes talents de conférencières sont indéniables. Je garderai en tête ta leçon de bienveillance avant tout envers soi-même quand on abat un travail acharné et passionné. Je remercie ensuite Natacha pour ce séjour en Bretagne qui a considérablement adouci ma rédaction, sans compter tous ces repas dominicaux, si plaisants, avec Jean-Pierre. Dédicace également à Aude et Pierre pour ces parties de gin. Agathe, quel plaisir cela a été de te résumer ma thèse. Grâce à ta patience et ta curiosité, tu as contribué à dépasser ma fatigue de rédaction au moment très délicat de la rédaction de l'introduction et de la conclusion. Je remercie tous les membres fondateurs de l'association DS4H ADAMS avec qui j'ai passé du très bon temps. Anne-Laure, j'ai admiré ton grand professionnalisme, ton esprit de communauté et ton ouverture d'esprit. Alexandre, Dorian, je vous suis reconnaissante de m'avoir soutenue lors du concours de thèse pour la bourse régionale et d'avoir pris le temps de garder un pied dans un projet de

recherche fondamentale en cours. Je n'oublie pas mes professeurs, Patrick Coquillard et Francine Diener, qui ont contribué de façon déterminante à m'orienter vers une bio-informatique théorique qui colle avec mon amour (longtemps contrarié) des mathématiques. Je remercie les membres de mon CSI, Fabienne et Sabine. Je remercie aussi Magali pour ta précieuse efficacité. Merci à tous ceux qui ont croisé ma route lors de ma thèse et qui ont contribué de près ou de loin à son aboutissement, j'ai appris en chacun de vous. Une liste non exhaustive : Emilien, Jonathan, Ophélie, Benjamin, Laetitia, Amélie, Rajeev, Nina, Jade, Camille. Merci à tous ces grands fous qui ont participé à une cagnotte de thèse indécente. Mes (avant) derniers remerciements vont à mes parents et au reste de ma famille avec qui il est toujours cocasse de décompresser lors de repas de fête.

Et enfin il y a Julien. On s'est connus jusqu'à présent uniquement dans ce contexte de préparation de mon doctorat, avec toutes les remises en question qu'elle a provoquées. Comme tu le répètes si bien, une bonne thèse est une thèse terminée et maintenant que le point final de cette expérience est atteint, je piétine d'impatience de voir ce que nous réserve l'avenir. Sans toi, tout aurait été plus difficile, pas infaisable, mais parfois insupportable. Cette aventure de thèse a marqué un tournant dans ma façon de penser et d'être, je suis heureuse que tu en aies été témoin. Ce que l'on a vécu est indélébile et tu as été une inspiration déterminante dans la définition de mes objectifs moraux élémentaires : l'intégrité, la capacité à apprendre de ses erreurs et l'art de dédramatiser la plupart des situations délicates. On a découvert ensemble des penseurs, l'escalade, des séries inoubliables et n'oublions pas le plus important : les aventures "poutrinesques" (adjectif qui nous est propre et qui renferme ô combien de délires, un jour je tenterai d'en faire un inventaire) qui ne cesseront, je pense, de me faire rire. Tu as somme toute été un partenaire plus qu'exceptionnel lors de cette thèse, d'une résistance à toute épreuve. La pudeur m'oblige à conclure. Comme disait Jankelevitch, restons encore dans le mystère de notre amour et, de la sorte, puisse t'il encore nous galvaniser et nous permettre de nous dépasser au sens de nous aider à nous construire comme on estime être le plus élégant.

~_(\ツ)_/

Table des matières

| | |
|--|-----------|
| Notations | 1 |
| 1 Introduction | 3 |
| 1.1 Définition du cycle cellulaire | 5 |
| 1.2 Le contrôle du cycle cellulaire | 6 |
| 1.2.1 Les acteurs moléculaires clés de la régulation du cycle cellulaire | 6 |
| 1.2.2 Polysémie du terme « checkpoint » dans le contexte du cycle cellulaire | 7 |
| 1.3 Motivations | 9 |
| 1.4 Vue d'ensemble de la thèse | 11 |
| État de l'art | |
| 2 Les formalismes pour modéliser la régulation de comportements cycliques | 15 |
| 2.1 Un préambule épistémologique : pourquoi formaliser ? | 17 |
| 2.2 Les systèmes équations différentielles ordinaires | 17 |
| 2.3 Les réseaux booléens aux réseaux multivalués | 22 |
| 2.4 Approche multivaluée avec multiplexes : le formalisme de modélisation de René Thomas | 24 |
| 2.4.1 Syntaxe abstraite : graphe d'influence avec multiplexes | 24 |
| 2.4.2 De la syntaxe à la sémantique : la notion de paramètre dynamique | 29 |
| 2.4.3 Sémantique : le graphe de transition asynchrone | 35 |
| 3 Une revue des modèles incontournables de la régulation du cycle cellulaire | 39 |
| 3.1 Préambule | 41 |
| 3.2 Les modèles continus pionniers | 43 |
| 3.2.1 Identification d'une structure de régulation : les switches bistables | 43 |
| 3.2.2 Deux propriétés intrinsèquement discrètes du phénomène de checkpoint | 45 |
| 3.3 Les modèles discrets et la formalisation du phénomène de checkpoint | 47 |
| 3.3.1 Un exemple de modélisation discrète des régulations impliquées dans certains checkpoints | 47 |
| 3.3.2 Un exemple de formalisation de certaines contraintes chronologiques au cours du cycle cellulaire | 48 |
| 3.3.3 Une première tentative de formalisation des phases du cycle cellulaire | 50 |
| 3.4 La formalisation des phases munies de checkpoints : la question reste ouverte | 52 |
| 3.4.1 Relier le mécanisme des checkpoints à ses deux propriétés dynamiques | 52 |
| 3.4.2 Les checkpoints intrinsèques et extrinsèques | 53 |
| 3.5 Bilan sur la modélisation des checkpoints | 54 |

Méthodes formelles utilisées

| | | |
|----------|---|-----------|
| 4 | Apport de deux méthodes formelles pour les réseaux de régulations discrets | 59 |
| 4.1 | Un préambule épistémologique : conjonctions et réfutations | 61 |
| 4.2 | Model-checking de formules CTL | 62 |
| 4.2.1 | Syntaxe de CTL | 63 |
| 4.2.2 | Sémantique de CTL | 63 |
| 4.2.3 | Principe du model-checking | 66 |
| 4.2.4 | CTL et les chemins équitables | 67 |
| 4.2.5 | Quelques exemples de propriétés formalisées en CTL (équitable) | 68 |
| 4.3 | Logique de Hoare génétiquement modifiée | 69 |
| 4.3.1 | Syntaxe | 70 |
| 4.3.2 | Sémantique | 71 |
| 4.3.3 | Règles d'inférence | 72 |
| 4.3.4 | Calcul de la plus faible précondition | 73 |
| 4.4 | Gestion de l'exhaustivité des modèles avec TotemBioNet | 74 |
| 4.4.1 | Entrées et sorties | 75 |
| 4.4.2 | Workflow | 75 |

Contributions de la thèse

| | | |
|----------|---|------------|
| 5 | Notre ensemble de modèles pour l'étude des phases du cycle cellulaire | 81 |
| 5.1 | Abstraction et question de modélisation | 83 |
| 5.2 | Les variables et leurs seuils discrets | 84 |
| 5.3 | Nos conditions initiales | 86 |
| 5.4 | Formalisation des régulations impliquées dans la progression du cycle cellulaire | 88 |
| 5.4.1 | Notre graphe d'influence conçu pour la définition d'un cycle cellulaire canonique | 88 |
| 5.4.2 | Une séquence canonique d'évènements de régulation au cours des phases du cycle cellulaire | 92 |
| 5.4.3 | Définition des multiplexes | 98 |
| 5.5 | Première identification des valeurs des paramètres | 102 |
| 5.5.1 | Vérification de la stabilité de la phase de quiescence | 103 |
| 5.5.2 | Formalisation d'un cycle cellulaire canonique comme un attracteur cyclique | 104 |
| 5.5.3 | Se poser les bonnes questions sur les checkpoints à l'aide de TotemBioNet | 104 |
| 5.5.4 | Identification manuelle de certains paramètres | 106 |
| 6 | Formalisation des phases du cycle cellulaire | 111 |
| 6.1 | Préambule | 113 |
| 6.2 | Définition syntaxique d'une phase canonique | 114 |
| 6.3 | Formalisation d'une phase à partir de sa définition canonique | 115 |
| 6.3.1 | Définition de l'hyperpavé d'un triplet de Hoare | 115 |
| 6.3.2 | Vers une formalisation d'une phase sous la forme d'un prédicat du premier ordre | 117 |
| 6.4 | Application à notre modèle du cycle cellulaire | 118 |

| | | |
|----------------|---|------------|
| 6.4.1 | Les phases canoniques de notre modèle | 118 |
| 6.4.2 | Les hyperpavés de nos phases canoniques | 120 |
| 6.4.3 | Les phases de notre modèle | 123 |
| 7 | Formalisation d'un checkpoint entre deux phases | 129 |
| 7.1 | Préambule | 131 |
| 7.2 | Le prédicat générique <i>séparation – temporelle</i> (π_i, π_{i+1}) | 132 |
| 7.2.1 | Les prédicats <i>peutClôre</i> et <i>peutInitier</i> | 132 |
| 7.2.2 | Le prédicat <i>estRequis</i> | 133 |
| 7.3 | Irréversibilité des transitions de phases | 135 |
| 7.4 | Application à notre modèle et nos phases du cycle cellulaire | 137 |
| 7.4.1 | Implémentation du prédicat <i>séparation – temporelle</i> en Prolog | 137 |
| 7.4.2 | Notre graphe de régulation et ses phases munies de checkpoint | 138 |
| 7.4.3 | La preuve de concept | 140 |
| 8 | Conclusion et perspectives | 143 |
| 8.1 | Bilan | 143 |
| 8.1.1 | Contributions méthodologiques | 144 |
| 8.1.2 | Contributions pour la modélisation du cycle cellulaire | 144 |
| 8.2 | Perspectives | 145 |
| 8.2.1 | Inférence de modèles à l'aide de nos prédicats | 145 |
| 8.2.2 | À propos de notre modélisation des checkpoints du cycle cellulaire des mammifères | 146 |
| 8.2.3 | Quelques applications en biologie | 146 |
| Annexes | | |
| A | Paramétrisation de notre GRBM du cycle cellulaire | 159 |

Liste des définitions

| | | |
|----|---|-----|
| 1 | Fonctions de Hill | 18 |
| 2 | Fonctions step de Heavyside | 21 |
| 3 | Graphe de régulation biologique avec multiplexe | 25 |
| 4 | Famille des paramètres dynamiques | 29 |
| 5 | État | 29 |
| 6 | Paramètre applicable | 30 |
| 7 | Réseau de régulation biologique avec multiplexe | 30 |
| 8 | Condition de Snoussi | 31 |
| 9 | Graphe de transition asynchrone | 35 |
| 10 | Arc non fonctionnel | 36 |
| 11 | Formule CTL | 63 |
| 12 | Structure de Kripke | 64 |
| 13 | Fonction d'étiquetage | 64 |
| 14 | Sémantique d'une formule CTL à l'aide d'une structure de Kripke | 65 |
| 15 | Relation de satisfaction d'une formule CTL par un modèle | 65 |
| 16 | Langage des assertions | 70 |
| 17 | Langage de chemin | 70 |
| 18 | Triplet de Hoare | 71 |
| 19 | Sémantique d'un chemin | 71 |
| 20 | Sémantique d'un triplet de Hoare | 72 |
| 21 | Axiomes de l'incrémentement et de la décrémentation | 72 |
| 22 | Règle d'inférence de la composition séquentielle | 73 |
| 23 | Règle d'inférence du chemin vide | 73 |
| 24 | Axiomes des bornes | 73 |
| 25 | La plus faible précondition <i>wp</i> | 74 |
| 26 | Chemin syntaxique élémentaire | 114 |
| 27 | Triplet de Hoare élémentaire | 114 |
| 28 | Hyperpavé d'un triplet de Hoare élémentaire | 115 |
| 29 | Prédicat <i>estAdmis</i> | 118 |
| 30 | Prédicats <i>peutInitier</i> et <i>peutClôre</i> | 132 |
| 31 | Prédicat <i>séparation – temporelle</i> | 133 |
| 32 | Prédicat <i>estRequis</i> | 134 |
| 33 | Prédicat <i>irréversible</i> | 135 |

Notations

| | |
|---------|---|
| CFCT | Composante Fortement Connexe Terminale |
| CPR | Complexes de Pré-Réplication |
| CTL | Computational Tree Logic |
| CYC/CDK | Complexe CYCline/Cyclin-Dependant Kinase |
| EDO | Équations Différentielles Ordinaires |
| GF | Growth Factors |
| GRBM | Graphe de Régulation Biologique avec Multiplexe |
| GTA | Graphe de Transition Asynchrone |
| LHGM | Logique de Hoare Génétiquement Modifiée |
| RRBM | Réseau de Régulation Biologique avec Multiplexe |
| SAC | Spindle Assembly Checkpoint |

CHAPITRE 1

Introduction

La thèse vise la formalisation d'un concept clé dans la régulation des systèmes biologiques : les checkpoints. La notion de checkpoint évoque un blocage d'un évènement particulier tant qu'un ensemble de tâches n'est pas réalisé. Les checkpoints biologiques les plus connus sont ceux intrinsèques au processus nécessaire à la survie de toute espèce, à savoir le cycle de division des cellules. Ce système physiologique fait partie des plus étudiés à la fois expérimentalement et mathématiquement ce qui conduit à naviguer dans une complexité grandissante. Dans cette introduction, nous prenons le parti pris de donner au lecteur non biologiste un aperçu de la complexité des checkpoints du cycle cellulaire qui implique de définir au préalable les concepts clés de la biologie du cycle cellulaire. Nous défendrons ensuite l'idée qu'un checkpoint biologique est un processus intrinsèquement discret qui repose sur un blocage plutôt que sur un retard d'un évènement de régulation biologique.

| | | |
|------------|--|-----------|
| 1.1 | Définition du cycle cellulaire | 5 |
| 1.2 | Le contrôle du cycle cellulaire | 6 |
| 1.2.1 | Les acteurs moléculaires clés de la régulation du cycle cellulaire | 6 |
| 1.2.2 | Polysémie du terme « checkpoint » dans le contexte du cycle cellulaire | 7 |
| 1.3 | Motivations | 9 |
| 1.4 | Vue d'ensemble de la thèse | 11 |

1.1 Définition du cycle cellulaire

La reproduction cellulaire est une caractéristique fondamentale de tous les êtres vivants. Le cycle cellulaire est une séquence d'évènements par lesquels les cellules se reproduisent tout en conservant l'intégrité de leur génome. L'évolution marque une tendance vers une complexification de ce processus, particulièrement chez les eucaryotes pluricellulaires [88], ce qui complique l'identification des déterminants de cette intégrité. Le cycle cellulaire est généralement décrit par la séquence des phases $G1 \rightarrow S \rightarrow G2 \rightarrow M$ où S et M sont les deux phases critiques, cf. figure 1.1. La phase S (pour « synthèse ») est une phase au cours de laquelle l'ADN est totalement répliqué.

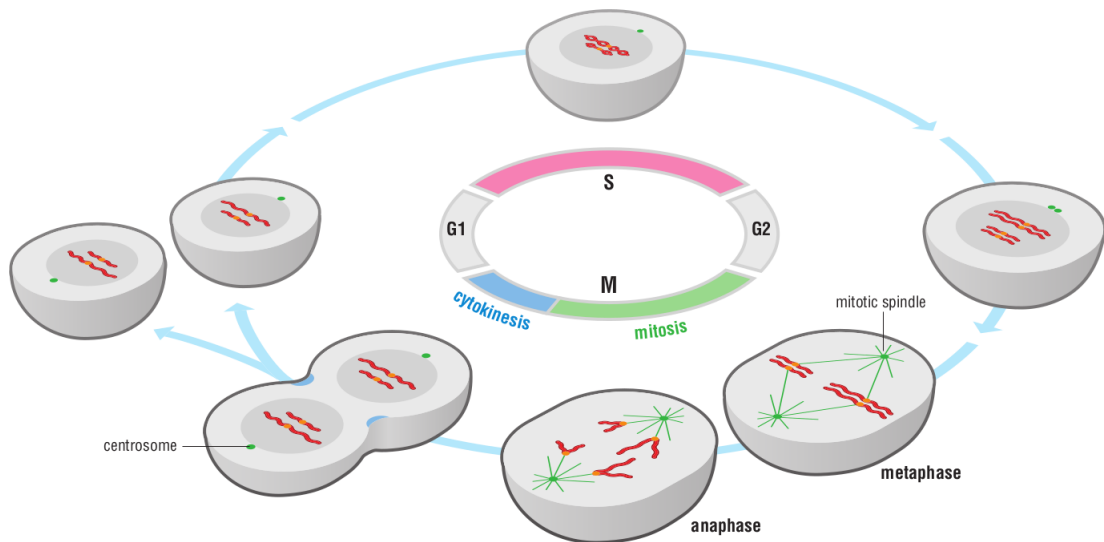


Figure 1.1 – **Le cycle cellulaire et ses quatre phases.** Les phases de répllication de l'ADN (S) et de division cellulaire (M) sont les deux phases critiques du cycle cellulaire qui doivent impérativement avoir lieu séparément. Figure extraite de [32].

La phase M (pour « Mitose ») est la phase durant laquelle la division d'une cellule mère en deux cellules filles a lieu. Expérimentalement, on distingue la phase dans laquelle se trouve une cellule en mesurant, par exemple à l'aide d'une technique de cytométrie de flux, son contenu en ADN. Une cellule en phase G1 contient deux exemplaires de chaque chromosome. On dit que la cellule est diploïde ($2n$). Une cellule en phase S voit son contenu en ADN doubler ($4n$). À la fin de la mitose, les deux cellules filles qui résultent de la division de la cellule mère sont de nouveau diploïdes.

La phase M est subdivisée en 4 sous-phases. La première est une phase de condensation de l'ADN sous la forme de chromosomes (prophase). La deuxième coïncide avec l'attachement des chromosomes au fuseau mitotique et à l'alignement des chromosomes le long d'une ligne imaginaire à équidistance des deux pôles opposés de la cellule (métaphase). Un chromosome est formé de deux chromatides soeurs (en rouge en figure 1.1) rattachées par une structure appelée la cohésine. À la toute fin de la métaphase, la cohésion entre les deux chromatides est détruite ce qui entraîne la séparation des chromatides soeurs (anaphase). Les chromatides soeurs sont ensuite déplacées, chacune à un pôle opposé, au niveau d'une structure appelée le centrosome (en vert en figure 1.1) et l'enveloppe nucléaire se reforme autour de l'ADN des deux futures cellules filles (télophase). La fin de la télophase est communément marquée par la division de la cellule mère en deux cel-

lules filles (cytokinèse).

Les phases S et M sont toutes deux précédées d'une phase de préparation nommée « gap ». La phase G1 est la phase qui précède la réplication et G2 celle qui précède la mitose. Enfin, en présence de conditions de croissance défavorables ou de signaux inhibiteurs provenant d'autres cellules, un cellule peut s'arrêter pendant de longues périodes dans un état de non-division prolongé.

1.2 Le contrôle du cycle cellulaire

Le processus du cycle cellulaire est soumis à des contrôles génétiques drastiques qui assurent la fidélité de la transmission de l'ADN de génération en génération de cellules. Le maintien de l'intégrité du génome tout au long du cycle cellulaire revêt une grande importance pour le développement des organismes et l'homéostasie des tissus et organes. Toute anomalie du cycle cellulaire peut avoir des conséquences sur ces processus fondamentaux. L'observation par microscopie du cycle cellulaire montre que ces deux processus se succèdent sans aucun chevauchement au cours d'un cycle cellulaire non pathologique [54]. De plus, lorsque ces deux processus ne sont pas séparés temporellement, l'intégrité de l'ADN est compromise pour les générations futures. En voici deux exemples. Une cellule qui entre précocement en mitose, c'est à dire avant que le génome ne soit répliqué en entier, hérite d'un ensemble incomplet de chromosomes. Une cellule qui entre correctement en mitose, c'est à dire une fois que la totalité du génome est répliquée, mais qui réalise une seconde fois la réplication peut également hériter d'un ensemble incomplet de chromosomes si la ségrégation des chromosomes a lieu avant la fin du second tour de réplication [55]. Le cycle cellulaire doit être contrôlé « en interne » pour assurer une alternance stricte de ses deux phases critiques et ainsi préserver l'intégrité de l'ADN. Nous donnerons ici les principaux mécanismes de contrôle, mis en lumière ces trente dernières années, qui :

1. rendent interdépendants les évènements de la phase S et de la phase M de sorte que le processus de mitose ne peut pas débiter tant que la réplication n'est pas terminée,
2. bloquent le cycle cellulaire et réparent l'ADN lors d'un stress génotoxique provoqué par la réplication ou la mitose elles-mêmes ou par des facteurs exogènes.

1.2.1 Les acteurs moléculaires clés de la régulation du cycle cellulaire

La progression du cycle cellulaire est contrôlée par les cyclines (abrégées *cyc*) formant un complexe moléculaire avec les kinases dépendantes des cyclines (abrégées *cdk*) [14]. Ces complexes sont notés *cyc/cdk* et on leur attribue des rôles distincts au cours du cycle cellulaire. Les complexes *cycD/cdk4-6* et *cycE/cdk2* déclenchent l'entrée au sein du cycle cellulaire lui-même depuis la phase de quiescence, les complexes *cycE-A/cdk2* participent au déclenchement de la phase S et les complexes *cycA-B/cdk1* quant à eux participent à l'initiation de la phase M (figure 1.2). En effet, les complexes *cycE/cdk2* et *cycA/cdk1* sont nécessaires respectivement à la réalisation de la transition G1/S et de la transition G2/M. Des expériences de perte de fonction ou d'inhibition chimique des complexes *cycE/cdk2* et *cycA/cdk1* sur des lignées cellulaires humaines conduisent à l'observation de blocages respectivement en phase G1 et G2 [38, 54]. Le contrôle du cycle cellulaire se base sur une activation séquentielle et transitoire de ces principaux complexes *cyc/cdk* et par extension des processus sous leur contrôle. Par exemple, l'activation de la phase M par le complexe mitotique *cycB/cdk1* est conditionnelle à l'achèvement de la réplication.

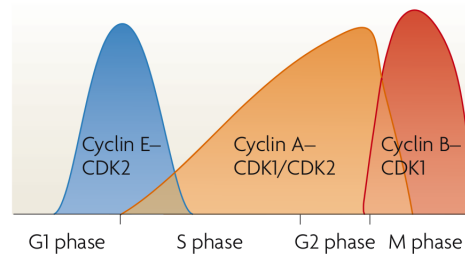


Figure 1.2 – **Les complexes cyc/cdk sont les acteurs moléculaires clés de la régulation du cycle cellulaire.** L'activité des complexes cyc/cdk représentée en ordonnée est restreinte à certaines fenêtres temporelles. Figure adaptée de [38].

1.2.2 Polysémie du terme « checkpoint » dans le contexte du cycle cellulaire

Si la cellule ne parvient pas à achever la réplication de l'ADN, un signal bloque l'initiation de la mitose. Le processus de réplication en cours d'exécution génère un signal semblable à un stress de réplication mais dans une condition non pathologique. Ce signal est capté par des protéines de surveillance de la réplication qui *bloquent* l'activation du complexe moléculaire initiateur de la mitose tant que cette dernière n'est pas terminée. Ces protéines « surveillantes » sont ATR et Chk1. Concrètement, ATR inhibe le complexe cycA/cdk1 qui est impliqué dans la préparation de la phase M. En effet, l'activation de ce complexe suffit à déclencher l'activation autonome du complexe cycB/cdk1 au moment de la transition G2/M [38, 84, 90]. ATR active en aval Chk1, une kinase qui participe notamment à la stabilisation des fourches de réplication en S. De surcroît, [90] suggèrent que la voie ATR/CHK1 repousse l'activation du complexe cycB/cdk1 *via* Cdc25/Wee1, c'est à dire de manière similaire à la réponse aux dommages de l'ADN. Dans les faits, [84] observent une première accumulation de la cycline B en phase S, chez des cellules traitées avec un inhibiteur de ATR ou de Chk1 et synchronisées au début de la phase S. Chez des cellules saines non perturbées, l'augmentation de la concentration de la cycline B (par extension de l'activité du complexe cycB/cdk1) est observée au moment de la préparation de la mitose en phase G2. L'inhibition du contrôle de l'activation du complexe cycA/cdk1 exercé par ATR et Chk1 au cours de la réplication provoque une véritable crise d'identité au cours du cycle cellulaire, où les phases S et G2 se chevauchent. Ce chevauchement conduit à une mitose prématurée, une réplication insuffisante et des dommages à l'ADN. Nous venons précisément de donner un exemple de mécanisme moléculaire assurant la préservation de l'intégrité de l'ADN, propriété que l'on attribue aux *checkpoints* du cycle cellulaire. Dans le langage courant, le terme « checkpoint » fait d'abord référence à *un point de passage qui peut être franchi uniquement si une condition est vérifiée*. Le corollaire de cette première description d'un checkpoint est la stabilité d'une phase tant que la condition du checkpoint n'est pas satisfaite [29, 86]. Le phénomène de checkpoint est donc intrinsèquement *discret*. Le checkpoint donné en exemple est S/G2 mais, historiquement, les premiers checkpoints du cycle cellulaire mis en évidence sont ceux situés aux transitions G1/S et G2/M.

Le checkpoint G1/S assure que la réplication de l'ADN n'a lieu qu'une fois par cycle. G1 est une phase de préparation de la réplication qui ne peut débuter que si (condition) les complexes de pré-réplication (abrégié par la suite CPR) sont tous recrutés aux origines de réplication. Ce complexe moléculaire est composé d'enzymes qui catalysent le déroulement des brins d'ADN,

les hélicases. Cette étape est indispensable aux étapes d'initiation et d'élongation de la réplication de l'ADN. Le chargement et l'activation des hélicases sont des événements cruciaux qui doivent avoir lieu avant l'initiation de la réplication, caractérisée quant à elle par le recrutement de l'ADN polymérase α au niveau de l'ADN simple-brin [25]. L'activation du CPR est sous le contrôle du complexe *cycE/cdk2*. En phase S, c'est à dire une fois le processus de réplication initié, les protéines nécessaires à l'assemblage des CPR sont inhibées. Ce processus est sous le contrôle du complexe *cycA/cdk2*, l'inhibiteur du complexe *cycE/cdk2* activateur des CPR. [56]. L'inhibition des CPR est une étape essentielle en phase S pour éviter plusieurs cycles de réplication et un épuisement des ressources en nucléotides. Le début du processus d'inhibition des CPR par *cycA/cdk2* doit naturellement avoir lieu avant la fin du processus d'assemblage et d'activation des CPR par la cycline E et la *cdk2* pour assurer une réplication totale du génome. À la lumière des dernières informations, le terme checkpoint évolue : *les événements tardifs (i.e. la mitose) du cycle cellulaire dépendent de l'achèvement d'événements plus précoces du cycle cellulaire (i.e. la réplication)* [7]. Dans ce cas précis, le checkpoint est dit *intrinsèque* au cycle cellulaire. Le *Spindle Assembly Checkpoint* (abrégié ensuite SAC) en est un autre exemple. Ce checkpoint coordonne deux phases de la mitose : la métaphase et l'anaphase. L'alignement des chromosomes le long de la plaque métaphasique (métaphase) doit être totalement achevé avant la séparation des deux chromatides soeurs des chromosomes (anaphase). La protéine responsable du déclenchement de l'anaphase (APC-*cdc20*) est inhibée tant que tous les chromosomes ne sont pas attachés au fuseau mitotique au cours de la métaphase [53].

Le checkpoint G2/M est un exemple de processus supplémentaire de contrôle de l'entrée en mitose. L'activité du complexe *cycB/cdk1* est réprimée au cours de la réplication *via* l'inhibition ATR-dépendante de son activateur, le complexe *cycA/cdk1*. Cette inhibition du complexe est renforcée par la protéine Wee1, une kinase qui inhibe l'activité du complexe *cycB/cdk1* en phase S et également en phase G2 [98]. La fonction de Wee1 dans le checkpoint G2/M apparaît clairement lorsque les cellules subissent des dommages à l'ADN. Il réprime l'activité des complexes *cyc/cdk* et plus particulièrement le complexe *cycB/cdk1*, le principal facteur de l'entrée en mitose, en réponse aux cassures simples ou doubles brins générées par exemple après une erreur lors du processus de réplication ou une exposition à des agents mutagènes. Wee1 participe à l'arrêt du cycle cellulaire en G2 ce qui permet la réparation de l'ADN avant la ségrégation des chromosomes [11]. Une cellule peut être soumise à des lésions de l'ADN provoquées par des facteurs environnementaux génotoxiques, par exemple des rayonnements ionisants tels que les UV. Ces lésions sont généralement des cassures de l'ADN repérées par des biosenseurs qui activent des voies de réparation de l'ADN. On retrouve ATR et Chk1 parmi les acteurs déterminants des voies de réponses aux dommages à l'ADN. Leur rôle consiste ici à stabiliser une phase tant que les dommages ne sont pas réparés [29]. En l'occurrence, Chk1 active Wee1 (l'inhibiteur de *cycB/cdk1*) et ATR inhibe Cdc25 (l'activateur de *cycB/cdk1*) et la transition vers la mitose est suspendue. La condition nécessaire à cette transition est directement l'absence de dommages à l'ADN. Le concept de checkpoint du cycle cellulaire s'enrichit : *un checkpoint bloque la progression du cycle cellulaire en réponse à des dommages à l'ADN exogènes en ciblant directement les régulateurs de la progression du cycle cellulaire*. On parlera de checkpoint *extrinsèque* du cycle cellulaire.

Intrinsèques ou extrinsèques, les checkpoints garantissent que les erreurs génétiques potentiellement létales ne seront pas transmises à la descendance de la cellule. Un dernier checkpoint subsiste : le point de restriction qui stabilise la phase G1 *via* la concentration de facteurs de

croissance. Il est par ailleurs possible de synchroniser des cellules en phase de quiescence (G0) ou en début de phase G1 par déprivation de facteurs de croissance. Le point de restriction marque le moment où une cellule devient insensible aux facteurs de croissance et s'engage irréversiblement dans le cycle cellulaire (figure 1.3). Cet engagement est directement influencé par des signaux extra-cellulaires et, de fait, le *checkpoint de restriction* est un garant du contrôle de la prolifération. L'action du checkpoint de restriction est restreinte temporellement pour enlever toute possibilité

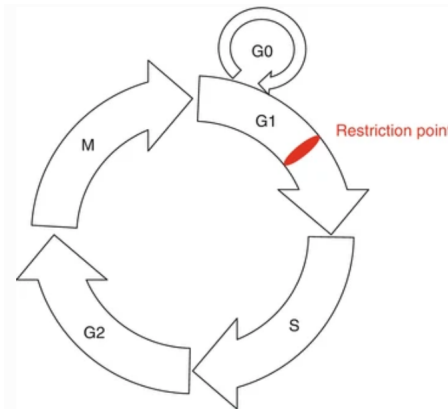


Figure 1.3 – **Le point de restriction canoniquement situé en G1.** Il correspond à la frontière qui sépare une portion du cycle cellulaire sensible à la privation en facteurs de croissance et une portion qui y est insensible. Figure adaptée de [61].

aux facteurs de croissance de contrôler la progression des phases S et M, et donc de potentiellement les interrompre alors qu'elles sont en cours d'exécution [15]. En d'autres termes, qu'importe la concentration en facteurs de croissance, les phases S et M ont obligatoirement lieu une fois que la cellule passe le point de restriction.

1.3 Motivations

Une compréhension profonde des mécanismes qui sous-tendent les checkpoints est motivée par la recherche de stratégies thérapeutiques basées sur une activation des checkpoints par des agents cytostatiques ou cytotoxiques, pour contrecarrer la prolifération incontrôlée des cellules cancéreuses [57, 103]. Du côté de la recherche fondamentale, plusieurs études expérimentales montrent un lien bilatéral entre le cycle cellulaire (et indissociablement ses checkpoints) d'une part et l'horloge circadienne [65, 75, 102] ou le métabolisme [31] d'autre part. En voici un résultat marquant. La protéine suppresseur de tumeur p53 est un acteur déterminant dans l'activation du checkpoint G1/S en cas de dommages à l'ADN. Il contrôle la prolifération cellulaire en amont de la phase S en cas de dommage à l'ADN et la disruption de ce checkpoint (et supposément des autres) est associée à une prolifération incontrôlée et une instabilité génomique, deux caractéristiques du cancer [57]. p53 s'avère sous le contrôle de l'horloge circadienne ce qui soulève la question de l'impact d'une disruption de l'horloge circadienne (*e.g.* le travail en horaires décalés) sur le cancer. En effet, le gène de la protéine p53 est muté dans plus de 50% des tumeurs humaines. De plus, il est désormais admis que la disruption du rythme circadien joue un rôle clé dans la genèse des tumeurs cancéreuses et, inversement, le cancer perturbe directement les rythmes circadiens [94].

Le rôle prédominant des checkpoints à la fois dans l'émergence et les traitements des cancers nécessite de se saisir d'urgence de la question de la formalisation des checkpoints. Une quantité déjà importante et grandissante de connaissances biologiques s'accumulent sous le terme de checkpoint du cycle cellulaire. Une approche basée sur de la modélisation mathématique est indispensable pour faire face à la complexité engendrée par le nombre de molécules et de voies de régulations qui explose. À titre indicatif pour *Homo sapiens*, au moins 1244 gènes sont annotés avec le terme « cell cycle checkpoint » au moment de l'écriture de cette thèse [105] et au minimum 274 protéines sont impliquées dans les checkpoints du cycle cellulaire [108]. Ceci sans compter les liens avec les autres systèmes physiologiques. Nous avons recensé cinq checkpoints du cycle cellulaire : les checkpoints G1/S, S/G2, G2/M, le SAC et le checkpoint de restriction. Ils forment les checkpoints intrinsèques du cycle cellulaire et les quatre premiers peuvent être activés par les checkpoints extrinsèques en réponse à un stress génotoxique. Les checkpoints G1/S, S/G2, G2/M et le SAC partagent une propriété que l'on généralise par la proposition suivante : *un checkpoint intrinsèque empêche tout évènement qui initie une phase donnée de se réaliser tant que tous les évènements de la phase précédente ne sont pas correctement terminés*. Il s'agit de la propriété minimale que doivent générer les checkpoints intrinsèques du cycle cellulaire pour assurer, en l'occurrence, une séparation temporelle des phases S et M.

Seule une modélisation mathématique est à même de prédire si un ensemble de voies de régulations, supposées impliquées dans les points de contrôle du cycle cellulaire, peut générer une dynamique qui reproduit au moins cette propriété fondamentale des checkpoints. De par son approche rigoureuse, un formalisme de modélisation adapté à la question de recherche peut relever des contradictions parmi les connaissances biologiques formalisées. Les contradictions d'un modèle donné forment le matériel de base à la prédiction de nouvelles expériences qui serviront à élucider de nouveaux mécanismes impliqués dans la question des checkpoints. En outre, l'enjeu de la modélisation des checkpoints ne se situe pas tant dans la formalisation des voies de régulations impliquées que dans la formalisation des phases et des checkpoints du cycle cellulaire. En pratique, les études de modélisation du cycle cellulaire concentrent leurs efforts dans la conception des voies de régulations et dans la simulation ou vérification de leurs propriétés dynamiques en vue de reproduire les comportements asymptotiques attendus. Typiquement on retrouve la présence d'un attracteur cyclique caractéristique du cycle cellulaire sous perfusion constante en facteurs de croissance, l'arrêt en phase de quiescence en absence de ces facteurs ou encore l'observation d'une trajectoire où les complexes cyc/cdk s'activent et s'inactivent dans un ordre attendu [28, 42, 74]. Pour autant à notre connaissance, aucune étude ne s'est encore consacrée à *la généralisation des contraintes dynamiques que génèrent les checkpoints le long du cycle cellulaire*. Une des raisons principales est qu'aucune étude ne s'est attelée à *la représentation mathématique d'une phase du cycle cellulaire* qui est un objet indissociable des checkpoints.

1.4 Vue d’ensemble de la thèse

Les **chapitres 2 et 4** définissent le matériel théorique sur lequel se fonde la thèse. Le chapitre 2 est dédié à un exposé des différents formalismes de modélisation utilisés pour représenter la dynamique des réseaux de régulation biologique. Le chapitre 4 est consacré à la définition de notre approche formelle dans le cadre de l’informatique théorique et qui est appliquée à la spécification et vérification formelle des checkpoints du cycle cellulaire.

Le **chapitre 3** est consacré à un état de l’art de la modélisation mathématique des checkpoints du cycle cellulaire. Huit modèles ont été sélectionnés pour témoigner de la façon dont la communauté scientifique se saisit de la question des checkpoints. Même si les études pionnières s’appuient historiquement sur des formalismes continus, les études les plus récentes tendent vers une discrétisation de la problématique. La revue de [40], mise à jour récemment [82], fût fondatrice pour cette thèse. Les auteurs concluent qu’un checkpoint est un phénomène intrinsèquement discret qui repose structurellement sur un couplage de *switchs bistables*. Un switch bistable est une boucle de rétroaction positive qui génère des états stables distincts et participe au cloisonnement de certains événements du cycle cellulaire (*e.g* ceux de la réplication) par rapport à d’autres (*e.g* ceux de la mitose). Ce phénomène de cloisonnement permet une *séparation temporelle* des phases S et M. L’étude de [86] montre, au travers d’un exemple, que si l’on peut admettre un switch bistable caractéristique d’une phase alors un couplage en série de plusieurs switchs bistables est capable de générer un engagement irréversible d’une phase vers la suivante. Nous appellerons ce phénomène *l’irréversibilité des transitions de phases*. À la lumière de l’état de l’art proposé dans ce chapitre, deux enjeux se dégagent nettement. Le premier est de définir correctement une phase du cycle cellulaire, définition qui sera nécessairement incluse dans celle des deux propriétés dynamiques d’un checkpoint. Le second enjeu est d’ordre méthodologique. Comment prouver qu’une certaine mécanistique des checkpoints, *e.g* un couplage de switchs bistables impliquant les régulateurs du cycle cellulaire, génère une certaine phénoménologie des checkpoints, *e.g* les phénomènes de séparation temporelle des phases et d’irréversibilité des transitions de phases ?

Le **chapitre 5** est dédié à une formalisation des connaissances statiques et dynamiques du cycle cellulaire. Elle est effectuée à l’aide du formalisme de modélisation de René Thomas. Les régulations considérées comme essentielles à la question biologique sont formalisées par un graphe d’influence inspiré par ceux de [40] et [69]. Un tel graphe couvre *a priori* un grand ensemble de modèles différents car les forces des activations et des inhibitions sont inconnues. On utilise une famille de paramètres dynamiques qui encode ces forces [8] et permet de s’affranchir d’un choix arbitraire d’une fonction d’activation par variable, donc d’un choix arbitraire de modèle. Dans les années 70, les modèles qualitatifs basés sur les mathématiques discrètes se sont avérés utiles pour comprendre les principales causalités qui régissent les phénotypes observés, et le cadre multivalué de René Thomas est devenu un classique pour les réseaux de régulation biologique. Il a gagné en puissance avec l’introduction des méthodes formelles au début des années 2000 [20]. Ce chapitre prouve que les propriétés dynamiques classiques du cycle cellulaire qui sont observées expérimentalement (*e.g* le phénomène de quiescence ou les oscillations des complexes cyc/cdk) peuvent être formalisées à l’aide de deux logiques temporelles incontournables en informatique théorique et adaptée au formalisme de René Thomas : la logique temporelle arborescente [20] et la logique de Hoare génétiquement modifiée [85]. La plate-forme logicielle TotemBioNet est choisie pour faire la jonction entre les connaissances statiques (*i.e.* le graphe d’influence) et les phéno-

mènes dynamiques classiques du cycle cellulaire. `TotemBioNet` identifie les paramétrisations qui satisfont toutes les propriétés dynamiques formalisées [96]. L'ensemble de modèles résultant encode les données expérimentales auxquelles notre définition de phase se confronte au chapitre 6.

Le **chapitre 6** se consacre à la généralisation du concept de phase indépendamment de la définition d'un modèle du cycle cellulaire. Trois définitions génériques sont proposées et varient selon leur degré de liberté. La définition d'une *phase canonique* est l'étape où sont fixées les hypothèses. Une phase est un multi-ensemble d'évènements de régulation biologique, assimilables à des évènements discrets, qui sont requis pour qu'une phase du cycle cellulaire ait lieu. La définition de l'*hyperpavé d'une phase canonique* représente le domaine de la phase au sein duquel aucune connaissance arbitraire sur l'ordre des évènements n'est présupposée. Une *phase* est finalement définie comme le résultat de la confrontation de l'hyperpavé d'une phase canonique aux données expérimentales. Cette confrontation repose sur l'application de la logique de Hoare génétiquement modifiée.

Le **chapitre 7** est le chapitre central où les deux propriétés fondamentales des checkpoints sont formalisées. Le prédicat de *séparation temporelle* est introduit et assure que tout évènement qui peut clore une phase n'est permutable avec aucun évènement qui peut initier la phase suivante. Une contrainte est générée pour sélectionner les modèles du cycle cellulaire qui satisfont cette propriété. La vérification de l'*irréversibilité des transitions de phases* est encodée directement par une formule CTL qui utilise la formule caractéristique d'une phase pour vérifier qu'aucune visite des autres phases n'est autorisée à partir de tout autre état que l'état final d'une phase. Cette conceptualisation des deux propriétés des checkpoints est enfin validée à l'aide de l'ensemble de modèles inférés au chapitre 5 : la preuve de concept est un succès s'il existe au moins une paramétrisation qui satisfait les deux propriétés intrinsèques des checkpoints.

État de l'art

CHAPITRE 2

Les formalismes pour modéliser la régulation de comportements cycliques

| | | |
|------------|---|-----------|
| 2.1 | Un préambule épistémologique : pourquoi formaliser ? | 17 |
| 2.2 | Les systèmes équations différentielles ordinaires | 17 |
| 2.3 | Les réseaux booléens aux réseaux multivalués | 22 |
| 2.4 | Approche multivaluée avec multiplexes : le formalisme de modélisation de René Thomas | 24 |
| 2.4.1 | Syntaxe abstraite : graphe d'influence avec multiplexes | 24 |
| 2.4.1.1 | De l'origine des seuils discrets de régulations biologiques | 25 |
| 2.4.1.2 | De l'utilité des multiplexes | 27 |
| 2.4.2 | De la syntaxe à la sémantique : la notion de paramètre dynamique | 29 |
| 2.4.2.1 | Définition de la famille des paramètres dynamiques à partir d'un GRBM | 29 |
| 2.4.2.2 | Signification de la valeur des paramètres dynamiques | 29 |
| 2.4.2.3 | La condition de Snoussi pour contraindre la valeurs des paramètres dynamiques | 30 |
| 2.4.2.4 | Une expérience de pensée pour contraindre la valeurs des paramètres dynamiques | 31 |
| 2.4.3 | Sémantique : le graphe de transition asynchrone | 35 |

2.1 Un préambule épistémologique : pourquoi formaliser ?

Une question majeure que se pose la communauté des chercheurs en biologie moléculaire est la suivante : quels sont les mécanismes de régulations biologiques qui *induisent* un comportement biologique donné ? La biologie expérimentale tente de répondre à cette question en suivant une approche *abductive* qui permet d'extraire un ensemble d'hypothèses sur les causalités du comportement biologique étudié, à partir de corrélations *inattendues*. L'approche abductive est une forme de raisonnement qui ne suit pas de cadre rigoureux et expose particulièrement les chercheurs aux erreurs d'interprétation des résultats expérimentaux.

Le vivant est un système complexe et multi-échelle qui rend la tâche d'investigation des mécanismes du vivant quasi impossible sans poser de cadres théoriques généraux à partir des observations particulières, en suivant une approche *inductive*. Les cadres formels de modélisation des réseaux de régulations biologiques en sont un exemple typique. Ils permettent de formaliser des concepts biologiques sur lesquels appliquer des approches *déductives* réalisant des preuves du type : *étant donné un réseau de régulation biologique, certains comportements biologiques sont nécessairement admis*. Cette étape de formalisation des connaissances et une approche déductive sont également utilisées dans le cas où les expériences biologiques ne sont pas en mesure de mettre directement à l'épreuve du vivant les hypothèses sur les causalités du comportement biologique étudié.

In fine, la formalisation des connaissances biologiques est une étape essentielle pour s'assurer de la correction d'un raisonnement et un choix *raisonné* du formalisme de modélisation est tout aussi essentiel. Tout formalisme doit être adapté à la question biologique posée pour être en mesure d'y répondre. Il existe deux courants majeurs de modélisation. Une approche quantitative qui représente de manière continue la concentration d'une espèce biochimique par des nombres réels et une approche qualitative qui abstrait sa concentration par un entier positif ou nul qui représente l'action de l'espèce biochimique sur une ou plusieurs de ses cibles. Nous présenterons le formalisme quantitatif le plus couramment utilisé pour modéliser les réseaux de régulations biologiques, à savoir les systèmes d'équations différentielles en section 2.2. De même, nous nous focaliserons sur les réseaux booléens en section 2.3 pour l'approche qualitative. Nous situerons chacun de ces formalismes présentés vis à vis des questions biologiques qu'ils peuvent traiter et des hypothèses qu'ils admettent. Ensuite, nous positionnerons en section 2.4 le formalisme multivalué de modélisation des réseaux biologiques introduit par René Thomas et Houssine Snoussi [4, 9, 35].

2.2 Les systèmes équations différentielles ordinaires

Les systèmes d'équations différentielles ordinaires constituent un formalisme de modélisation classique de la dynamique d'un graphe de régulation biologique, en des termes purement quantitatifs. De nombreux modèles du cycle cellulaire utilisent ce formalisme c'est pourquoi nous lui dédions une section à part entière. Une équation différentielle ordinaire (abrégée EDO) permet typiquement de représenter la variation au cours du temps de la concentration d'une entité biologique notée x_i , i allant de 1 à n (e.g. *ARNm*, protéine, métabolite, une espèce biologique). Une EDO relie la dérivée $\frac{dx_i}{dt}$ à une fonction f_i qui décrit les mécanismes biologiques qui influencent x_i , fonction quasiment toujours *non linéaire*. On représente en général une EDO par espèce biologique participant au réseau de régulation biologique. Formellement pour un système d'EDO de

dimension n [18] :

$$\frac{dx_i}{dt} = f_i(x), \quad 1 \leq i \leq n$$

où x est un vecteur de concentrations positives ou nulles (x_1, \dots, x_n) . La fonction $f_i : \mathbb{R}_+^n \rightarrow \mathbb{R}_+$ définit les mécanismes de régulations de x_i :

$$f_i(x) = r(\bar{x}_i) - \delta_i x_i$$

où \bar{x}_i est le vecteur de concentrations des prédécesseurs de x_i dans le graphe de régulation, δ_i est le taux de dégradation de x_i et $r : \mathbb{R}^m \rightarrow \mathbb{R}$ est le plus souvent une somme de fonctions de régulation traduisant chacune l'effet de l'un des m régulateurs de x_i . Cette somme suppose que les x_j agissent indépendamment les uns des autres et les fonctions de régulation classiquement utilisées sont les *fonctions de Hill* (définition 1).

Une fonction de Hill est une approximation d'une courbe sigmoïde caractéristique d'un phénomène observé en biochimie : la *coopérativité positive* (la fixation d'une molécule régulatrice x sur une molécule cible y améliore l'affinité de x pour y). Les fonctions de régulations h^+ et h^- de la définition 1 possèdent un *paramètre de coopérativité* noté n qui confère à ces fonctions une forme sigmoïdale lorsque $n > 1$, illustrée en figure 2.1.

Définition 1: Fonctions de Hill

Étant donné un *activateur* x et sa cible y , la fonction de Hill h^+ qui décrit le processus d'activation de y par x (noté $x \xrightarrow{+} y$) est définie par :

$$h_{\theta,n}^+(x) = \frac{x^n}{x^n + \theta^n}$$

Étant donné un *inhibiteur* x et sa cible y , la fonction de Hill h^- qui décrit le processus d'inhibition de y par x (noté $x \xrightarrow{-} y$) est définie par :

$$h_{\theta,n}^-(x) = \frac{\theta^n}{x^n + \theta^n}$$

n est appelé le paramètre de *coopérativité* et θ le paramètre de *seuil*.

Plus le paramètre de coopérativité est grand plus la pente de la sigmoïde est élevée. Plus le nombre de molécules x augmente plus l'effet sur y augmente jusqu'à atteindre une valeur limite. Cette dernière traduit le phénomène de saturation de la molécule y qui ne possède qu'un nombre fini de récepteurs pour x . L'effet sur y du régulateur x dans le cas d'une activation (resp. inhibition) noté $h_{\theta,n}^+$ (resp. $h_{\theta,n}^-$) converge vers 1 (resp. 0) lorsque x tend vers $+\infty$. Les fonctions de régulations h^+ et h^- possèdent également un *paramètre de seuil* noté θ qui définit, lorsque n tend vers $+\infty$, une valeur de x à partir de laquelle l'effet de la régulation de y devient maximal.

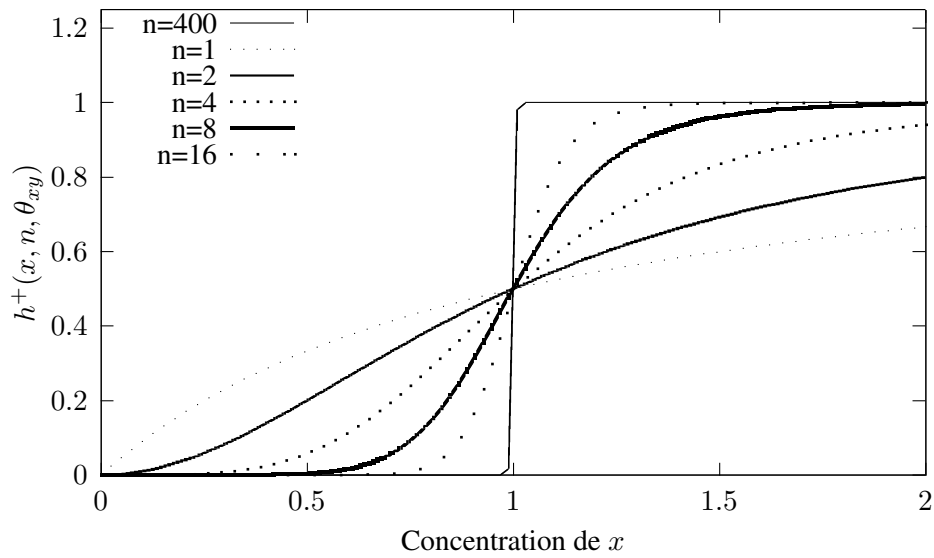
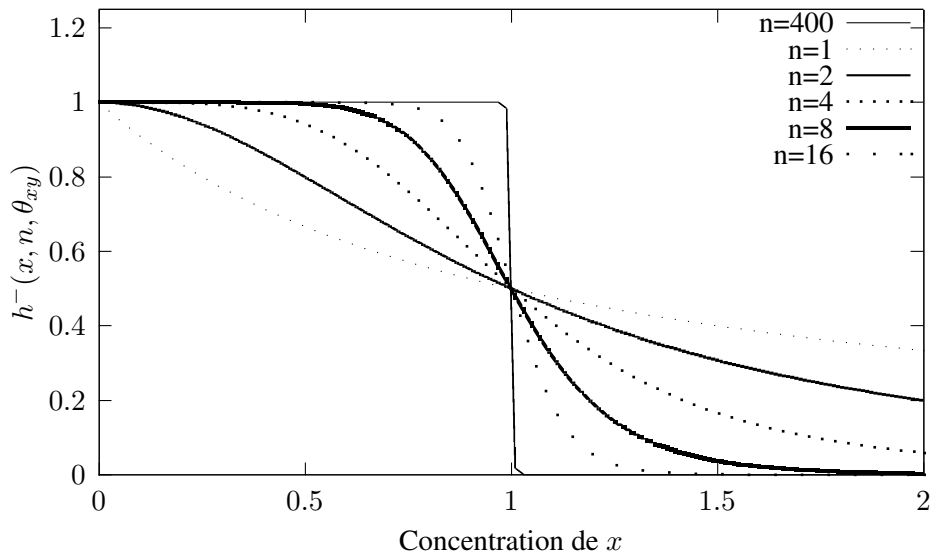
(a) La fonction de régulation positive $h^+(x, \theta_{xy}, n)$ (b) La fonction de régulation négative $h^-(x, \theta_{xy}, n)$

Figure 2.1 – **Les fonctions de Hill h^+ et h^- .** Elles associent la concentration d'un régulateur x à l'effet de sa régulation sur y de paramètre $\theta_{xy} = 1$ et pour $n = 1, 2, 4, 8, 16$ et 400 . θ est le point d'inflexion. Lorsque $n = 1$, la fonction h^+ est égale à l'expression donnée par Michaelis-Menten où θ_{xy} représente une constante d'affinité.

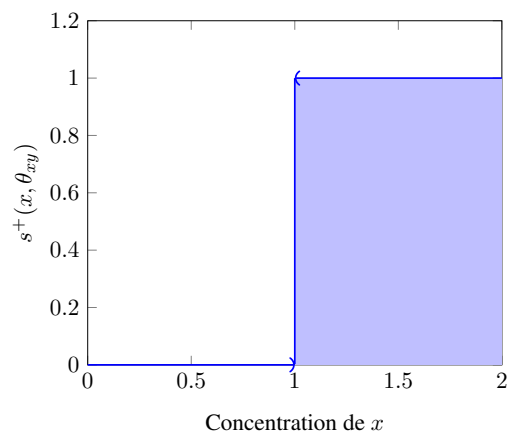
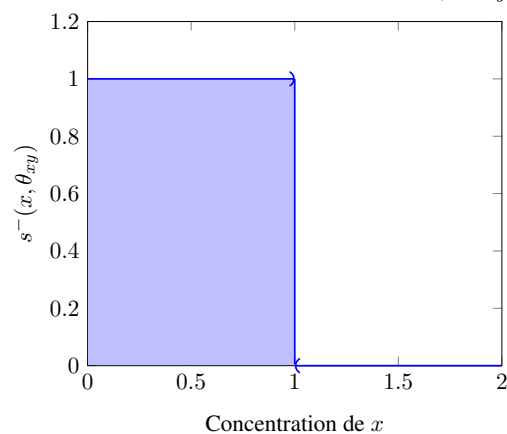
(a) La fonction de régulation positive $s^+(x, \theta_{xy})$ (b) La fonction de régulation négative $s^-(x, \theta_{xy})$

Figure 2.2 – **Les fonctions de step de Heavyside s^+ et s^- de paramètre $\theta_{xy} = 1$** Elles décrivent logiquement l'effet de la régulation de x sur y . La valeur $y = 1$ représente sa valeur maximale normalisée.

La fonction step de Heavyside, donnée en définition 2, est une approximation discrète de la fonction de Hill lorsque n tend vers $+\infty$ et fournit une information qualitative : x est inactif en dessous de la valeur θ_{xy} ce qui se traduit par aucun effet sur y , sinon x agit au maximum sur y à partir de la valeur θ_{xy} . Une illustration de la fonction step de Heavyside est donnée en figure 2.2. Cette vision qualitative fait d'ailleurs écho à l'intuition première du formalisme qualitatif de René Thomas (*cf.* section 2.4), qui est utilisé dans les travaux de modélisation de cette thèse.

Définition 2: Fonctions step de Heavyside

Étant donné un *activateur* x et sa cible y , la fonction step s^+ est définie par :

$$s^+(x, \theta) = \begin{cases} 0 & \text{si } x < \theta_{xy} \\ 1 & \text{si } x \geq \theta_{xy} \end{cases}$$

Étant donné un *inhibiteur* x et sa cible y , la fonction step s^- est définie par :

$$s^-(x, \theta) = \begin{cases} 1 & \text{si } x < \theta_{xy} \\ 0 & \text{si } x \geq \theta_{xy} \end{cases}$$

θ_{xy} est le paramètre de *seuil* de la fonction step.

En général, les systèmes d'EDO qui abstraient les mécanismes de régulations biologiques ne peuvent pas être résolus analytiquement, étant donnée la non-linéarité de la fonction r . Nonobstant, des méthodes de résolutions numériques permettent d'obtenir des solutions approchées, en partant d'une condition initiale.

Prenons l'exemple d'un oscillateur moléculaire basique à deux variables x et y . Ces deux variables possèdent chacune un taux de synthèse basal (respectivement σ_x et σ_y) et un taux de dégradation (respectivement δ_x et δ_y). Considérons une rétroaction négative : x active y et y inhibe x . Le système d'EDO à deux dimensions qui formalise ce réseau de régulation, à l'aide d'une fonction de Hill, est le suivant :

$$\begin{cases} \frac{dx}{dt} = \sigma_x \times \frac{x^{n_x}}{x^{n_x} + \theta_{xy}} - \delta_x x \\ \frac{dy}{dt} = \sigma_y \times \frac{\theta_{yx}}{\theta_{yx} + x^{n_x}} - \delta_y y \end{cases}$$

En plus de la solution analytique, les paramètres *cinétiques* de ce système d'EDO sont inconnus, à savoir les paramètres σ_i , δ_i , les paramètres de seuil θ_{xy} , θ_{yx} et les paramètres de coopérativité n_x et n_y . L'estimation de ces paramètres se heurte à des limitations.

Nous citerons d'abord le besoin d'une grande quantité de données expérimentales pour aborder le problème d'identifiabilité des paramètres. Généralement, obtenir ces données est un processus coûteux expérimentalement en temps et financement. Un paramètre associé à une variable v est dit structurellement non-identifiable si la variation de sa valeur n'a aucun effet sur la trajectoire de v . La valeur estimée d'un paramètre non identifiable à partir de données expérimentales n'aurait donc aucune interprétation biologique. Le choix des variables à expliciter au sein du système d'EDO finit donc toujours par limiter la taille des modèles. Cette limitation est parfaitement illustrée par Jon Von Newmann : « *With four parameters I can fit an elephant, and with five I can make him wiggle his trunk* ».

L'hypothèse du continu admet que les espèces biochimiques se trouvent en quantité infinie et implique de générer une grande quantité de données expérimentales sur l'activité des espèces biochimiques d'intérêt. De plus, cette hypothèse ne permet pas d'exploiter des connaissances sur

l'hétérogénéité des populations de cellules que l'analyse des données de séquençage sur cellule unique peut apporter. Les modèles qualitatifs (booléens ou multivalués) constituent une abstraction simplificatrice des modèles continus mais permettent de prouver la satisfiabilité de comportements asymptotiques caractéristiques des aspects fondamentaux du vivant tels que les cycles limites (*e.g.* processus homéostatiques) et la multistationnarité (phénomènes épigénétiques ou de différenciation cellulaire). Le pouvoir d'abstraction du discret ouvre les portes de méthodes formelles de l'informatique théorique à la fois pour formaliser et vérifier des propriétés biologiques purement qualitatives (*e.g.* les checkpoints du cycle cellulaire) qu'il est fondamental d'étudier avant d'affiner la problématique sur des aspects plus quantitatifs, *e.g.* les effets dose-réponse d'une substance chimique sur le vivant ou encore la synchronisation de la chronométrie des régulations biologiques.

2.3 Les réseaux booléens aux réseaux multivalués

Introduits dans la fin des années 60 par Kauffman [3], les réseaux booléens sont aujourd'hui classiquement utilisés pour étudier les propriétés dynamiques qualitatives d'un réseau d'influences biologiques. Nous mentionnions plus haut les comportements asymptotiques omniprésents dans le vivant tels que les attracteurs cycliques et la multistationnarité. Chaque entité biochimique modélisée est représentée par une variable booléenne v . Lorsqu'elle est en concentration suffisante (resp. insuffisante) pour agir sur ses cibles, elle est caractérisée par $v = 1$ (resp. $v = 0$). L'espace des états d'un réseau booléen à n variables est donc défini par $\{0, 1\}^n$ et chaque variable v se voit munie d'une fonction logique d'activation f_v pour construire les transitions entre ces états. Cette fonction caractérise la nature des coopérations entre les variables pour activer v . Elle s'exprime dans une logique propositionnelle restreinte à la conjonction, disjonction et négation. Considérons les fonctions logiques d'activation suivantes, associées à chacune des deux variables booléennes x et y :

$$\begin{aligned} f_x &: \neg y \\ f_y &: x \wedge y \end{aligned}$$

La fonction f_x exprime que l'activation de x a lieu lorsque y est inactif et f_y formalise l'activation de y par le complexe formé par x et y . On peut également abstraire le graphe d'influence suivant où les noeuds représentent les variables et les arcs entre les noeuds les influences positives ou négatives qu'exercent une variable sur la synthèse d'une autre. L'étiquette $+$ et $-$ représentent respectivement une activation et une inhibition :

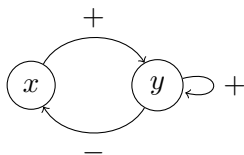


Figure 2.3 – Le graphe d'influence qui abstrait le réseau booléen défini par f_x et f_y

Cette représentation permet de ne pas considérer un modèle *a priori*, qui est défini par un graphe de transitions à partir de l'ensemble des fonctions d'activation f_v . Deux interprétations des réseaux booléens se confrontent classiquement. Une modélisation qui adopte une stratégie asynchrone suppose que chaque variable, qui doit s'activer dans un état, se met à jour indépendamment des

autres. La stratégie synchrone consiste quant à elle à mettre à jour en même temps toutes les variables qui doivent s'activer. Des stratégies intermédiaires ont été proposées depuis. Par exemple, [28] propose d'assigner des classes de priorités aux variables booléennes ce qui permet d'obtenir un système de transitions mélangeant transitions synchrones et asynchrones. Plus récemment, une sémantique plus permissive a été introduite par [100] qui prend en compte des états transitoires de variables en cours d'activation ou d'inhibition interprétable au choix par 0 ou 1. Il est d'ailleurs à noter que le choix de la sémantique d'un réseau booléen a un impact sur les propriétés d'atteignabilité et de stabilité qu'il peut admettre. Le choix d'une sémantique est donc une hypothèse de modélisation qu'il convient de discuter dans toute conclusion d'une étude de modélisation à base de réseaux booléens. Un réseau booléen synchrone à la *Kauffman* ne prend pas en compte le fait que l'ordre total de mise à jour des variables est quasiment inconnu en biologie. En biologie, il est invraisemblable que deux variables passent toujours leurs seuils d'activation (resp. inactivation) exactement au même moment. Une sémantique asynchrone à la *Thomas* prend en compte tous les ordres de mises à jour possibles donc tous les chemins d'exécution possible à partir d'un état. Cette sémantique est considérée plus réaliste d'un point de vue biologique en revanche il en découle, non pas une sous-, mais une sur-approximation des états atteignables par rapport aux les connaissances biologiques.

Lorsqu'une variable active deux cibles, il est peu probable que les deux cibles soit pour autant activées exactement en même temps. Un moyen d'y remédier est de considérer un réseau multivalué. Ils permettent d'exprimer plus finement l'activité d'une variable. Une variable est munie d'une borne $b_v \in \mathbb{N}^*$. Il est donc possible de définir dans quel ordre une variable v va activer ou inhiber ses cibles, en la munissant de plusieurs seuils distincts de régulation. C'est ce que propose l'approche multivaluée de René Thomas présentée dans la section 2.4, qui s'avère être la plus appropriée pour représenter explicitement les différents niveaux qualitatifs des complexes cyclines/cdk en fonction d'états de phosphorylation bien distincts.

L'analyse des graphes de transitions (synchrones ou asynchrones) permet d'identifier une composante fortement connexe terminale (CFCT) caractéristique d'un attracteur cyclique (e.g. le cycle cellulaire), ou de plusieurs de ces composantes caractéristiques d'une multistationnarité. La limitation principale réside dans l'analyse de la pertinence biologique de ces CFCT. Par exemple, le cycle cellulaire est caractérisé par une CFCT d'un réseau booléen, par [87] et [28]. En revanche, l'identification des chemins admissibles biologiquement au sein de cette composante est une tâche fastidieuse sachant que le nombre d'états augmente exponentiellement en fonction du nombre de variables. Cette tâche est d'autant plus fastidieuse que les connaissances biologiques sur le comportement d'un système sont souvent partielles.

En outre, lorsque les règles logiques de mise à jour des variables ne sont pas connues, un problème d'apprentissage logique se pose : *quels sont les modèles qui satisfont un ensemble de propriétés donné ?* Typiquement, la méthode *Answer Set Programming* permet de caractériser un ensemble exhaustif de modèles logiques associés à un réseau de signalisation [60] ou d'identifier un ensemble de modèles booléens solutions de propriétés biologiques formalisées, par exemple [68, 97]. Houssine Snoussi introduit quant à lui la notion de ressource et une famille des paramètres logiques notés $K_{v,\omega}$ qui associe à chaque variable toutes les configurations possibles des ressources de v à partir d'un graphe d'influence. Une ressource notée ω est ici abstraite comme la présence d'un activateur et l'absence d'un inhibiteur. Par exemple, le graphe d'influence 2.3 admet que la variable y peut être *potentiellement* activée par la présence de x ou l'absence de y . La variable

y possèdent 4 paramètres qui gouvernent sa dynamique : $K_{y,\emptyset}$, $K_{y,\{x\}}$, $K_{y,\{y\}}$, $K_{y,\{x,y\}}$, et x en possède deux : $K_{x,\emptyset}$ et $K_{x,\bar{y}}$. Dans le cadre booléen, $K_{v,\omega}$ est également une variable booléenne. L'ensemble de ces paramètres symboliques associé à une variable v permet de ne présupposer aucune règle de mise à jour f_v . Lorsque chaque paramètre est instancié, la table de vérité des paramètres $K_{v,\omega}$ donne l'expression de la fonction f_v . Par exemple la table de vérité suivante ci-dessous permet de déduire $f_y = x \wedge y$:

| x | y | ω | $K_{y,\omega}$ |
|-----|-----|-------------|----------------|
| 0 | 0 | \emptyset | 0 |
| 1 | 0 | x | 0 |
| 0 | 1 | y | 0 |
| 1 | 1 | $\{x, y\}$ | 1 |

Table 2.1 – **Inférence d'une règle d'activation d'un variable à partir de la table de vérité de ses paramètres** $K_{v,\omega}$. Ce jeu de paramètres n'est autre que la table de vérité de la conjonction $x \wedge y$.

y est activée uniquement lorsque ses deux ressources sont activées. [20] ont introduit un apprentissage symbolique de modèles à base de model-checking de formules CTL à partir d'un ensemble exhaustif de modèles déduit à partir d'un graphe d'influence et de la famille des paramètres $K_{v,\omega}$. L'objectif de l'apprentissage est d'identifier les modèles (c'est à dire les fonctions f_v dans le cadre booléen) qui satisfont des propriétés temporelles exprimées en CTL (*e.g.* atteignabilité et stabilité), sans présupposer arbitrairement des valeurs de paramètres lorsque les connaissances biologiques manquent.

2.4 Approche multivaluée avec multiplexes : le formalisme de modélisation de René Thomas

Ce formalisme définit mathématiquement un ensemble de régulations biologiques sous la forme d'un *graphe de régulation biologique avec multiplexe*, introduit en section 2.4.1. Sa dynamique est définie en section 2.4.3 par un *graphe de transition asynchrone* pour lequel le comportement de chaque variable, soumise à toutes les configurations possibles de régulations, est défini par une famille de *paramètres dynamiques* introduit en section 2.4.2. Cette section propose deux niveaux de description : d'une part un niveau intuitif qui illustre les besoins de formalisation en biologie et d'autre part un niveau purement formel, proposé dans les encadrés de définition.

2.4.1 Syntaxe abstraite : graphe d'influence avec multiplexes

La syntaxe abstraite du formalisme permet de spécifier des régulations entre des variables biologiques sous la forme d'un graphe de régulation biologique avec multiplexe (définition 3. Les variables biologiques représentent des molécules (ou des groupes de molécules) du vivant telles que typiquement des gènes, des transcrits, des protéines, des complexes moléculaires de natures variées, ou des variables abiotiques telles que la température, le pH ou encore le niveau d'oxygène. L'ensemble des interactions entre les différentes variables est représenté mathématiquement par un graphe orienté où les sommets sont soit des variables (représentées graphiquement par des

ronds) soit des multiplexes (représentées graphiquement par des rectangles), comme en figure 2.5. Ces derniers formalisent des processus de régulations biologiques plus ou moins complexes à l'aide d'un sous ensemble de la logique du premier ordre où chaque variable, notée v , représente une entité biologique interprétée par un niveau de concentration discret.

L'introduction des multiplexes [35] dans la syntaxe du formalisme de René Thomas enrichit une première approche multivaluée [9] qui introduit la notion fondamentale de *seuil discret* de régulation. Ces seuils sont en nombre fini et nous n'utilisons pas de quantificateurs, par conséquent le nombre d'atomes est fini et la logique des multiplexes est propositionnelle. Nous proposons d'abord d'expliquer d'où viennent ces seuils qui permettent de définir les briques de base de régulations biologiques (littéraux). Enfin nous montrerons la puissance d'expression des multiplexes pour décrire un large panel de régulations complexes par des combinaisons de ces littéraux, en suivant la logique propositionnelle.

Définition 3: Graphe de régulation biologique avec multiplexe

Un *graphe de régulation biologique avec multiplexe* (GRBM) est un triplet $\mathcal{G} = (V, M, E)$ où :

- V est un ensemble fini de variables v munies d'une borne $b_v \in \mathbb{N}^*$
- M est un ensemble fini de symboles m munis d'une formule φ_m sur le langage \mathcal{L} défini inductivement par :
 - pour toute variable $v \in V$ et pour tout entier n tel que $n \in \llbracket 0, b_v \rrbracket$, $(v \geq n) \in \mathcal{L}$
 - si $(\varphi, \psi) \in \mathcal{L}^2$ alors $\neg\varphi \in \mathcal{L}$, $\varphi \wedge \psi \in \mathcal{L}$ et $\varphi \vee \psi \in \mathcal{L}$
- E est l'ensemble des arcs où $E \subseteq M \times V$.
On note $(m \rightarrow v)$ un élément de E et $E^{-1}(v)$ est l'ensemble des prédécesseurs de v .

2.4.1.1 De l'origine des seuils discrets de régulations biologiques

En 1973, René Thomas propose une approche purement booléenne [4] parallèlement à Kauffman. L'**idée fondatrice** est la même :

- $x = 0$ représente un *intervalle* de concentrations de x trop faibles pour produire un effet sur y , à ne pas confondre avec une concentration de x nulle.
- $x = 1$ représente un *intervalle* de concentrations de x pour lesquelles on observe un effet de la régulation sur y .

La régulation de la transcription d'un gène par un facteur de transcription est un exemple typique en biologie. Dans ce cas, la variable x décrit la concentration du facteur de transcription et y la concentration du produit du processus de transcription du gène y , à savoir son ARN messager.

L'approche booléenne de René Thomas se base sur une discrétisation de l'axe de concentration continu de chaque variable x sur lequel est placé le seuil θ_{xy} . Étant donné un système à n variables, l'espace de concentration à n dimensions est découpé 2^n hyperpavés au sein desquels chaque variable agit uniformément sur ses cibles. En pratique, il est très rare de trouver une estimation des seuils de concentration θ_{xy} mesurés expérimentalement. Généralement, les résultats d'ex-

périences permettent plutôt d'extraire des informations d'une nature qualitative, comme illustré par René Thomas lui-même : « *Note that biologists known to be allergic to any formal treatment nevertheless invariably say (and write) : "in the absence of immunity, this, in the presence of immunity, that; at low temperature, this, at high temperature, that ...". This is a typically Boolean attitude.* » [9].

Cette approche purement booléenne a le défaut de limiter grandement le pouvoir d'expression des régulations entre plusieurs variables biologiques, parce qu'une variable peut agir sur plusieurs cibles à des seuils qu'il faut distinguer ou parce qu'une variable peut agir différemment sur une même cible à des concentrations différentes. Le cycle cellulaire possède un exemple typique : le cas du complexe cyclineE/Cdk2 (abrégé *cycE/cdk2* et introduit dans le chapitre précédent). Ce complexe biologique possède deux seuils d'actions distincts sur ses cibles, à l'origine de trois valeurs de concentration : $cycE/cdk2 = 1$ correspond à un état du complexe *monophosphorylé*, $cycE/cdk2 = 2$ à un état *hyperphosphorylé*, et enfin le complexe n'agit significativement sur aucune cible lorsque $cycE/cdk2 = 0$. Dans le premier cas, le complexe *cycE/cdk2* est capable d'activer uniquement sa propre phosphorylation, alors que dans le deuxième il est capable, *de surcroît*, d'activer la transcription de la cycline A.

Houssine Snoussi et René Thomas proposèrent donc en 1991 un formalisme **multivalué** [9] de modélisation des réseaux de régulations biologiques. Si une variable possède n cibles alors elle est associée à un domaine discret fini de borne maximale $b_v = n$, dans le pire des cas. Il n'est pas à exclure que v puisse réguler *au même seuil* plusieurs cibles et dans ce cas b_v est strictement inférieur à n . Lorsqu'aucune information biologique ne permet de discriminer les différents seuils d'actions d'une variable, il est tentant de choisir le nombre maximal de seuil mais il est souvent préférable opter pour l'hypothèse la plus simple d'après le principe du Rasoir d'Ockham : choisir le nombre minimal de seuils à défaut de donnée expérimentale.

Étant donné un ensemble de variables V , la discrétisation de l'espace de concentration du système continu résulte en un découpage en $\prod_{v \in V} (b_v + 1)$ hyperpavés dans lesquels les régulations sont constantes. Un exemple est proposé en figure 2.4. Considérons un système à deux dimensions où la variable x possède deux seuils de régulations notés respectivement θ_x^1 et θ_x^2 (variable multivaluée) et la variable y n'en possède qu'un seul noté θ_y^1 (variable booléenne). La variable x peut donc prendre que trois valeurs discrètes :

$$\begin{cases} x = 0 & \text{si } x < \theta_x^1 \\ x = 1 & \text{si } \theta_x^1 \leq x < \theta_x^2 \\ x = 2 & \text{si } x \geq \theta_x^2 \end{cases}$$

La variable y se résume au cas booléen abordé en début de section :

$$\begin{cases} y = 0 & \text{si } y < \theta_y^1 \\ y = 1 & \text{si } y \geq \theta_y^1 \end{cases}$$

Au sein de l'espace continu de concentration d'un système biologique, l'endroit précis où se trouve chaque seuil de régulation de chaque variable est *in fine* une préoccupation mineure dans le formalisme de René Thomas. Ce qui est primordial est uniquement *l'ordre dans lequel chaque variable passe ses seuils de régulation*, si l'on s'intéresse uniquement au comportement qualitatif du système.

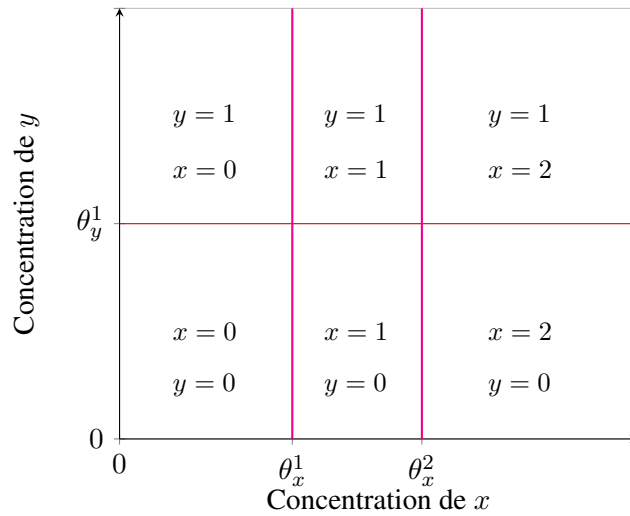
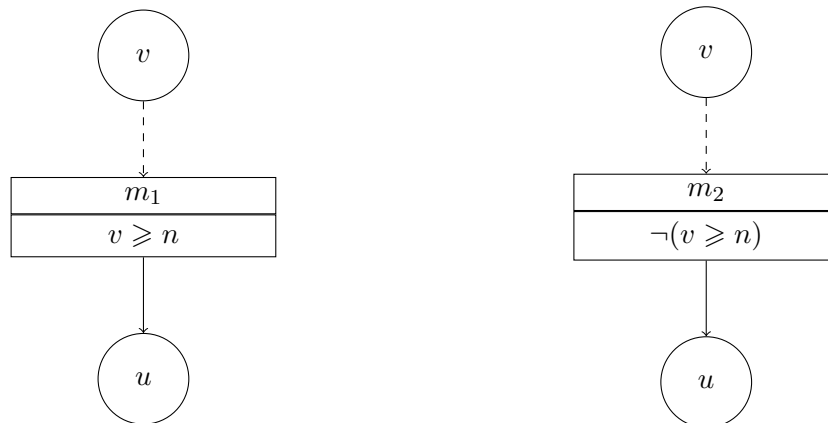


Figure 2.4 – **Exemple de discrétisation d'un espace de concentration à 2 dimensions à partir de l'unique seuil de régulation de y (noté θ_y^1) et des deux seuils de x (notés θ_x^1 et θ_x^2).** L'espace de concentration continu est découpé en 6 états où l'action de chaque variable sur ses cibles est uniforme et où chaque état abstrait une région de l'espace de concentration continue.

2.4.1.2 De l'utilité des multiplexes

Un multiplexe est défini par un couple formé par un identificateur noté m et une formule en logique propositionnelle notée φ_m . Lorsque φ_m est réduite à un atome (sans négation), elle définit l'action positive qu'exerce une variable sur une ou plusieurs cibles.



(a) Le multiplexe m_1 représente la présence de l'activateur v : il atteint ou dépasse son seuil n .

(b) Le multiplexe m_2 représente l'absence de l'inhibiteur v : il n'atteint pas le seuil n .

Figure 2.5 – **Les littéraux du langage des multiplexes.** L'arc $m_1 \rightarrow u$ (resp. $m_2 \rightarrow u$) représente l'activation (resp. l'inhibition) de u par v . Formellement, les arcs de v vers m_1 (resp. vers m_2) sont superflus car v apparaît explicitement dans la formule du multiplexe m_1 (resp. m_2).

Dans un GRBM, un arc de la forme $m \rightarrow u$ tel que $\varphi_m = (v \geq n)$ représente l'activation de u par v au seuil n , cf. figure 2.5a. Pour décrire une inhibition, il suffit d'utiliser la négation de la formule atomique, à savoir $\neg(v \geq n)$, cf. figure 2.5b. Un multiplexe dont la formule est un atome (resp. la négation d'un atome) décrit donc la présence d'un activateur (resp. l'absence d'un inhibiteur), à un seuil donné. La puissance d'expression des multiplexes vient de leur capacité à combiner plusieurs littéraux (atomes ou leurs négations) en utilisant la conjonction et la disjonction de la logique propositionnelle, pour représenter des phénomènes de coopération/concurrence. En voici quelques exemples :

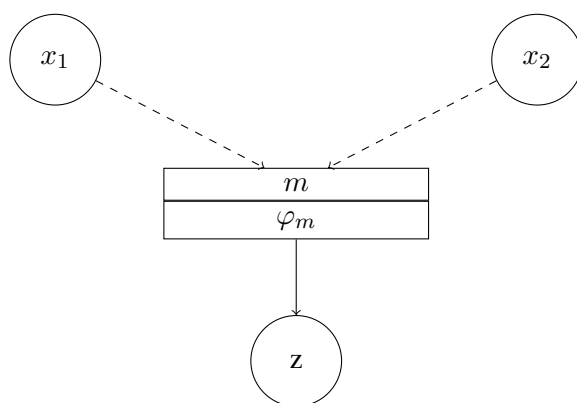


Figure 2.6 – **Combinaison de l'action deux régulateurs x_1 et x_2 sur une cible commune z .** Le multiplexe m de formule φ_m formalise la nature des phénomènes de coopération/concurrence entre les x_i .

Les complexations : La formation d'un complexe est classiquement décrite au niveau syntaxique par une *conjonction* de plusieurs littéraux. Considérons par exemple dans la figure 2.6 que x_1 et x_2 sont deux activateurs de z mais uniquement sous la forme d'un complexe x_1x_2 . Le complexe formé est formalisé par $\varphi_m \equiv (x_1 \geq 1) \wedge (x_2 \geq 1)$. Pour décrire l'action d'un complexe inhibiteur, il suffit de définir une conjonction de littéraux négatifs.

Les phénomènes d'inhibition de régulation : Une conjonction permet également de décrire l'inhibition de l'action d'un régulateur, par un autre régulateur. Cette inhibition est en général formalisée, dans le cas de deux régulations, par la conjonction d'un littéral et d'un littéral négatif. Considérons cette fois dans la figure 2.6 (avec $n = 2$) que x_2 inhibe l'activation de z par x_1 . Cette inhibition est formalisée par $\varphi_m \equiv (x_1 \geq 1) \wedge \neg(x_2 \geq 1)$.

Les regroupements de régulations de même nature : Une *disjonction inclusive* de plusieurs littéraux permet de regrouper l'action de plusieurs régulateurs lorsqu'ils agissent *de manière indépendante* sur une cible commune. Si x_1 et x_2 agissent indépendamment sur z , on écrira $\varphi_m \equiv (x_1 \geq 1) \vee (x_2 \geq 1)$.

Les phénomènes de compétition : Une *disjonction exclusive* permet de formaliser des phénomènes de compétition. Si x_1 et x_2 entrent en compétition pour activer z , on écrira $\varphi_m \equiv (x_1 \geq 1 \vee x_2 \geq 1) \wedge \neg(x_1 \geq 1 \wedge x_2 \geq 1)$. Cette formule signifie que x_1 et x_2 ne

peuvent agir positivement sur z que s'ils ne sont pas actifs en même temps. Un exemple typique est lorsque x_1 et x_2 sont deux gènes qui produisent deux protéines X_1 et X_2 qui ont une forte affinité et que le complexe X_1 - X_2 ne peut plus agir sur z .

Les multiplexes permettent d'abstraire un large éventail de mécanismes de régulations génétiques tels que typiquement la transcription, la traduction, la (dé)phosphorylation, la dégradation protéolytique. Nous renvoyons le lecteur au chapitre 5 pour des exemples plus sophistiqués, avec notre graphe de régulation avec multiplexe de la progression du cycle cellulaire des mammifères.

Dans les quatre exemples précédents, la variable z a un unique prédécesseur alors qu'une représentation classique lui donnerait deux prédécesseurs. Une conséquence directe est la réduction du nombre de dynamiques que peut générer un GRBM, cf. section 2.4.2. Plus généralement, lorsqu'une variable v possède n prédécesseurs, elle est *potentiellement* soumise à l'influence de 2^n configurations possibles de régulations. Le niveau syntaxique ne détermine pas si un multiplexe a un effet sur la dynamique de la variable v , c'est à dire sur la valeur vers laquelle v est attirée.

2.4.2 De la syntaxe à la sémantique : la notion de paramètre dynamique

2.4.2.1 Définition de la famille des paramètres dynamiques à partir d'un GRBM

Le GRBM fournit une description syntaxique du système où chaque variable v est soumise à l'influence d'un ensemble de multiplexes prédécesseurs. La notion de *paramètres dynamiques* est le point de jonction entre une description syntaxique et une description dynamique : les paramètres sont dénotés $K_{v,\omega}$ où ω est un ensemble de *multiplexes* prédécesseurs de v . L'ensemble des paramètres associé à une variable peut être extrait directement à partir du GRBM en exhibant toutes les combinaisons possibles des multiplexes prédécesseurs de v . Chaque variable v possède 2^n paramètres, où n est le degré entrant de v dans un GRBM.

Définition 4: Famille des paramètres dynamiques

Étant donné un GRBM $\mathcal{G} = (V, M, E)$, on note \mathcal{K} la famille des *paramètres dynamiques* constituée des symboles de la forme $K_{v,\omega}$ où v parcourt V et ω parcourt l'ensemble des parties de $E^{-1}(v)$.

$K_{v,\omega}$ est appelé le paramètre de v pour l'ensemble de ressources ω .

2.4.2.2 Signification de la valeur des paramètres dynamiques

La dynamique d'un GRBM est définie par des transitions entre états, un état étant défini par la valeur courante de chacune des variables du graphe :

Définition 5: État

Étant donné un GRBM \mathcal{G} , un *état* est une substitution $\eta : V \rightarrow \mathbb{N}$ telle que pour tout $v \in V$, $\eta(v) \in \llbracket 0, b_v \rrbracket$. On note ζ l'ensemble des états.

Dans chaque état noté η , toute variable v est substituée par une valeur appartenant à son domaine discret : on peut donc identifier pour toute variable v l'unique ensemble des régulations auxquelles

elle est *effectivement* soumise. Tout multiplexe prédécesseur de v tel que sa formule caractéristique φ_m est satisfaite dans η est appelé *ressource* de v dans η . Ainsi, une variable ne possède qu'un paramètre applicable par état, *cf.* définition 6.

Définition 6: Paramètre applicable

Étant donné un GRBM \mathcal{G} et un état η , un paramètre $K_{v,\omega} \in \mathcal{K}$ est dit *applicable* en η si ω est l'ensemble des $m \in E^{-1}(v)$ tels que φ_m est satisfaite pour la substitution η .

La valeur entière du paramètre $K_{v,\omega}$ définit la valeur vers laquelle est attirée une variable v lorsque ω est l'ensemble de ses ressources. Le codomaine de $K_{v,\omega}$ est donc fini et réduit au domaine discret de v . La *dynamique locale* en η est déterminée par une instanciation du paramètre applicable $K_{v,\omega}$ pour tout v . On parle de réseau de régulation biologique avec multiplexe lorsque chaque paramètre $K_{v,\omega}$ est substitué par une valeur appartenant à son domaine discret.

Définition 7: Réseau de régulation biologique avec multiplexe

Un *réseau de régulation biologique avec multiplexe* (RRBM) est un couple $\mathcal{R} = (\mathcal{G}, \sigma)$ où \mathcal{G} est un GRBM et $\sigma : \mathcal{K} \rightarrow \mathbb{N}$ est une substitution telle que pour tout $K_{v,\omega} \in \mathcal{K}$, $\sigma(K_{v,\omega}) \in \llbracket 0, b_v \rrbracket$. σ est appelé un paramétrage du GRBM.

Dans le cadre de la thèse, on appelle *un modèle* une occurrence de couple (\mathcal{G}, σ) .

La substitution σ permettra de formaliser la *dynamique globale* du système par un *graphe de transitions* où l'ensemble des sommets représente l'ensemble des états discrets du système et où les transitions entre états sont définies à partir des valeurs $\sigma(K_{v,\omega})$, *cf.* section 2.4.3.

L'étape critique lors de la modélisation de réseaux de régulations biologiques est *l'identification des paramètres* $K_{v,\omega}$, c'est à dire trouver la ou les substitutions σ en accord avec les connaissances biologiques (si elles existent). L'introduction de la famille des paramètres \mathcal{K} permet d'utiliser des méthodes formelles classiques de l'informatique théorique, telles que le *model-checking* ou la *logique de Hoare*, pour réaliser un apprentissage symbolique des valeurs des paramètres $K_{v,\omega}$ à partir des connaissances biologiques, *cf.* chapitre 4 pour une introduction à ces méthodes formelles. Le nombre de paramétrages augmente exponentiellement en fonction de la taille du GRBM. Plus précisément, ce nombre est égal à $\prod_{v \in V} (b_v + 1)^{2^{d^-(v)}}$, où b_v est la borne maximale de v et $d^-(v)$ son degré entrant. Ainsi une utilisation *raisonnable* de ces outils théoriques repose sur une réduction *au maximum* du domaine des paramètres $K_{v,\omega}$.

2.4.2.3 La condition de Snoussi pour contraindre la valeurs des paramètres dynamiques

Plusieurs méthodes existent pour contraindre le domaine des paramètres $K_{v,\omega}$, préalablement à l'utilisation du model-checking et de la logique de Hoare. Dans un premier temps, Houssine Snoussi a introduit en 1989 la condition de monotonicité suivante [8] : toute ressource ajoutée à l'ensemble des ressources d'une variable ne peut pas faire diminuer son point focal. Dit en d'autres termes : une absence de ressource doit être une absence, une présence de ressource doit

être une présence mais ne doit pas devenir une absence.

Définition 8: Condition de Snoussi

Un RRBM $\mathcal{R} = (\mathcal{G}, \sigma)$ satisfait la *condition de Snoussi* pour une variable v si :

$$\forall \omega \in E^{-1}(v), \forall \omega' \in E^{-1}(v), \omega \subseteq \omega' \Rightarrow \sigma(K_{v,\omega'}) \geq \sigma(K_{v,\omega})$$

Toute variable n'est pas nécessairement concernée par la condition de Snoussi. Il existe des cas où l'ajout d'une ressource à une variable v diminue son point focal car une interaction non prévue entre plusieurs de ses ressources, au stade de la conception du GRBM, la rend inactive.

2.4.2.4 Une expérience de pensée pour contraindre la valeurs des paramètres dynamiques

D'autre part, une méthode empirique de contrainte du domaine des paramètres $K_{v,\omega}$ a émergé de plusieurs décennies de travail de modélisation de l'équipe *Bio-info Formelle* de l'IS3, sur des questions biologiques variées. Il s'agit d'une pure **expérience de pensée** dont le scénario est le suivant :

Étant donné une variable v qui reçoit *en entrée* une combinaison de ressources ω supposées constantes *ad vitam eternam*, on se demande quelle(s) cible(s) v peut réguler *en sortie*. Cela conduit au moins à prendre en compte les deux cas suivants :

- **Lorsque v est inactive** ($v = 0$). Ce cas permet de savoir si les ressources ω sont suffisantes pour que v agisse sur une ou plusieurs cibles.
- **Lorsque v est saturée** ($v = b_v$). Ce cas permet de savoir si v peut maintenir sa valeur maximale.

La condition de Snoussi peut de plus aider la démarche. Ce scénario est réitéré pour tout ω .

On note que se demander vers quelle valeur converge v , à ω fixé, revient à se demander quels seuils v est capable de passer et donc sur quelle(s) cible(s) peut agir v étant donné ω .

Cette expérience nécessite une expertise *biochimique* qui ne peut pas être directement fournie par une expérience biologique. La biologie expérimentale est nécessairement systémique. Il est donc impossible de fixer et d'isoler l'action d'un régulateur sur une variable, ne serait-ce que parce que les résultats des expériences prennent en compte l'ensemble des processus biologiques de la cellule. Il n'est donc pas possible de rendre complètement inactif un régulateur et de fixer ω comme contexte de l'expérience. Pour répondre à l'expérience de pensée, un expert est mis à contribution sur les aspects biochimiques et un modélisateur mène l'expérience en veillant à ce que l'expert ne raisonne pas de manière systémique.

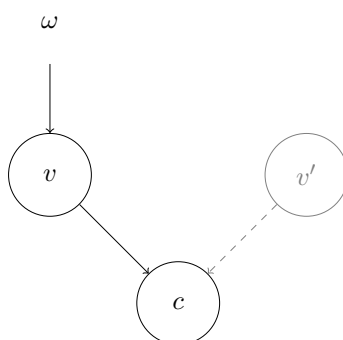


Figure 2.7 – **Un exemple de raisonnement systémique au cours de l'expérience de pensée.** Ici l'erreur consiste à considérer qu'une cible c peut être effectivement régulée par un régulateur v étant donné ses ressources ω mais en incluant implicitement une autre variable v' qui régule également c .

Considérons à titre d'exemple le GRBM en figure 2.8 qui représente une boucle positive et un cycle négatif sur x . La variable x est multivaluée et peut prendre trois valeurs (0, 1 ou 2), y est une variable booléenne. Ce GRBM génère exactement 324 modèles possibles.

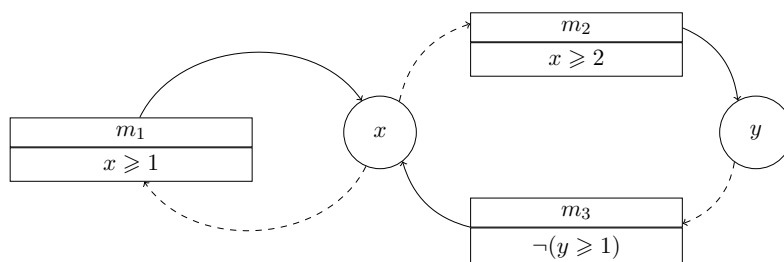


Figure 2.8 – **Un exemple de GRBM pour illustrer l'expérience de pensée** Il formalise une auto-activation de x au seuil 1, une activation de y par x au seuil 2 ainsi qu'une inhibition de x par y au seuil 1.

Admettons que l'enjeu de l'expérience soit de savoir quel effet exerce chacun de ces cycles sur le comportement qualitatif de x . Premièrement, le GRBM permet d'identifier 6 paramètres dynamiques. Quatre paramètres gouvernent la dynamique de x : $K_{x,\{m_1,m_3\}}$, $K_{x,\{m_1\}}$, $K_{x,\{m_3\}}$ et $K_{x,\emptyset}$. Deux paramètres gouvernent la dynamique de y : $K_{y,\{m_2\}}$ et $K_{y,\emptyset}$. Recentrons désormais l'expérience de pensée sur la variable x et considérons toutes ses combinaisons d'influences possibles sous la forme de 4 scénarii illustrés en figure 2.9.

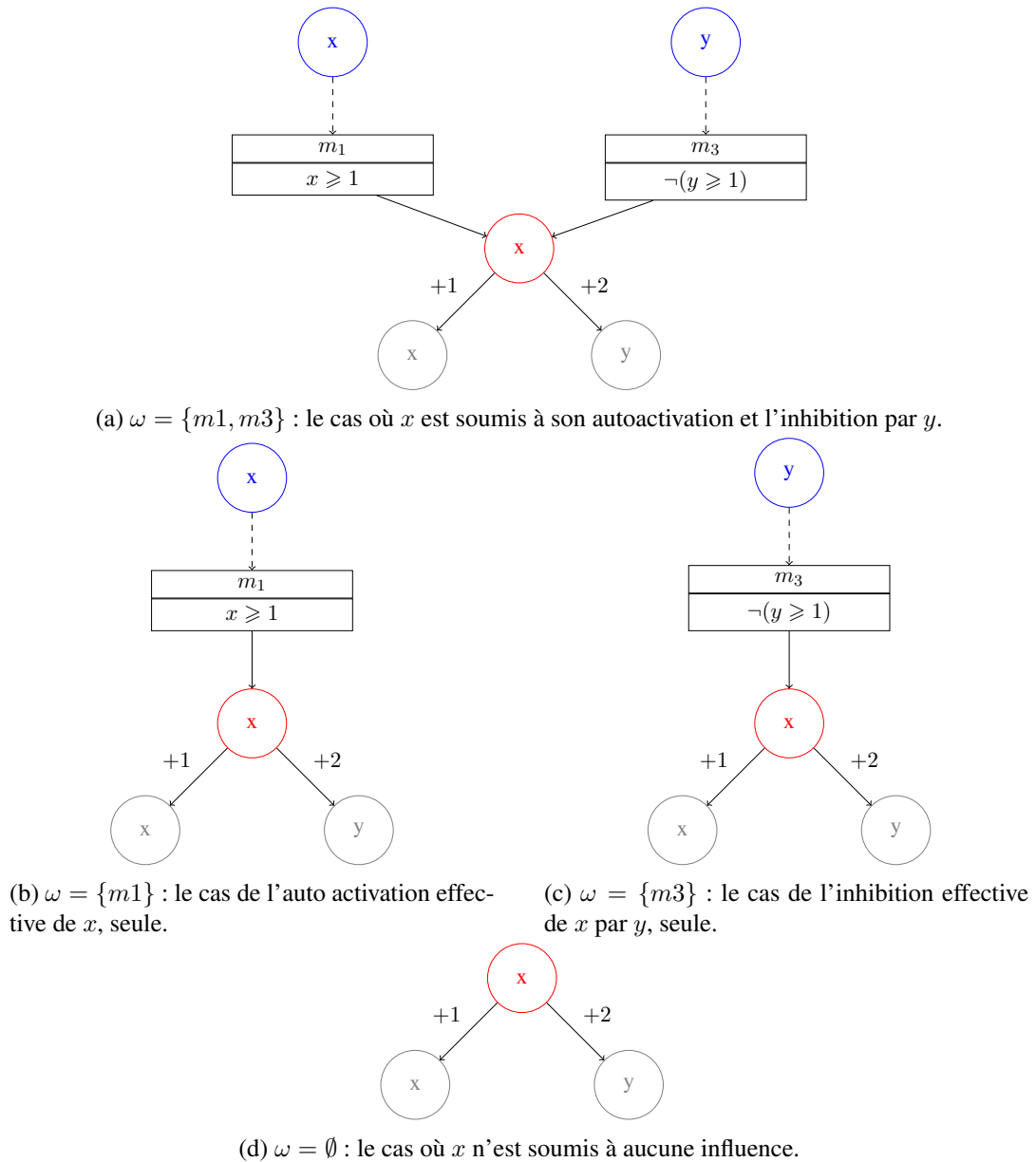


Figure 2.9 – **Illustration du cadre de l'expérience de pensée pour contraindre le domaine des paramètres de Thomas $K_{x,\omega}$ pour tout ω de x .** La variable x est placée au centre en rouge, ses régulateurs sont placés au dessus en bleu et ses cibles en dessous en gris. Les étiquettes des arcs prédécesseurs des cibles de x représentent soit une activation (signe +) soit une inhibition (signe -) à un seuil discret de x donné. Dans les cas a et b , la variable x est représentée trois fois : en bleu en tant que *ressource*, en rouge en tant que *cible de la ressource* et en gris comme *cible de la régulation de x* . La subtilité de l'expérience réside dans le fait que x est considéré actif *ad vitam eternam* uniquement dans le rôle de ressource.

$\mathbf{K}_{x,\{m_1,m_3\}}$: Ce paramètre décrit le contexte où x s’auto-active (au seuil 1) et *n’est pas* inhibé par y (au seuil 1), cf. figure 2.9a. La question posée ici est de savoir vers quelle valeur est attiré x en présence de toutes ses ressources en considérant une condition initiale où x est actif ($x = 0$). *Idem* en considérant une condition initiale où x est saturé ($x = 2$). Admettons que notre expert biologiste nous indique dans le premier cas que x possède assez de ressources pour atteindre au moins son premier seuil de régulation. Le modélisateur formalise cette information par la contrainte $K_{x,\{m_1,m_3\}} \geq 1$. Admettons dans le second cas que les ressources de x permettent de maintenir x dans son état de saturation. Cette information est formalisée par la contrainte $K_{x,\{m_1,m_3\}} = 2$. Dans le cadre de cet exemple, nous choisissons la plus forte contrainte c’est à dire $K_{x,\{m_1,m_3\}} \geq 2$.

$\mathbf{K}_{x,\{m_1\}}$: Ce paramètre décrit le contexte où à la fois l’activateur x et l’inhibiteur y sont présents, ce qui implique que ω n’inclut pas le multiplexe ressource m_3 , cf. Figure 2.9b. Admettons qu’une procédure similaire au scénario précédent nous donne la contrainte $K_{x,\{m_1\}} = 1$.

$\mathbf{K}_{x,\{m_3\}}$: Ce paramètre décrit le contexte où à la fois l’activateur x et l’inhibiteur y sont absents, ce qui implique que ω n’inclut pas le multiplexe ressource m_1 , cf. Figure 2.9c. Ce contexte est appelé *le cas basal* de la régulation de x puisqu’aucune régulation inhibitrice ou activatrice n’est effective. Admettons qu’une procédure similaire au premier scénario nous donne la contrainte $K_{x,\{m_3\}} = 1$.

$\mathbf{K}_{x,\emptyset}$: Enfin ce paramètre décrit le contexte où aucun multiplexe n’est ressource, autrement dit : aucun activateur de x n’est présent et aucun inhibiteur de x n’est absent, cf. Figure 2.9d. Admettons qu’une procédure similaire au scénario précédent nous donne la contrainte $K_{x,\emptyset} = 0$.

Finalement, la dynamique de x est déterminée par l’ensemble des paramètres fixés $K_{x,\{m_1,m_3\}} = 2$, $K_{x,\{m_1\}} = 1$, $K_{x,\{m_3\}} = 1$ et $K_{x,\emptyset} = 0$ et permet de conclure que :

- La coopération des ressources m_1 et m_3 est suffisante pour l’activation de x à son second seuil de régulation.
- La ressource m_3 est nécessaire à x pour lui permettre d’atteindre son premier seuil de régulation mais pas suffisante pour passer son second seuil, ce qui empêche l’inhibition de y par x .
- Il en est de même pour la ressource m_1 .

D’après les valeurs de paramètres, on déduit que l’activation de x au seuil 1 est caractérisée par la fonction logique $\varphi_{m_1} \vee \varphi_{m_3}$, à distinguer de l’activation de x au seuil 2 qui elle est caractérisée par $\varphi_{m_1} \wedge \varphi_{m_3}$. La table de vérité fournie par les paramètres $K_{v,\omega}$ permet donc de connaître la fonction de mise à jour d’une variable, à seuils distincts dans le cas multivalué.

In fine, puisque tous les paramètres dynamiques de x ont été instanciés et qu’aucun des paramètres de y ne l’a été alors il reste 4 modèles possibles. Cette expérience de pensée permet d’accumuler des informations sur la *dynamique locale* du système étudié en contraignant directement à la main le domaine des paramètres $K_{v,\omega}$. On note $\phi_{K_{v,\omega}}$ la conjonction de toutes ces contraintes biologiques. $\phi_{K_{v,\omega}}$ ne dépend ni de v ni de ω , c’est la conjonction de tous.

2.4.3 Sémantique : le graphe de transition asynchrone

Comme mentionné dans la section précédente, le graphe de transition asynchrone est un objet mathématique *déduit d'un RRBM* et représente sa dynamique globale. Les sommets sont les états du système et les arcs représentent les transitions entre états construites à partir d'un paramétrage σ des $K_{v,\omega}$.

Définition 9: Graphe de transition asynchrone

Étant donné un RRBM $\mathcal{R} = (\mathcal{G}, \sigma)$, son graphe de transition asynchrone (GTA) est défini de la façon suivante :

- l'ensemble des sommets est l'ensemble des états possibles ζ de \mathcal{R} ,
- il existe une transition $\eta \rightarrow \eta'$ s'il existe une variable $v \in V$ telle que $\sigma(K_{v,\omega}) \neq \eta(v)$, où $K_{v,\omega}$ est le paramètre applicable de v en η , et :
 1. $\eta'(v) = \eta(v) + 1$ si $\sigma(K_{v,\omega}) > \eta(v)$
 2. $\eta'(v) = \eta(v) - 1$ si $\sigma(K_{v,\omega}) < \eta(v)$
 3. pour tout $v' \in V$, $v' \neq v \Rightarrow \eta'(v') = \eta(v')$

La sémantique choisie dans le cadre du formalisme de René Thomas possède deux caractéristiques : elle est **asynchrone** (condition 3 de la définition 9) et **unitaire** (conditions 1 et 2 de la définition 9).

La première caractérise la désynchronisation de l'évolution des variables : lorsqu'une variable évolue, aucune autre variable n'évolue simultanément. L'idée intuitive est que la probabilité que deux variables atteignent leur seuil de régulation *simultanément* est nulle, notamment étant donné la non exactitude des instruments de mesure de concentrations continues. La discrétisation de l'espace de concentration ne consiste donc pas uniquement en un découpage en hyperpavés (au sein desquels l'ensemble des ressources est constant), s'y ajoute la désynchronisation d'une dynamique continue.

La seconde caractéristique impose que tout arc d'un GTA ne peut relier que deux sommets voisins qui ne diffèrent que d'une seule composante. Ainsi lorsqu'une variable v est attirée par une valeur $\eta'(v)$ qui n'est pas celle de son état courant $\eta(v)$ et qui diffère de plus d'une unité, par exemple $\eta'(v) = \eta(v) + 2$, v doit nécessairement passer par l'état $\eta''(v) = \eta(v) + 1$ dans lequel elle est potentiellement soumise à l'influence d'autres ressources qui ne permettent pas forcément d'atteindre l'état $\eta'(v)$.

Considérons à titre d'exemple le graphe d'interaction de la figure 2.8, noté \mathcal{G}_1 et les deux paramétrages suivant :

- $\sigma_1 = \{K_{x,\emptyset} = 0, K_{x,m1} = 0, K_{x,m3} = 1, K_{x,m1,m3} = 2, K_{y,\emptyset} = 0, K_{y,m2} = 1\}$
- $\sigma_2 = \{K_{x,\emptyset} = 0, K_{x,m1} = 0, K_{x,m3} = 1, K_{x,m1,m3} = 2, K_{y,\emptyset} = 0, K_{y,m2} = 0\}$.

Chacun des paramétrages prend en considération les résultats de l'expérience de pensée sur les paramètres dynamiques de x réalisée en section 2.4.2 à l'exception de $K_{\{x,m1\}}$ qui prend la valeur 0 pour illustrer l'asynchronisme de la dynamique dans l'état caractérisé par $x = 1 \wedge y = 1$. Toutes choses égales par ailleurs, les deux paramétrages diffèrent d'une seule valeur de paramètre : $K_{y,m2}$.

Les modèles $(\mathcal{G}_1, \sigma_1)$ et $(\mathcal{G}_1, \sigma_2)$ permettent de construire deux graphes de transition asynchrone d'ordre 6, représentés en figure 2.10, qui produisent deux comportements qualitativement bien distincts. Le modèle $(\mathcal{G}_1, \sigma_1)$ génère une *oscillation* autour des seuils $x = 2$ et $y = 1$ alors que le modèle $(\mathcal{G}_1, \sigma_2)$ génère un *état stable* caractérisé par $x = 2 \wedge y = 0$.

On remarque que le paramétrage σ_2 du graphe de régulation de la figure 2.8 met en évidence que l'ajout de la ressource m_2 (présence de l'activateur x au seuil 2) à n'importe quel ensemble de ressources de y ne produit aucun effet sur y . On dit dans ce cas que l'arc $m_2 \rightarrow y$ est *non fonctionnel*.

Définition 10: Arc non fonctionnel

Étant donné un RRBM $\mathcal{R} = (\mathcal{G}, \sigma)$, un arc de \mathcal{G} de la forme $m \rightarrow v$ est *non fonctionnel* si : $\forall \omega \in E^{-1}(v) \setminus m, \sigma(K_{v, \{\omega, m\}}) = \sigma(K_{v, \omega})$

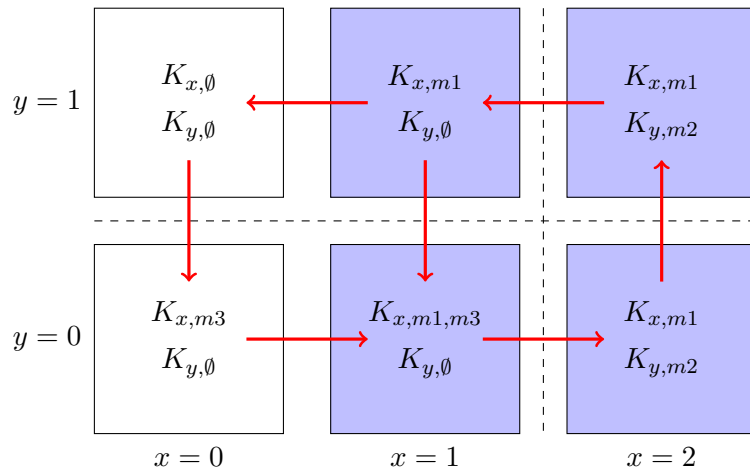
Des propriétés dynamiques du système biologique peuvent être observées directement à partir d'un graphe de transition asynchrone généré par un modèle (\mathcal{G}, σ) , telles que classiquement l'atteignabilité d'un ensemble d'états, l'existence d'un état ou cycle stable, ou encore la présence d'un de plusieurs bassins d'attraction.

Dans le formalisme multivalué de René Thomas, deux *propriétés générales* sur les RRBM peuvent être déduites directement à partir de la dynamique d'un RRBM :

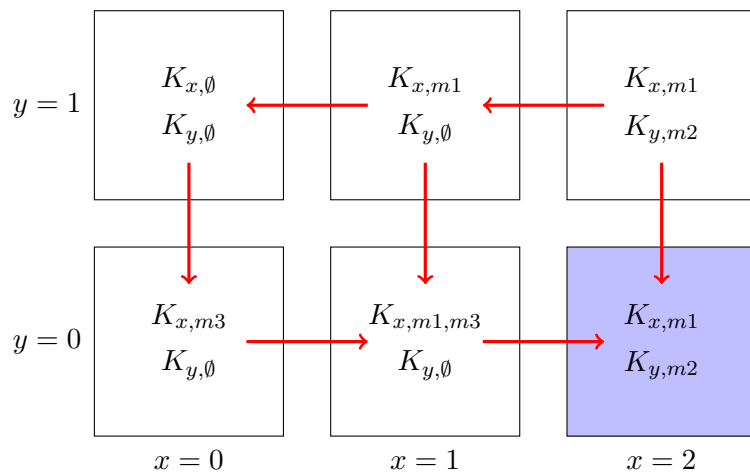
- À propos de la présence de plusieurs états stables dans un GTA : une condition nécessaire à cette observation est la présence d'au moins un cycle positif dans le RRBM. Un cycle positif comporte un nombre *pair* d'inhibitions (dont tous les arcs inhibiteurs et activateurs sont fonctionnels). Cette propriété est d'autant plus d'intérêt qu'elle s'applique à des phénomènes de différenciations cellulaires (*e.g.* hématopoïèse, développement embryonnaire, morphogénèse végétale) ou épigénétiques (*e.g.* le cas d'étude de *pseudomonas aeruginosa* [20]).
- La seconde propriété énonce que la présence d'un cycle négatif dans un RRBM est une condition nécessaire à la présence d'une oscillation dans un GTA. Un cycle négatif comporte un nombre *impair* d'inhibitions (de nouveau dont tous les arcs inhibiteurs et activateurs sont fonctionnels). Les processus oscillants sont omniprésents en biologie, caractéristiques de l'homéostasie. Le cycle cellulaire en est un exemple lui-même.

Ces deux conjectures ont été prouvées, respectivement par [33] et [49].

In fine, le formalisme de René Thomas représente la dimension temporelle de manière *chronologique* : seule des successions d'états sont représentées. Toute information chronométrique ne peut donc pas être exploitée dans ce formalisme et toutes les questions sur des délais de régulations sont déléguées à des formalismes dédiés : soit purement quantitatifs comme abordés en section 2.2 soit hybrides tels que l'extension hybride de l'approche multivaluée de René Thomas elle-même.



(a) Graphe de transition asynchrone engendré à partir du modèle $(\mathcal{G}_1, \sigma_1)$. On observe une oscillation de part et d'autre des seuils $x = 2$ et $y = 1$ représentés conceptuellement par les droites en pointillés.



(b) Graphe de transition asynchrone engendré à partir du modèle $(\mathcal{G}_1, \sigma_2)$. L'oscillation issue du paramétrage σ_1 n'existe pas dans ce modèle puisque l'arc activateur de x vers y dans le graphe de régulation de la figure 2.8 est non fonctionnel. Ce modèle admet à la place un état stable caractérisé par la formule $x = 2 \wedge y = 0$.

Figure 2.10 – Deux graphes de transition asynchrone engendrés à partir du graphe d'interaction de la figure 2.8 et de deux paramétrages. Les états sont représentés par des carrés et les transitions par des flèches orientées de l'état source vers l'état cible. Le système modélisé étant de dimension deux, chaque état ne possède que deux paramètres : un qui s'applique à x et un à y .

CHAPITRE 3

Une revue des modèles incontournables de la régulation du cycle cellulaire

| | | |
|------------|--|-----------|
| 3.1 | Préambule | 41 |
| 3.2 | Les modèles continus pionniers | 43 |
| 3.2.1 | Identification d'une structure de régulation : les switches bistables | 43 |
| 3.2.2 | Deux propriétés intrinsèquement discrètes du phénomène de checkpoint | 45 |
| 3.3 | Les modèles discrets et la formalisation du phénomène de checkpoint | 47 |
| 3.3.1 | Un exemple de modélisation discrète des régulations impliquées dans certains checkpoints | 47 |
| 3.3.2 | Un exemple de formalisation de certaines contraintes chronologiques au cours du cycle cellulaire | 48 |
| 3.3.3 | Une première tentative de formalisation des phases du cycle cellulaire | 50 |
| 3.4 | La formalisation des phases munies de checkpoints : la question reste ouverte | 52 |
| 3.4.1 | Relier le mécanisme des checkpoints à ses deux propriétés dynamiques | 52 |
| 3.4.2 | Les checkpoints intrinsèques et extrinsèques | 53 |
| 3.5 | Bilan sur la modélisation des checkpoints | 54 |

3.1 Préambule

« *Essentiellement, tous les modèles sont faux, mais certains sont utiles.* » - Georges Box.

Le cycle cellulaire est un processus bien caractérisé d'un point de vue phénoménologique. Ses différentes phases (G1, S, G2 et M) sont décrites à l'échelle cellulaire en observant les structures que forment la cellule et sa ploïdie, à l'aide de la microscopie ou de techniques de tri de cellules comme la cytométrie en flux. Chaque phase peut aussi être décrite par une succession d'évènements moléculaires (*e.g.* activation ou inhibition de régulateurs biochimiques) qui sont corrélés aux évènements cellulaires du cycle et qui ont lieu dans un ordre précis afin de préserver l'intégrité de l'ADN. Un exemple typique est la préparation de la phase S. Au cours de la phase G1, les complexes de pré-réplication (abrégié CPR) sont recrutés exhaustivement à leurs origines de réplication respectives. Ces recrutements sont initiés par les complexes cyclineE/cdk2. L'intégrité génétique de la cellule est préservée lorsque le recrutement exhaustif des CPR a lieu *avant* leur activation, qui est quant à elle activée par le complexe cyclineA/cdk2, sans quoi la cellule s'expose à des mutations délétères. Ces mécanismes de préservation, que l'on appelle *checkpoint*, veillent au bon déroulement des évènements du cycle cellulaire dans un ordre précis, de sorte que l'ADN soit totalement dupliqué, qu'il n'y ait pas plus d'une copie par cycle de réplication et que l'ADN soit réparti équitablement entre les deux cellules filles.

Ces mécanismes de contrôle sont encodés dans la structure des régulations impliquant les principaux régulateurs du cycle cellulaires et des boucles de rétroactions couplées de manière complexe [44]. La régulation du cycle cellulaire a été largement étudiée depuis la découverte des complexes cyclines/cdks par Tim Hunt et Paul Nurse en 2001, prix Nobel de physiologie ou de médecine. Expérimentalement, l'enjeu est d'identifier les principaux acteurs de la régulation du cycle cellulaire que l'on peut corréliser à des phénomènes cellulaires, comme le passage du point de restriction ou la transition G1/S par exemple. L'enjeu de modélisation réside quant à lui dans la gestion de la complexité et la caractérisation des structures de régulation impliquées dans l'émergence des phénomènes checkpoints.

Ce chapitre propose un tour d'horizon des modèles mathématiques qui traitent du contrôle du cycle cellulaire et de leur manière de se saisir de la question des checkpoints. L'approche historique est celle de la modélisation par des systèmes d'équations différentielles ordinaires (section 3.2) mais elle s'est rapidement confrontée à un processus coûteux d'ajustement des paramètres aux données expérimentales. Des approches qualitatives ont suivi et ont prouvé leur capacité à reproduire les principaux comportements du cycle cellulaire dans un univers purement discret (section 3.3). Ces approches se focalisent sur l'étude des motifs d'interactions qui sont nécessaires ou suffisants pour obtenir des comportements fondamentaux. À la lumière de ces approches, nous démontrerons que les approches purement discrètes s'avèrent être les plus adaptées pour formaliser et vérifier les propriétés dynamiques caractéristiques des checkpoints (section 3.4). La table 3.1 en donne un aperçu global.

| Modèle | Formalisme | Formalisation d'une phase | Formalisation d'un checkpoint intrinsèque |
|--------|----------------------------------|--|--|
| [86] | Booléen | Motif conditionnellement stable (MCS) d'un réseau étendu | Condition suffisante pour la stabilité du MCS (logique propositionnelle restreinte à la conjonction et à la négation.) |
| [87] | Booléen | Partition d'un graphe résumé d'un graphe de transition asynchrone | Irréversibilité des transitions entre deux phases adjacentes |
| [74] | Multivalué | État stable (quiescence uniquement) | Formule CTL qui encode l'inexistence de certains chemins dans un graphe de transitions |
| [69] | Multivalué puis hybride | Un chemin arbitraire de la logique de Hoare génétiquement modifiée | Non abordée |
| [63] | Multivalué | Non abordée | La voie ATM/ATR-p38MAP dans le cas de dommages à l'ADN (checkpoint extrinsèque). |
| [42] | Continu (EDO) | État (transitoirement) stable (G1, S et M) | Voie ATR/Chk1 (checkpoint de réplication de l'ADN) |
| [40] | Continu (EDO) | État (transitoirement) stable (G1 et S/G2/M uniquement) | Switch bistable |
| [22] | Équations linéaires par morceaux | Non abordée | Une formule CTL qui vérifie le passage obligatoire dans un état checkpoint avant la transition vers la phase suivante. |

Table 3.1 – **Synthèse de la formalisation des phases et de ses checkpoints au sein de modèles marquants dans la compréhension du phénomène et de ses causes.** La formalisation d'une phase est indissociable de celle des checkpoints. La question de la représentation et la validation biologique du domaine d'une phase reste une question ouverte. Aussi, aucune étude à notre connaissance ne s'est penchée sur la caractérisation de toutes les phases du cycle cellulaire (G1,S,G2 et M) et des contraintes dynamiques que sous-tendent les checkpoints.

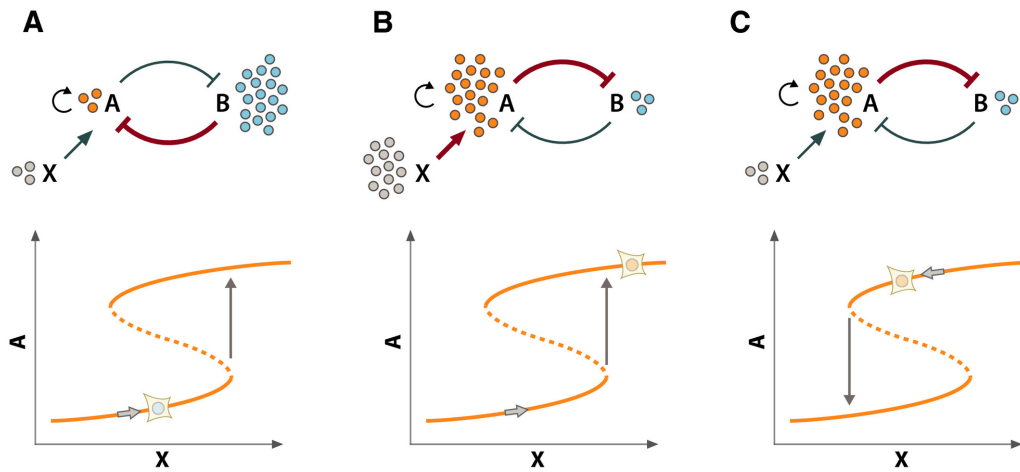


Figure 3.1 – **Le mécanisme de switch stable et sa propriété d’hystérésis : un premier pas vers la compréhension des checkpoints.** Les variables A et B sont munies d’un seuil d’activation qui, lorsqu’il est franchi, permet au système de se stabiliser exclusivement dans un ou l’autre des états stables. Le premier état stable est caractérisé par A inactif et B actif (partie A). Dans cet exemple, l’activation X permet à A d’atteindre une concentration suffisante pour inhiber B ce qui plonge le système dans l’état stable caractérisé par l’inactivation de B et l’activation de A (partie B). L’ajout du signal d’activation de X vers A fait basculer le système *irréversiblement* dans un des états stables qu’exhibe le switch bistable. Dans l’exemple choisi, l’auto-activation de A et l’inhibition de B par A sont suffisantes pour maintenir A stablement actif, qu’importe la valeur de concentration de X (partie C de la figure). Cette figure est adaptée de [93].

3.2 Les modèles continus pionniers

La modélisation par système d’équations différentielles ordinaires (abrégées EDO) a d’abord été l’approche la plus commune et a permis d’expliquer structurellement l’émergence des propriétés dynamiques du cycle cellulaire. Nous mentionnerons la séparation temporelle de la réplication et de la mitose et l’uni-directionnalité du cycle cellulaire. Un système d’EDO représente de manière continue les variations de concentrations des entités biochimiques, concernées par la question biologique, comme une fonction continue de la concentration de ses régulateurs, *cf.* chapitre 2 section 2.2.

3.2.1 Identification d’une structure de régulation : les switches bistables

Les modèles phares sont [23, 40, 42]. Ils mettent en évidence que la régulation des principaux complexes cycline/cdk est non linéaire et repose sur un mécanisme de *switch stable*. Un switch bistable est une boucle de rétroaction positive (nombre pair d’inhibitions) qui génère deux états stables *distincts*. La figure 3.1 est un exemple jouet où deux états stables distincts sont générés par une inhibition mutuelle entre deux variables notées A et B . Le modèle [42] exhibe également deux états stables distincts : une cellule en phase de quiescence et une cellule prolifération. Une cellule quiescente est ici associée à une inactivation constante des différents complexes cycline/cdk, à savoir les cyclineD/cdk4-6, cyclineE/cdk2, cyclineA/cdk2 et cyclineB/cdk1. Une cellule en prolifération est associée à une oscillation autoentrenue de ces complexes. Le passage d’un état de

quiescence vers un cycle cellulaire auto-entretenu est abrupte et est contrôlé par un switch bistable impliquant les facteurs de croissance, *cf.* figure 3.2. Le mécanisme de switch stable est modélisé

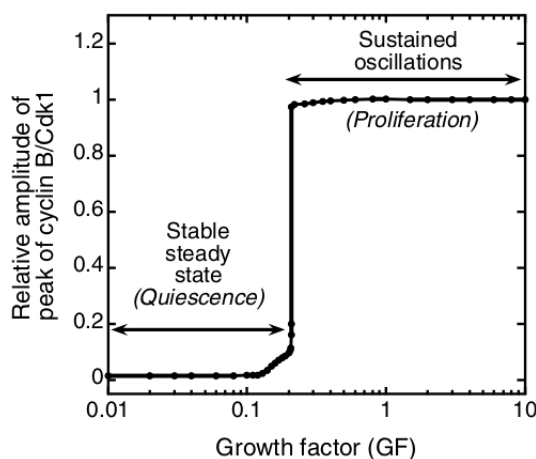


Figure 3.2 – L’activité *switch-like* des complexes régulateurs du cycle cellulaire illustrée avec le complexe cyclineB/cdk1. L’effet des facteurs de croissances (GF) sur l’activité du complexe cyclineB/cdk1 peut être approximé par une fonction seuil (*e.g.* figure 2.2) ou par une fonction logique. Cette figure est adaptée de [42].

systématiquement *via* une fonction de Hill (définition 1 page 18) et une paramétrisation appropriée permet de reproduire un phénomène d’*ultrasensibilité*. Ce phénomène traduit le fait qu’un régulateur possède un seuil de concentration au delà duquel sa cible bascule *quasi-instantanément* d’un état inactif vers un état actif.

La modélisation de [40] a poussé le degrés d’abstraction du phénomène de contrôle du cycle cellulaire à son paroxysme jusqu’alors. Une phase y est représentée d’un point de vue phénoménologique par un état stable auto-entretenu. L’intégrité de la phase est assurée par un *checkpoint* qui empêche le chevauchement avec la phase suivante en maintenant transitoirement la stabilité de la phase courante. Les auteurs se représentent structurellement un checkpoint par un switch bistable qui génère la stabilité d’un état caractéristique d’une phase. L’état reste stable tant qu’un certain ensemble d’évènements de régulation, *i.e.* activation ou inhibition brutale de la concentration d’un régulateur, n’a pas eu lieu. c’est uniquement lorsque tous les évènements requis ont eu lieu que la transition vers une phase peut avoir lieu. Nous mentionnions précédemment que la phase G1 ne peut pas débuter tant que le complexe cyclineE/cdk2 n’a pas activé tous les complexes de pré-réplication. Si l’on suit l’exemple fourni par [40] en figure 3.3, la phase G1 est munie d’un checkpoint si elle reste stable justement tant que l’activation du complexe cyclineE/cdk2 (représenté ici par la variable SK , les *starting kinases*) n’a pas eu lieu. La phase G1 est ici caractérisée par une inactivation des complexes cyclineA/cdk1-2 et cyclineB/cdk1, abstraits tout deux par la variable CDK , qu’ils opposent au bloc de phases S/G2/M caractérisé par la variable CDK active.

La partie inférieure de la figure 3.3 met en évidence que la cellule ne peut pas basculer d’un état G1 vers un état S tant que SK n’est pas activé ce qui indique que l’activation de SK est requise avant celle de CDK . C’est uniquement lorsque SK est activée en fin de G1 que le système bascule vers l’état S/G2/M. Cette bascule de G1 vers S repose sur l’activation par SK du switch

bistable formé entre *CDK* et les *Enemies*, une variable qui représente les principaux inhibiteurs de cycline (cf. la partie supérieure de la figure 3.3), à la faveur de l'activation de *CDK*. L'état S/G2/M est ensuite capable de s'auto-entretenir à partir du moment où les *CDK* inhibent leur propre inhibiteur, dans ce cas les *Enemies*.

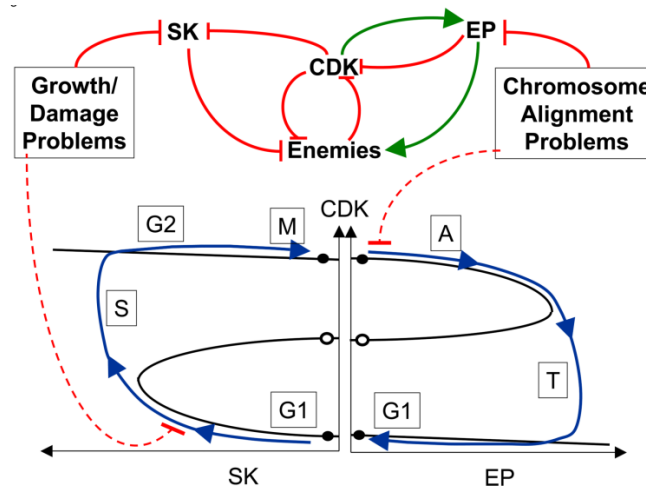


Figure 3.3 – Graphique des états stables du modèle de [40] exhibant les phases G1, S/G2/M et les transitions G1/S et métaphase/anaphase (bas) et les switches stables représentant les mécanismes de checkpoints (haut). La boucle positive formée entre *CDK* et les *Enemies* génère deux états stables. La faible concentration de *CDK* est associée à G1 alors que sa forte activité est associée à S/G2/M. L'activation de *SK* permet l'activation stable de *CDK*. Une fois la *CDK* activée, elle active sa propre répression en activant *EP*. L'activation de cette dernière active la sortie de la mitose en inhibant directement *CDK* ce qui permet de boucler sur l'état qui caractérise la phase G1. Lorsqu'un dommage à l'ADN est détecté en phase G1, la transition G1/S est de surcroît bloquée via l'inhibition de *SK* par la variable « Growth / Damage Problems ». Cette figure est adaptée de [40].

3.2.2 Deux propriétés intrinsèquement discrètes du phénomène de checkpoint

Les modèles continus que nous avons décrits précédemment ont permis d'exhiber des propriétés dynamiques du cycle cellulaire purement qualitatives qui émergent d'un couplage en série des switches bistables formés par les complexes cycline/cdk et leurs inhibiteurs. Ces propriétés dynamiques qui caractérisent le **phénomène de checkpoint** sont les suivantes :

1. La *séparation temporelle des phases* est assurée par un switch bistable qui sépare deux phases successives et contrôle la transition d'une phase vers la phase suivante. Ces dernières peuvent être caractérisées par des états mutuellement exclusifs d'un point de vue logique. Les transitions de phases sont des activations ou inhibitions abruptes des complexes cycline/cdk ou de leurs inhibiteurs ce qui marque une transition nette d'une phase vers une autre. Ces activations et inhibitions abruptes sont d'ailleurs assimilables à des événements discrets. C'est uniquement lorsque les phases sont distinctes que l'ensemble

des évènements appartenant à chaque phase peuvent être isolés et que deux phases adjacentes ne se chevauchent pas. Un switch bistable maintient la stabilité d'une phase jusqu'à ce que tous les évènements requis au cours d'une phase aient eu lieu. Le cas échéant, la phase enclenche sa propre déstabilisation et la transition vers la phase suivante a lieu.

2. *L'unidirectionnalité* du cycle cellulaire est assurée par le phénomène d'hystérésis qui émerge d'un switch bistable. Le signal qui permet à une cellule de basculer dans une phase n'est pas celui qui déstabilise cette même phase. Autrement dit, la transition d'une phase vers la suivante est *irréversible* car est déclenchée par un autre signal que celui qui a permis de l'initier. Le couplage en série de switch bistables à l'aide de boucle de rétroactions négatives (nombre impair d'inhibitions) génère un phénomène de stabilisation en série des différentes phases qui induisent chacune leur propre déstabilisation lorsque tous les évènements requis ont eu lieu. Nous illustrerons cette propriété avec les deux premiers points de contrôle du cycle cellulaire. Le premier switch stable du cours du cycle cellulaire est activé par des facteurs de croissances et permet d'engager irréversiblement la cellule en G1 *via* la stabilisation autonome de la cyclineE/cdk2. Ce switch stable est formé une boucle positive entre E2F et le complexe cyclineE/cdk2 [93]. Ce complexe, une fois stabilisé, active à la fois sa propre inhibition et la stabilisation de la phase S en activant stablement la cyclineA/cdk1 [40]. Trois autres points de contrôles sont ensuite également couplés en séries : la transition S/G2, G2/M et métaphase/anaphase.

[40] et [93] mettent en exergue que les checkpoints reflètent des mécanismes d'inter-dépendances entre des évènements discrets. Ils ne consistent pas à ralentir une phase si un évènement crucial n'a pas lieu mais à la bloquer totalement tant que cet évènement n'a pas lieu. Ces phénomènes de checkpoint ne sont en revanche pas formalisés puisque le choix d'un formalisme continu ne fournit pas le langage approprié pour exprimer des contraintes de dépendances logiques entre évènements discrets, ou au minimum complexifie inutilement le problème de la formalisation cette contrainte. Pour autant, un premier effort de formalisation des phases du cycle cellulaire a été réalisé, ne serait-ce que parce qu'un checkpoint est une notion indissociable d'une phase : sa formalisation au préalable est nécessaire.

[42] est l'étude qui offre la granularité la plus fine. Les variables cyclineD/cdk4-6, cyclineE/cdk2, cyclineA/cdk2 et cyclineB/cdk1 permettent de représenter respectivement les phases G1, S, et M par un état actif (transitoirement) stable. L'étude de [40] identifie l'évènement de régulation qui permet d'atteindre le point de contrôle de G1 et donc d'effectuer la transition G1/S. L'activation de *SK* est donc requise avant celle de *CDK*. Cette contrainte exprimée dans des termes logiques expriment une caractéristique fondamentale du phénomène de checkpoint. Pour autant, la formalisation de cette contrainte sur la chronologie des évènements n'est pas formulée dans le cadre continu ou alors uniquement dans des termes chronométriques. Le modèle de [42] en est un parfait exemple : la voie ATR/Chk1 a été ajoutée au système d'EDO pour contraindre la période des phases G1 et S afin qu'elles ne se chevauchent pas. Or ce n'est pas l'ajout d'un délai qui doit rendre le phénomène de checkpoint opérationnel. Le phénomène prend racine dans la structure même des régulations. Les formalismes logiques tels que les réseaux booléens ou le formalisme de René Thomas sont les plus appropriés pour représenter une phase par un ensemble d'évènements discrets *et* des contraintes logiques sur les transitions entre ces phases. En revanche, on souligne

d'avance qu'il n'existe pas, à notre connaissance, de modèles qualitatifs dédiés à la formalisation des phases donc des checkpoints.

3.3 Les modèles discrets et la formalisation du phénomène de checkpoint

[28] furent les premiers à prouver qu'un formalisme de modélisation logique est capable de reproduire les propriétés asymptotiques clés du cycle cellulaire, exhibées par les modèles de [23]. Ce modèle à base d'EDO a été traduit en un réseau booléen générique du cycle cellulaire qui traduit les mécanismes de contrôle de l'entrée dans le cycle cellulaire communs à la levure et aux mammifères. L'analyse du graphe de transitions révèle les mêmes attracteurs mais transposés dans un monde discret. La phase de quiescence est représentée par un état stable du graphe de transition, déduit des fonctions booléennes associées aux variables de [23]. L'oscillation auto-entretenu des complexes cycline/cdk est représentée par une composante fortement connexe terminale. Cet attracteur complexe est associé au cycle cellulaire lui-même, en constante perfusion d'une concentration suffisante en facteurs de croissance. Ce modèle ne vérifie pas une propriété indissociable du cycle cellulaire, à savoir la correcte succession des phases G1, S, G2 et M. Les contraintes qui opèrent au niveau des transitions critiques du cycle cellulaire ne sont donc pas abordées.

3.3.1 Un exemple de modélisation discrète des régulations impliquées dans certains checkpoints

Le modèle de [87] propose d'étendre l'analyse de l'attracteur complexe de [28] à l'aide des graphes résumés (figure 3.4). Les auteurs définissent des k -partitions d'un attracteur com-

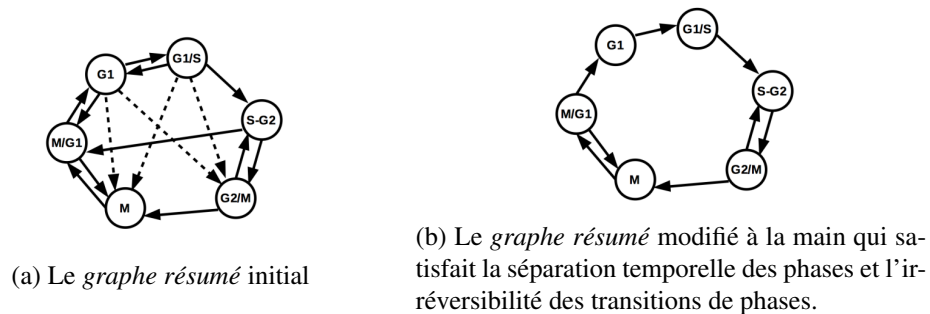


Figure 3.4 – Un *graphe résumé* du cycle cellulaire proposé par [87] à partir du modèle booléen de [28]. Les marqueurs utilisés pour définir une phase sont les états (actif ou inactif) des variables $E2F$, les cyclines E, A et B, Cdc20 et Cdh1. Cette figure est adaptée de [87].

plexe, où les partitions sont non vides et mutuellement exclusives, et k est un entier. La disjonction des partitions assure que les phases ne se chevauchent pas, c'est à dire la séparation temporelle des phases : une des caractéristiques du phénomène de checkpoint mentionnées plus haut. Les états appartenant au cycle cellulaire ont été classés dans six partitions à partir de la définition d'une formule caractéristique (restreinte à la logique propositionnelle) de G1, G1/S, S/G2, G2/M, M et M/G1. Chacune encode un état particulier de $E2F$, cyclineE/cdk2, cyclineA/cdk1-2 et cyclineB/cdk1 et sont définies de sorte d'être mutuellement exclusives. Ces partitions sont

en réalité des semi-partitions puisque les auteurs exhibent une *partition inconnue* non vide qui rassemble tous les états qui n'ont pas été classés dans une des phases ou portions de phase mentionnées précédemment, *i.e* qui ne satisfont aucune des formules caractéristiques des phases. À partir du graphe de transition et des six partitions, les auteurs reconstruisent les transitions entre les différents phases et ont pour objectif de vérifier le seul ordonnancement correct, c'est à dire $G1 \rightarrow G1/S \rightarrow S/G2 \rightarrow G2/M \rightarrow M \rightarrow M/G1$, soit celui qui garantit que les évènements d'une phase ont tous eu lieu avant d'atteindre la phase suivante et l'unidirectionnalité du cycle cellulaire. À l'aide de l'algorithme REVEAL [12], les auteurs infèrent un réseau booléen à partir du graphe résumé modifié qui vérifie la séquence correcte définie plus haut, *cf.* figure 3.4. Les auteurs ne se sont en revanche pas focalisés sur la spécification des contraintes qu'une phase doit vérifier avant d'effectuer sa transition irréversible vers la phase suivante. En effet, ces contraintes s'appliquent sur des évènements qui doivent avoir lieu dans un certain ordre et les graphes résumés n'ont pas le pouvoir d'expression adapté pour identifier les évènements (*i.e* les transitions) clés non permutables. Une phase reste toujours une boîte noire, en termes d'évènements, dont le domaine est fixé arbitrairement.

3.3.2 Un exemple de formalisation de certaines contraintes chronologiques au cours du cycle cellulaire

Le modèle de [74] utilise la logique CTL, présentée dans le chapitre suivant, pour formaliser un chemin caractéristique du cycle cellulaire, à partir d'une version mise à jour du modèle de [28]. Le cycle cellulaire est représenté mathématiquement par une formule CTL de la forme :

$$E[G1 \cup S \wedge E[S \cup G2 \wedge E[G2 \cup M \wedge E[M \cup G1]]]]$$

qui peut être schématisée par :

$$G1 \xrightarrow{G1/S} S \xrightarrow{S/G2} G2 \xrightarrow{G2/M} M \xrightarrow{M/G1} G1$$

où :

- $G1, S, G2$ et M sont des formules caractéristiques des états marqueurs de la phase correspondante,
- E correspond au quantificateur existentiel de chemins (d'un graphe de transition),
- \cup est un opérateur temporel binaire, par exemple $\varphi \cup \psi$ (lire « φ until ψ ») traduit le long d'une séquence d'états qu'une formule propositionnelle φ est vraie jusqu'à atteindre un état où une autre formule propositionnelle ψ est vraie, *cf.* figure 3.5.

De nouveau, les auteurs ne se focalisent pas sur la définition des formules caractéristiques des phases mais s'interrogent *formellement* sur l'existence d'une séquence *particulière* d'états des complexes cyclineE/cdk2 (noté E pour faire simple), cyclineA/cdk1 (noté A) et cyclineB/cdk1 (noté B) au sein du graphe de transitions associé à leur modèle :

$$(E = 0 \wedge A = 0 \wedge B = 0) \xrightarrow{E^+} (E = 1 \wedge A = 0 \wedge B = 0) \xrightarrow{A^+} (E = 1 \wedge A = 1 \wedge B = 0)$$

$$\xrightarrow{E^-} (E = 0 \wedge A = 1 \wedge B = 0) \xrightarrow{B^+} (E = 0 \wedge A = 1 \wedge B = 1)$$

$$\xrightarrow{A^-} (E = 0 \wedge A = 0 \wedge B = 1) \xrightarrow{B^-} (E = 0 \wedge A = 0 \wedge B = 0)$$

où $(E = 0 \wedge A = 0 \wedge B = 0)$ spécifie une cellule en début de G1. Certaines transitions exhibées par cette séquence sont associées à l'initiation d'une phase, par exemple A^+ initie la phase S et

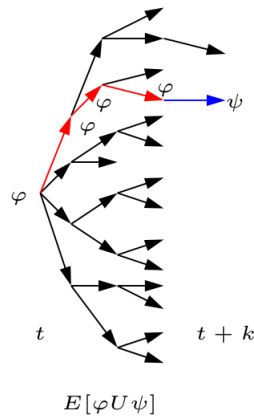


Figure 3.5 – **Contrainte sur l'ordre des phases exprimée à l'aide d'une formule CTL et de la représentation arborescente des chemins d'un graphe de transitions** La formule $E[\varphi U \psi]$ encode une contrainte sur l'ordonnancement de φ et ψ qui assure leur séparation temporelle. [22] et [74] caractérisent un checkpoint du cycle cellulaire par la formule $\neg E[(\neg\varphi) U \psi]$, cf. le corps de texte pour plus de détails.

$B+$ la phase M. L'existence de cette séquence particulière observée expérimentalement est vérifiée automatiquement (et très efficacement) à l'aide du model-checker Nu-SMV. Des expériences de perte de fonction *in silico* ont permis de mettre en évidence que le blocage de l'activation de certaines variables empêche certaines transitions de phases. Par exemple, la simulation du système après la perte de fonction du complexe cyclineE/cdk2 montre un arrêt de la cellule avant la phase S, c'est à dire avant la transition $A+$. Une collection de perturbation a permis aux auteurs d'identifier certains évènements de régulations requis avant la transition d'une phase vers la suivante. En outre, le choix initial d'une sémantique asynchrone révèle qu'il existe d'autres transitions réalisables à partir de certains états de leur séquence du cycle cellulaire. L'analyse des transitions alternatives le long de la séquence caractéristique du cycle cellulaire définie plus haut a permis de vérifier que certains évènements n'ont pas lieu avant d'autres, par exemple l'activation de E et A ne sont pas permutable ce qui permet de séparer temporellement G1 et S. Les checkpoints sont également étudiés sous le prisme de l'irréversibilité de certains évènements. Par exemple, l'irréversibilité de l'inactivation de E est validée par leur modèle. L'efficacité de l'algorithme de model-checking utilisé a permis aux auteurs de vérifier une collection de séquences définies *a priori*, dont certaines sont admises biologiquement et d'autres non.

Cette étude, inspirée de celle de [22], inaugure la formalisation logique d'un checkpoint avec la formule CTL $\neg E[(\neg\varphi) U \psi]$ qui vérifie que la visite des états caractérisés par φ est requise obligatoirement avant ceux caractérisés par ψ . Elle signifie littéralement qu'il n'existe aucun chemin qui permet d'atteindre les états caractérisés par ψ sans passer obligatoirement avant par ceux caractérisés par φ . La formalisation au préalable de φ et ψ est essentielle et repose plutôt sur une caractérisation des évènements requis lors d'une phase et avant l'initiation de la phase suivante. La question des évènements qui composent une phase n'est pas établie dans cette étude de modélisation ce qui implique que la question de la satisfiabilité des checkpoints ne peut pas être élucidée. Cette étude vise plutôt à prouver la pertinence des méthodes formelles dans la formalisation et la

vérification automatique des propriétés dynamiques listées par [28]. L'enjeu est désormais de se focaliser sur une définition raisonnable de φ et ψ , c'est à dire du domaine d'une phase du cycle cellulaire. La formule caractéristique du domaine d'une phase est déduite d'un ensemble d'évènements et d'un état initial qui lui sont associés.

3.3.3 Une première tentative de formalisation des phases du cycle cellulaire

La partie discrète de l'étude de modélisation hybride du cycle cellulaire de [69] offre justement un premier exemple de modélisation d'une phase par une séquence d'évènements requis à partir d'un état initial, à l'aide de la logique de Hoare génétiquement modifiée. Ils utilisent la syntaxe de la logique de Hoare pour représenter un évènement discret par une incrémentation ($v+$) ou décrémentation ($v-$) d'une variable biologique. Un état initial est représenté par une formule propositionnelle. Les auteurs ont transposé le modèle de [40] dans un monde discret à l'aide du formalisme multivalué de René Thomas. Une des limitations du graphe d'interactions du modèle de [40], en ce qui concerne la formalisation des phases, est le choix des variables qui ne permettait pas de distinguer les phases S, G2 et M. Les auteurs proposent donc de scinder *CDK* en deux variables : *A* représente le complexe cyclineA/cdk1-2 impliqué dans l'initiation et la terminaison de la phase S et *B* le complexe cyclineB/cdk1 impliqué dans l'initiation de la phase M. Les phases sont formalisées par :

$$\begin{aligned} G1 &: sk+; sk+; en- \\ S &: a+; sk-; sk-; b+ \\ G2 &: a-; ep+ \\ M &: en+; b-; ep- \end{aligned}$$

où l'état initial du cycle cellulaire (celui de G1) est formalisé par une formule qui caractérise l'inactivation de tous les complexes cycline/cdk et de *EP* et l'activation de *En*. Il s'agit, à notre connaissance, de la première étude qui explicite un ensemble d'évènements requis lors d'une phase.

De nouveau dans cette étude, l'existence de ce chemin est vérifié formellement mais cette fois à l'aide de la logique de Hoare génétiquement modifiée. Cette méthode formelle réalise automatiquement la preuve de correction d'un triplet de Hoare de la forme $\{precondition\} \text{ chemin } \{postcondition\}$ où la précondition encode, dans l'exemple de [69], un unique état initial, le chemin est une séquence d'évènements discrets et la postcondition correspond à l'état atteint après exécution du chemin. Un triplet de Hoare correct vérifie, étant donné un graphe de transitions, que l'état caractérisé par la *postcondition* est atteignable à partir d'un état satisfaisant la *precondition* en passant par le *chemin*. Nous ne détaillerons pas davantage puisque le chapitre suivant est dédié à la présentation formelle de cette méthode. Un ensemble d'évènements requis par phase est ici explicité et validé. En revanche, les contraintes de séparation temporelle des phases et d'irréversibilité ne sont pas définies alors même que la notion d'évènements requis lors d'une phase est introduite.

La séparation temporelle impose en fait que *l'évènement qui clôt une phase ne puisse pas avoir lieu avant celui qui initie la phase suivante*. Dans la continuité de cette étude, cette thèse se focalisera sur la formalisation de cette contrainte en suivant une *approche incrémentale*. Au préalable, une étape essentielle est le raffinement de la formalisation des phases proposée initialement par [69]. Nous avons réalisé ce raffinement dans l'article [96]. Ces travaux préliminaires ont été participé à la formulation de la question principale de la thèse, c'est pourquoi nous le décrivons dans cette section.

L'article part du modèle de [69] et enchaîne sur la question des ordonnancements des évènements au sein d'une phase. Cette étude a pour origine un constat des auteurs de [74] : les connaissances biologiques admettent plusieurs ordonnancements au sein de leur modèle à sémantique asynchrone. Dans la même veine, l'introduction des paramètres $K_{v\omega}$ a le pouvoir de relâcher encore plus la contrainte sur l'ordre des évènements au sein d'une phase, au sens où aucune fonction logique d'activation des variables n'est présupposée ce qui implique qu'aucun modèle est arbitrairement choisi, cf. chapitre 2 section 2.3. Une fois seulement après s'être affranchi d'un choix arbitraire d'un unique modèle et d'un unique ordonnancement, les connaissances biologiques peuvent être interrogées formellement afin d'exhiber *tous* les ordonnancements des évènements au sein d'une phase qui sont *admis biologiquement*, et donc quels évènements peuvent clore une phase et initier la phase suivante (exemple en table 3.2).

Les connaissances biologique du modèle de [69] ont été interrogées formellement à l'aide de l'outil TotemBioNet, décrit au chapitre suivant. Très succinctement, TotemBioNet est une plate-forme logicielle qui assiste le modélisateur dans la conception et la paramétrisation de graphe de régulation multivalué à *la Thomas*, en combinant efficacement la logique de Hoare génétiquement modifiée et du model-checking de formules CTL. Il énumère efficacement tous les modèles associés à un graphe de régulations donné. L'astuce principale réside dans la capacité de la logique de Hoare génétiquement à réduire les bornes des domaines des paramètres $K_{v,\omega}$, à l'aide d'une contrainte issue du calcul automatique de la preuve de correction de la plus faible précondition, cf. chapitre 4. C'est grâce à cet outil et aux quantificateurs de chemin de la logique de Hoare que l'on identifie les séquences des évènements requis lors d'une phase qui sont admises par un graphe de régulation donné. Sur les 100 800 modèles du graphe de régulation de [69] qui vérifient la condition de Snoussi (définition 8), 676 modèles vérifient le cycle cellulaire représenté par le triplet :

$$\{G1_{init}\} \underbrace{sk+; sk+; en-}_{G1}; \underbrace{a+; sk-; sk-; b+}_{S}; \underbrace{a-; ep+}_{G2}; \underbrace{en+; b-; ep-}_{M} \{G1_{init}\}$$

tel que $G1_{init} \equiv (sk = 0 \wedge a = 0 \wedge b = 0 \wedge ep = 0 \wedge en = 1)$. L'étude conclut que les phases G1 et G2, telles que définies, admettent tous les ordonnancements des évènements qui leur sont associés, ce qui suggère de considérer une phase non pas comme une séquence d'évènements mais un multi-ensemble d'évènements. L'initiation des phases S et M sont quant à elle totalement contraintes par les connaissances biologiques, cf. table 3.2 Les évènements clés du cycle cellulaire contraints par les checkpoints sont identifiés. En revanche, l'enjeu consiste toujours à définir raisonnablement la contrainte elle-même qui assure *aucune permutation de ces évènements clés*.

| Phase | Évènement qui peut l'initier | Évènement qui peut la clore |
|-------|------------------------------|-----------------------------|
| G1 | $sk+$ ou $en-$ | $sk+$ ou $en-$ |
| S | $a+$ | $b+$ |
| G2 | $ep+$ ou $a-$ | $a-$ ou $ep+$ |
| M | $en+$ | $ep-$ |

Table 3.2 – Les évènements qui peuvent initier et clore les phases définies initialement par [69]. Un checkpoint rend non permutable tout évènement qui peut clore une phase et initier la phase suivante. Par exemple, $sk+$ ou $en-$ peut clore G1 et $a+$ est le seul évènement qui peut initier la phase S. $sk+$; $sk+$; $a+$ ou $sk+$; $en-$; $a+$ est nécessairement un chemin que l'on ne doit pas pouvoir déduire du graphe de régulation de [69] à partir de l'état initial $G1_{init}$. L'étude de [96] explicite les évènements non permutable au cours d'une phase, en vue de la validation des checkpoints G1/S, S/G2 et G2/M.

3.4 La formalisation des phases munies de checkpoints : la question reste ouverte

La table 3.1 recense les modèles mentionnés dans les deux sections précédentes. Les premiers modèles génériques de la régulation du cycle cellulaire ont permis d'identifier la structure nécessaire à l'émergence d'une séquence de phases munies de checkpoints. Cette structure est un couplage en série de switchs stables par des boucles de rétroaction négatives. Puisque le choix d'un formalisme continu ne permet pas d'identifier un ensemble exhaustif d'états stables générés par ces structures, ni la séparation temporelle des phases ni l'irréversibilité des transitions de phases ne peuvent pas être prouvées.

3.4.1 Relier le mécanisme des checkpoints à ses deux propriétés dynamiques

Récemment, les travaux de [86] proposent d'y remédier avec une étude qualitative poussée des propriétés dynamiques des switchs stables. Ils identifient exhaustivement l'ensemble des états stables associés à un graphe de régulation qui incluent les switchs stables caractérisés par [23, 40, 42]. Ils mettent en évidence que ces états sont conditionnellement stables et expliquent que la condition logique qui rend ces états stables est un checkpoint. À partir d'un réseau booléen, ils énumèrent des *motifs conditionnellement stables* à l'aide d'un graphe étendu. Ce dernier encode à la fois la structure des régulations et les règles logiques d'activation des variables du réseau booléen. Chaque état d'une variable ($v = 1$ noté ici v et $v = 0$ noté $\neg v$) forme potentiellement un noeud du graphe étendu, si l'état de la variable intervient dans une fonction logique du réseau booléen. Un noeud composite, noté \circ , formalise l'opérateur \wedge et ses prédécesseurs sont ses opérands. Un arc et ses prédécesseurs représentent la condition suffisante pour l'observation des cibles de l'arc. Par exemple, dans la figure 3.6, le noeud composite et ses prédécesseurs représentent la condition suffisante pour la stabilité de l'état $x = 1 \wedge y = 1$. Ce type de graphe étendu permet d'identifier l'ensemble exhaustif des motifs conditionnellement stables c'est à dire un ensemble de noeuds d'un graphe étendu qui maintiennent stablement leur état si une condition est satisfaite. Une phase est ensuite associée *a posteriori* à un de ces motifs conditionnellement stables et ils explicitent la condition logique qui caractérise son « checkpoint ». Lorsque cette condition est satisfaite, le système bascule stablement dans une phase qui reste stable jusqu'à ce que la négation

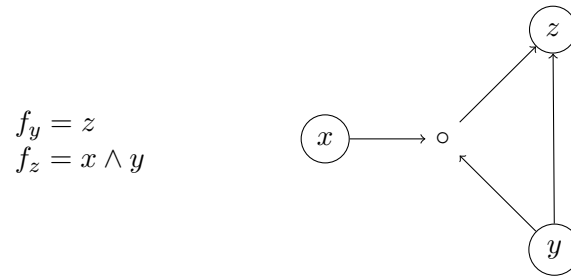


Figure 3.6 – **Exemple jouet de graphe étendu qui explicite un motif conditionnellement stable.** $y \longleftrightarrow z$ est un motif stable puisque l’inhibition réciproque de y et z suffit à maintenir l’état $y = 1 \wedge z = 1$ mais conditionnellement à la présence de x qui représente la condition qu’un motif doit satisfaire pour devenir stable.

de la condition du checkpoint soit satisfaite et déclenche la déstabilisation du motif [29]. En outre, ces résultats font écho aux travaux de [74] au sens où la perte de fonctions des événements requis lors d’une phase rend la condition du motif conditionnellement stable vraie et implique un blocage de la phase. Les auteurs ont ainsi pu expliciter, à l’aide de leur modèle, la condition du passage du point de restriction en G1, la transition S/G2 et la transition métaphase/anaphase. Ce modèle offre une vision plus claire des mécanismes déterminants dans l’émergence d’un checkpoint.

L’étude de [86], couplée à celle de [87], marque la tendance typique qui consiste d’abord à calculer les attracteurs (stables ou conditionnellement stables) d’un modèle (choisi arbitrairement) et à ensuite les associer à la phase qui lui est la plus proche. [74] et [69] font de la formalisation du cycle cellulaire, de ses phases et des checkpoints une question préalable mais elle est étudiée partiellement. Seule une formalisation convaincante de ces phénomènes dynamiques permettra de vérifier de manière convaincante qu’une structure de régulation (*e.g* un switch stable) peut générer ces phénomènes dynamiques (*e.g* checkpoint). Or à ce jour, nous n’avons recensé aucune étude de modélisation dédiée à une formalisation raisonnable des phénomènes *en eux-mêmes* de séparation temporelle des phases et d’irréversibilité des transitions de phases. En revanche, cet état de l’art offre quelques pistes. Une phase munie d’un checkpoint est d’abord formalisée par une formule CTL de la forme $E[\varphi U \psi]$ par [74] et [22]. φ est la formule caractéristique d’une phase et ψ la formule caractéristique des états « checkpoint », *i.e* les états par lesquels il est nécessaire de passer avant d’effectuer la (ou les) transitions vers la phase suivante. C’est à partir de cet état que la transition d’une phase vers une autre peut uniquement avoir lieu, puisqu’il vérifie que tous les événements requis lors d’une phase ont eu lieu. [69] proposent d’utiliser la syntaxe des événements de la logique de Hoare génétiquement modifiée pour représenter cet ensemble d’événements requis. À partir des ensembles d’événements caractéristiques d’une phase, [96] identifient tous les événements clés non permutables, en relâchant toutes les contraintes définies arbitrairement.

3.4.2 Les checkpoints intrinsèques et extrinsèques

Jusqu’ici, tous les travaux présentés admettent l’hypothèse que les processus de réplication et de mitose se déroulent correctement une fois activés. Certains modèles, typiquement celui de [63],

se focalisent sur la modélisation des voies activées lors d'un dommage à l'ADN au cours des différentes phases du cycle cellulaire. Nous distinguerons dans cette thèse deux types de checkpoints et nous focaliserons uniquement sur la modélisation du premier. Nous appellerons *checkpoints intrinsèques*, les mécanismes qui génèrent un ordonnancement correct des événements le long du cycle cellulaire. Nous nommerons *checkpoints extrinsèques*, les mécanismes qui s'ajoutent aux checkpoints intrinsèques et qui sont activés spécifiquement lorsqu'un dommage à l'ADN est détecté. Les checkpoints extrinsèques activent les checkpoints intrinsèques tant que l'ADN n'est pas réparé. La figure 3.3 page 45 fournit deux exemples de checkpoints extrinsèques. Le premier est la voie impliquée dans la détection des dommages à l'ADN lors de la réplication. Cette voie est nommée « *Growth / Damage Problems* » dans la figure et correspond à la voie ATR/Chk1 qui inhibe la transition G1/S tant que l'ADN n'est pas réparé. Le second exemple est l'activation du SAC lorsqu'un problème dans l'alignement des chromosomes le long de la plaque métaphasique est détecté, qui consiste à inhiber la transition métaphase/anaphase. Pour qu'un checkpoint extrinsèque soit vérifié par un modèle, un checkpoint intrinsèque doit être nécessairement formalisé et vérifié avant.

3.5 Bilan sur la modélisation des checkpoints

À la lumière de ce chapitre, nous résumons notre compréhension des checkpoints de la façon suivante : les checkpoints intrinsèques créent des dépendances structurelles entre les événements de régulations du cycle cellulaire. Globalement, ces structures prennent la forme d'une série de switch bistables reliés par des boucles de rétroactions négatives. Ces dépendances permettent de *bloquer* le début d'une phase tant que la phase précédente n'est pas terminée. Si un événement d'une phase est bloqué par des protéines senseurs de dommages à l'ADN, le cycle cellulaire doit se mettre en arrêt. Les phénomènes de checkpoints reposent par nature sur des *concepts discrets*. La thèse vise principalement à modéliser de façon générique les deux phénomènes engendrés par les checkpoints intrinsèques : la séparation temporelle des phases et l'irréversibilité des transitions de phases. Les phases doivent donc être au préalable formalisées. L'étude de modélisation du cycle cellulaire de [69] qui inspire notre graphe de régulation pour l'étude des phases (chapitre 5) prouve que le cycle cellulaire peut être représenté de manière générique par des concepts discrets. Nous réaliserons cette preuve de concept sur les checkpoints des phases du cycle cellulaire. Pour ce faire, on se focalise sur les checkpoints interphases qui sont prototypiques. Ils contrôlent l'initiation ou la terminaison d'une phase.

L'introduction nous apprend que les checkpoints du cycle cellulaire ne se situent pas nécessairement entre deux phases adjacentes. Il existe deux checkpoints intraphases : la transition métaphase/anaphase au sein de la phase M et le passage du point de restriction. Le premier représente un enjeu de recherche, fastidieux et encore peu résolu, qui ne peut pas cohabiter avec celui de la preuve de concept qui fait l'objet de cette thèse. Le second est le checkpoint de restriction situé canoniquement au sein de la phase G1. Il contrôle la transition G0/G1 en fonction de la concentration en facteurs de croissance. Une concentration insuffisante bloque une cellule en G0. Une concentration suffisante permet à la cellule d'entrer en G1 mais une portion précoce reste dépendante des facteurs de croissance. Au sein de cette thèse, la représentation mathématique des facteurs de croissance par une variable d'environnement facilite grandement la vérification du checkpoint de restriction. Une variable d'environnement est sous le contrôle total du modélisa-

teur : sa concentration est fixée . Ce phénomène de checkpoint de restriction est incontournable et modélisable en CTL, c'est pourquoi nous le considérons comme une propriété classique du cycle cellulaire (qui a été vérifiée au chapitre 5). *In fine*, tous ces arguments expliquent que nous nous concentrerons sur les checkpoints G1/S, S/G2 et G2/M au moment de la preuve de concept de la thèse (chapitre 7). Elle consiste à identifier un modèle mathématique discret du cycle cellulaire qui satisfait nos spécifications des deux propriétés des checkpoints intrinsèques que dégage l'état de l'art.

Méthodes formelles utilisées

CHAPITRE 4

Apport de deux méthodes formelles pour les réseaux de régulations discrets

| | | |
|------------|--|-----------|
| 4.1 | Un préambule épistémologique : conjonctions et réfutations . . . | 61 |
| 4.2 | Model-checking de formules CTL | 62 |
| 4.2.1 | Syntaxe de CTL | 63 |
| 4.2.2 | Sémantique de CTL | 63 |
| 4.2.3 | Principe du model-checking | 66 |
| 4.2.4 | CTL et les chemins équitables | 67 |
| 4.2.5 | Quelques exemples de propriétés formalisées en CTL (équitable) | 68 |
| 4.3 | Logique de Hoare génétiquement modifiée | 69 |
| 4.3.1 | Syntaxe | 70 |
| 4.3.2 | Sémantique | 71 |
| 4.3.3 | Règles d'inférence | 72 |
| 4.3.4 | Calcul de la plus faible précondition | 73 |
| 4.4 | Gestion de l'exhaustivité des modèles avec TotemBioNet . . . | 74 |
| 4.4.1 | Entrées et sorties | 75 |
| 4.4.2 | Workflow | 75 |

4.1 Un préambule épistémologique : conjonctions et réfutations

Le formalisme de René Thomas permet de représenter des connaissances biologiques dans un modèle logique. Le chapitre 2 décrit l’approche de modélisation choisie dans le cadre de la thèse qui consiste à formaliser un ensemble de régulations biologiques par un graphe de régulation biologique avec multiplexe (GRBM). La dynamique d’un GRBM noté \mathcal{G} est représentée par un graphe de transition asynchrone qui se déduit d’une paramétrisation σ de \mathcal{G} . Une occurrence (\mathcal{G}, σ) est appelé un modèle dans le contexte de la thèse. La taille des GRBM ne cesse d’augmenter depuis les premiers travaux de modélisation en biologie, l’étape critique se situe donc essentiellement dans l’identification des paramétrisations qui reproduisent des propriétés biologiques.

Des outils adaptés aux formalismes booléens permettent de calculer, pour un modèle donné, les ensembles d’états appartenant à une composante fortement connexe terminale [78] ou à un bassin d’attraction [81]. En revanche, la stratégie qui consiste à analyser l’ensemble des modèles un à un à l’aide de ces outils est fastidieuse, étant donnée l’explosion combinatoire du nombre de paramétrisations.

La stratégie qui consiste à faire un choix arbitraire d’un modèle, sans connaissance biologique à l’appui, est proscrite. Même si certains paramètres $K_{v,\omega}$ peuvent être identifiés en réalisant une expérience de pensée assistée d’un expert biochimiste, il est très rare de réussir à identifier une unique paramétrisation à l’issue de cette expérience. Comme le souligne le chapitre 2 en section 2.4.2.4, la biologie expérimentale est une science qui étudie le vivant dans sa globalité et le paramètre $K_{v,\omega}$ formalise un contexte local : celui de la régulation de v étant donné un ensemble fixé de ressources ω . Le philosophe Willard Van Orman Quine le formule d’ailleurs par : « *Nos énoncés sur le monde extérieur affrontent le tribunal de l’expérience non pas individuellement mais collectivement* » [6].

La stratégie sélectionnée pour cette thèse consiste préférentiellement à réaliser un apprentissage automatique des paramétrisations qui satisfont des propriétés encodées formellement. Deux méthodes formelles classiques du domaine de l’informatique théorique ont été sélectionnées : le *model-checking* et la *logique de Hoare*. Le *model-checking* vérifie qu’un modèle satisfait une propriété biologiques formalisée à l’aide d’une logique temporelle : CTL. Cette méthode formelle a déjà fait ses preuves dans la modélisation de réseaux de régulations biologiques du cycle cellulaire [22, 74].

Cette méthode permet de sélectionner les modèles, parmi tous ceux possibles, qui satisfont des propriétés biologiques dynamiques caractérisées par des formules en CTL. La *logique de Hoare génétiquement modifiée* formalise plus spécifiquement une série temporelle particulière, observée expérimentalement ou déduite indirectement d’un état de l’art. Un enrichissement de l’ensemble des règles d’inférences de la logique de Hoare classique permet de construire un problème de satisfaction de contraintes sur les paramètres appartenant à la famille \mathcal{K} . La résolution de ce problème permet de réduire l’ensemble des paramétrisations possibles à uniquement celles qui vérifient l’existence de la trace spécifiée. Cette logique a notamment permis d’identifier un ensemble de modèles qui satisfont une trace du cycle cellulaire observée expérimentalement [69].

L’outil TotemBioNet identifie de manière automatique la valeur des paramètres $K_{v,\omega}$ en combinant efficacement ces deux méthodes formelles qui ont été adaptées au formalisme de René Thomas. Cet outil permet de sélectionner l’ensemble des modèles *cohérents biologiquement*,

c'est à dire ceux qui satisfont la conjonction des propriétés dynamiques formalisées. Lorsque de nouvelles propriétés biologiques doivent être vérifiées, l'ensemble des autres propriétés testées jusque-là doivent l'être également. *In fine*, l'étape de vérification formelle de propriétés biologiques sur un ensemble exhaustif de modèles a pour objectif de réfuter un GRBM, ce qui est le cas lorsque la conjonction des propriétés dynamiques testées est insatisfiable. Ceci n'est pas sans rappeler le critère de scientificité de Karl Popper : « *Une théorie [en l'occurrence ici un modèle] qui n'est réfutable par aucun événement [en l'occurrence ici une propriété biologique formalisée logiquement] qui se puisse concevoir est dépourvue de caractère scientifique* » [1].

Le choix de ces deux méthodes formelles, parmi d'autres comme les méthodes d'interprétation abstraite pour ne citer qu'elle, répond au besoin de formalisation des propriétés dynamiques en biologie. CTL permet de définir des propriétés sur les états du systèmes de transitions et s'est avérée efficace pour représenter et vérifier des comportements asymptotiques, *i.e* la présence d'un attracteurs cycliques ou d'un ou plusieurs états stables, caractéristiques des comportements du vivant, *i.e* l'homéostasie [74, 76, 69] ou les processus de différenciation cellulaire [10, 13, 20]. La logique de Hoare génétiquement modifiée permet quant à elle de formaliser des chemins particuliers appartenant à des attracteurs cycliques dont la présence a été vérifiée à l'aide d'un algorithme de model-checking.

4.2 Model-checking de formules CTL

D'après la sémantique de René Thomas, plusieurs transitions peuvent avoir lieu à partir d'un état d'un graphe de transition asynchrone (abrégé GTA et formalisé en définition 9 page 35). Plusieurs chemins sont par définition possibles à partir d'un état donné. CTL (*Computational Tree Logic*) est une logique temporelle qui permet de raisonner sur le futur d'un état initial donné, en prenant en compte le non déterminisme admis dans le formalisme de René Thomas.

L'ensemble des états d'un GTA est fini mais le futur d'un état initial est représenté par un arbre d'exécutions infinies qui permettra de vérifier des comportements asymptotiques. CTL introduit des règles syntaxiques qui permettent de formaliser ces comportements. C'est l'objet de la sous-section 4.2.1. La sémantique de CTL est donnée par la relation de satisfaction $\eta \models \varphi$ où η est un état d'un GTA et φ est une propriété CTL bien formée. L'idée centrale du model-checking qui consiste à identifier les états qui satisfont une formule CTL φ . Si tous les états d'un GTA satisfont φ alors on a identifié une paramétrisation σ d'un GRBM qui satisfait le comportement formalisé.

Le sous-ensemble de paramétrisations σ menant à des modèles ne satisfaisant pas φ est caractérisé par $\sigma \not\models \varphi$. Si l'ensemble $\{\sigma \models \varphi\}$ est vide et que l'on conclue que φ traduit correctement le comportement biologique observé alors le GRBM est réfuté. Les définitions de la syntaxe et de la sémantique de CTL qui suivent ainsi que le principe du model-checking sont largement inspirés de [79].

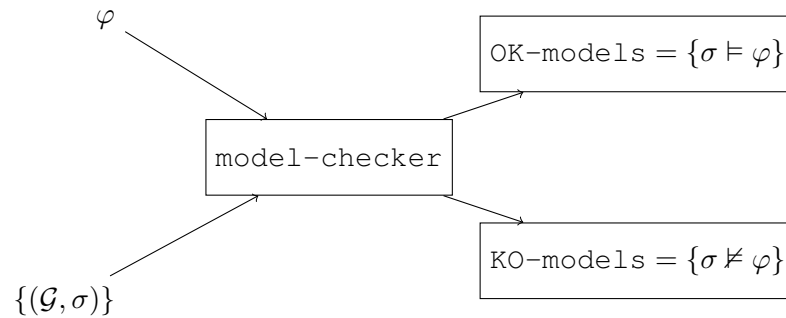


Figure 4.1 – **Vérification de modèle de Thomas.** Sélection (OK-models) et réfutation (KO-models) de modèles (\mathcal{G}, σ) à l’aide d’un model-checker et d’une formule CTL φ .

4.2.1 Syntaxe de CTL

Définition 11: Formule CTL

Étant donné un GRBM $\mathcal{G} = (V, M, E)$, une formule CTL bien formée est définie inductivement comme suit :

- une formule atomique est de la forme \top , \perp ou $v \geq n$ où $v \in V$ et $n \in \llbracket 0, b_v \rrbracket$,
- si φ et ψ sont des formules, alors $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$, $\varphi \Rightarrow \psi$ sont des formules et $AX(\varphi)$, $EX(\varphi)$, $AF(\varphi)$, $EF(\varphi)$, $AG(\varphi)$, $EG(\varphi)$, $A[\varphi \cup \psi]$ et $E[\varphi \cup \psi]$ sont aussi des formules.

Les opérateurs temporels de CTL (G, F, X, U) sont toujours précédés d’un quantificateur de chemin (A, E). Un quantificateur de chemin immédiatement suivi d’un opérateur temporel est appelé une modalité temporelle. Sans modalité temporelle, la langage CTL se réduit au langage des multiplexes (définition 3 page 25).

4.2.2 Sémantique de CTL

D’après la sémantique de René Thomas, plusieurs chemins peuvent se déployer à partir d’un état donné noté η_0 . Si on se borne aux premiers successeurs possibles de η_0 , il y a autant de chemins que de variables qui peuvent se mettre à jour à partir de η_0 . Considérons en table 4.1 un unique chemin p qui se déploie à partir de η_0 . Un opérateur temporel de CTL spécifie les états le long du chemin p qui doivent satisfaire la formule.

| | | |
|---|----------|--|
| G | Globally | φ est satisfaite dans tous les états du chemin p . |
| F | Future | φ est satisfaite dans le futur mais sans précision, c'est à dire dans un état quelconque succédant directement ou indirectement à l'état initial, voire ce dernier (le futur inclut le présent). |
| X | NeXt | φ est satisfaite dans l'état succédant directement à l'état initial. |
| U | Until | φ est satisfaite depuis l'état initial jusqu'à ce que ψ soit satisfaite. Le futur inclut le présent : il se peut que ψ soit satisfaite dès l'état initial donc φ peut ne jamais être satisfaite. |

Table 4.1 – Sémantique de la seconde lettre des modalités temporelles en CTL.

On admet que les formules φ et ψ sont deux formules CTL bien formées. Ainsi intuitivement, la formule $\text{AF}(\varphi)$ signifie qu'à partir de l'état courant, tous les chemins mèneront à un état dans lequel φ sera vraie. Similairement, $\text{E}(\varphi\text{U}\psi)$ signifie qu'il existe un chemin qui atteindra un état où ψ sera vraie et avant d'atteindre cet état, tous les états satisferont φ . Le lecteur non familier avec CTL pourra se référer à [79].

L'interprétation d'une formule CTL bien formée se fait à l'aide d'une structure de Kripke. Il s'agit d'une représentation générique d'un système de transitions, *e.g.* un GTA, duquel on déroule à partir de chacun de ses états un ensemble fini de chemins infinis.

Définition 12: Structure de Kripke

Soit AP un ensemble non vide de propositions atomiques, une structure de Kripke est un triplet de la forme $K = (\zeta, R, L)$ où :

- ζ est un ensemble fini d'états,
- $R \subseteq \zeta \times \zeta$ est une relation totale de transitions telle que $\forall \eta \in \zeta, \exists \eta' \in \zeta \mid (\eta, \eta') \in R$,
- $L : \zeta \rightarrow 2^{AP}$ est une fonction d'étiquetage qui associe à chaque état le sous-ensemble des formules atomiques que cet état satisfait.

Une structure de Kripke est munie d'une fonction d'étiquetage L qui associe à chaque état le sous-ensemble des formules atomiques qu'il satisfait, définition 12. Dans le cadre des réseaux de régulations biologiques, on considère un ensemble de formules atomiques qui permettent de spécifier l'état de chacune des variables. Le choix adopté est $AP = \{v \geq 0 \mid v \in V \text{ et } n \in \llbracket 0, b_v \rrbracket\}$. La structure de Kripke que nous utiliserons par la suite est construite sur le GTA auquel on ajoute la fonction d'étiquetage L définie en définition 12.

Définition 13: Fonction d'étiquetage

Étant donné un GRBM $\mathcal{G} = (V, E, M)$, le sous-ensemble des propositions atomiques satisfaites par un état η de \mathcal{G} est donné par la fonction d'étiquetage L définie par :

$$L : \eta \rightarrow \{v \geq n \mid v \in V \text{ et } n \in \llbracket 0, \eta(v) \rrbracket \text{ et } \eta(v) \geq n\}$$

L'ensemble des formules atomiques est fini puisque l'ensemble des états d'une variable l'est. Une fois l'ensemble des formules atomiques et une formule CTL φ définis, on introduit la relation de satisfaction \models qui relie η_0 et φ , définition 14.

Définition 14: Sémantique d'une formule CTL à l'aide d'une structure de Kripke

Soient un modèle représenté par une structure de Kripke $K = (\zeta, R, L)$ et un état initial η_0 . La sémantique de CTL est définie inductivement par :

- $\eta_0 \models \varphi$ si $\varphi \in L(\eta_0)$,
- $\eta_0 \models \neg\varphi$ si $\eta_0 \not\models \varphi$,
- $\eta_0 \models (\varphi \wedge \psi)$ si $\eta_0 \models \varphi$ et $\eta_0 \models \psi$,
- $\eta_0 \models AX(\varphi)$ si pour tous les successeurs η_1 de η_0 , $\eta_1 \models \varphi$,
- $\eta_0 \models EX(\varphi)$ s'il existe un successeur η_1 de η_0 tel que $\eta_1 \models \varphi$,
- $\eta_0 \models AF(\varphi)$ si pour tout chemin $\eta_0; \eta_1; \dots; \eta_i; \dots$, il existe un j tel que $\eta_j \models \varphi$,
- $\eta_0 \models EF(\varphi)$ s'il existe un chemin $\eta_0; \eta_1; \dots; \eta_i; \dots$, il existe un j tel que $\eta_j \models \varphi$,
- $\eta_0 \models AG(\varphi)$ si pour tout chemin $\eta_0; \eta_1; \dots; \eta_i; \dots$, pour tout $j \geq 0$ tel que $\eta_j \models \varphi$,
- $\eta_0 \models EG(\varphi)$ s'il existe un chemin $\eta_0; \eta_1; \dots; \eta_i; \dots$, il existe $j \in \mathbb{N}$ tel que $\eta_j \models \varphi$,
- $\eta_0 \models A[\varphi \cup \psi]$ si pour tout chemin $\eta_0; \eta_1; \dots; \eta_i; \dots$, il existe un j tel que $\eta_j \models \psi$ et quelque soit $i < j$, $\eta_j \models \varphi$
- $\eta_0 \models E[\varphi \cup \psi]$ s'il existe un chemin $\eta_0; \eta_1; \dots; \eta_i; \dots$, il existe un j tel que $\eta_j \models \psi$ et quelque soit $i < j$, $\eta_j \models \varphi$

Il ne reste plus qu'à introduire la relation de satisfaction qui relie un modèle $\mathcal{R} = (\mathcal{G}, \sigma)$ et φ une formule CTL.

Définition 15: Relation de satisfaction d'une formule CTL par un modèle

Étant donné un RRBM $\mathcal{R} = (\mathcal{G}, \sigma)$ représenté par une structure de Kripke $K = (\zeta, R, L)$ et φ une formule CTL, une paramétrisation σ satisfait φ si tous les états η de \mathcal{G} la satisfont :

$$\frac{\forall \eta \in \mathcal{G} : \eta \models \varphi}{\sigma \models \varphi}$$

On note alors $\sigma \models \varphi$. Par abus de langage, on écrit $\mathcal{R} \models \varphi$ si $\sigma \models \varphi$.

Les mécanismes de régulations encodés dans un graphe \mathcal{G} sont suffisants pour déduire un comportement dynamique caractérisé par φ si et seulement si φ est satisfiable. On dit alors que \mathcal{G} est cohérent biologiquement vis-à-vis de φ et on notera $\mathcal{G} \models \varphi$. On a alors l'équivalence suivante :

$$\mathcal{G} \models \varphi \equiv \exists(\mathcal{G}, \sigma) \mid \sigma \models \varphi$$

Les propriétés classiquement vérifiées sont des propriétés de sûreté (spécifient des comportements qui ne doivent jamais arriver) et de vivacité (spécifient des comportements qui finiront toujours par arriver) et en biologie l'enjeu est de définir une collection de propriétés que l'on suppose suffisantes pour qualifier un GRBM de *cohérent biologiquement*, étant données les variables de \mathcal{G} choisies *a priori* et des propriétés du système étudié. Ces propriétés doivent toujours être reliées à des observations expérimentales à l'étape de vérification de la cohérence biologique de \mathcal{G} mais, par la suite, des hypothèses pourront être testées.

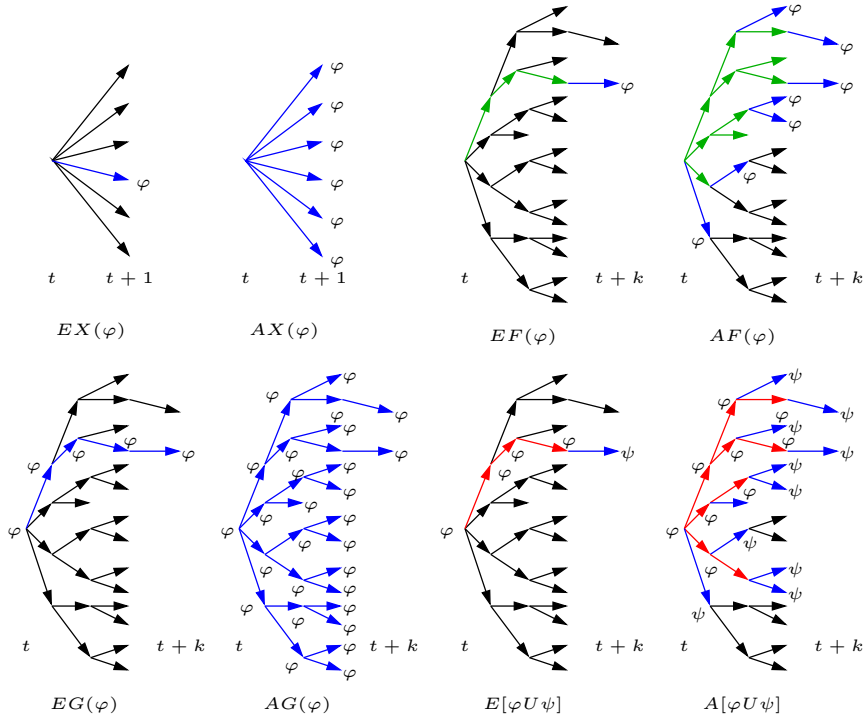


Figure 4.2 – **Illustration de la sémantique de CTL introduite en définition 14.** Un arbre d'exécution (*computation tree*) se déploie à partir d'un état initial d'une structure de Kripke $K = (\zeta, R, L)$ (définition 12). Chaque chemin déployé est une séquence infinie d'états $\eta_0; \eta_1; \dots$ telle que $\eta_0 \in \zeta$ et $(\eta_i, \eta_{i+1}) \in R, \forall i \geq 0$. Pour l'exemple, un seul état initial est considéré.

4.2.3 Principe du model-checking

Le model-checking est un algorithme qui permet de savoir si une formule CTL φ est satisfiable en calculant l'ensemble des états d'une structure de Kripke K qui satisfont φ . Sans perte de généralité, la syntaxe de CTL est réduite aux opérateurs $\neg, \wedge, \text{EX}, \text{EU}, \text{AF}$ puisque l'on admet ces équivalences :

$$\begin{aligned} \text{AG}(\varphi) &\equiv \neg \text{EF}(\neg \varphi) \\ \text{AX}(\varphi) &\equiv \neg \text{EX}(\neg \varphi) \\ \text{EG}(\varphi) &\equiv \neg \text{AF}(\neg \varphi) \\ \text{EF}(\varphi) &\equiv \text{E}[\top \cup \varphi] \\ \text{A}[\varphi \cup \psi] &\equiv \neg(\text{E}[\neg \psi \cup (\neg \varphi \wedge \neg \psi)] \vee \text{EG}(\neg \psi)) \end{aligned}$$

Initialement, tout état de K est étiqueté par le sous ensemble des propositions atomiques de φ qu'il satisfait. Un état est étiqueté par une formule de la forme $\neg \varphi$ s'il n'est pas étiqueté par φ . Une formule de la forme $\varphi \wedge \psi$ est associée à tout état étiqueté à la fois par φ et ψ . Une formule de la forme $\text{EX}(\varphi)$ est associée à tous les prédécesseurs de tous les états étiquetés par φ .

Pour étiqueter les états par une formule de la forme $\text{AF}(\varphi)$, on procède comme suit :

1. tout état déjà étiqueté par φ est étiqueté par $\text{AF}(\varphi)$,
2. tout état tel que tous ses successeurs est déjà étiqueté par $\text{AF}(\varphi)$,
3. l'étape 2 est répétée jusqu'à ce que l'étiquetage des états converge.

Pour cette fois étiqueter les états par une formule de la forme $E[\varphi \cup \psi]$:

1. tout état déjà étiqueté par ψ est étiqueté par $E[\varphi \cup \psi]$,
2. tout état déjà étiqueté par ϕ et pour lequel au moins un successeur a été étiqueté par $E[\varphi \cup \psi]$ est aussi étiqueté par $E[\varphi \cup \psi]$,
3. l'étape 2 est répétée jusqu'à ce que l'étiquetage des états converge.

L'étape finale consiste à renvoyer l'ensemble des états satisfaisant φ . Le model-checker NuSMV implémente un algorithme de model-checking symbolique efficace pour CTL, conçu par [16], qui repose sur une représentation symbolique (plutôt qu'explicite dans l'algorithme naïf présenté plus haut) de la structure de Kripke qui est représentée de manière plus compacte à l'aide d'un OBDD (*Ordered Binary Decision Diagrams*).

4.2.4 CTL et les chemins équitables

Le model-checker Nu-SMV considère la sémantique CTL qui admet des chemins *non équitables*. Un chemin p est dit non équitable s'il passe une infinité de fois par un état qui possède plusieurs successeurs mais qu'il ne passe qu'un nombre fini de fois par certains de ses successeurs. L'équitabilité des chemins est vérifiée s'ils explorent toutes les possibilités. En conséquence, les formules de vivacité de la forme $\text{AF}(\varphi)$, où φ spécifie un attracteur, sont systématiquement évaluées à faux par Nu-SMV dès qu'il existe une composante fortement connexe différente de l'attracteur mais à partir duquel il est possible d'atteindre l'attracteur.

Par exemple, considérons un graphe de transition asynchrone qui génère une oscillation des variables x et y autour du point $(x = 1, y = 1)$ et un état stable caractérisé par $x = 2 \wedge y = 1$. Il saute aux yeux en observant la figure 4.3 que cet état stable est atteignable quelque soit l'état initial considéré. Admettons que l'on souhaite vérifier par model-checking que le système finit toujours par se stabiliser vers la saturation de x à l'aide de la formule $\text{AF}[\text{AG}(x = 2)]$. Nu-SMV conclut que cette formule est fausse puisque le cycle $\eta_4 \rightarrow \eta_3 \rightarrow \eta_2 \rightarrow \eta_1 \rightarrow \eta_4$ peut être emprunté une infinité de fois, alors que deux chemins sont pourtant possibles depuis l'état η_4 . Le chemin qui emprunte une infinité de fois les états $\eta_4 \rightarrow \eta_3 \rightarrow \eta_2 \rightarrow \eta_1 \rightarrow \eta_4$ est donc un chemin non équitable. Pour être en mesure de vérifier que $\eta_4 \rightarrow \eta_5 \rightarrow \eta_6$ finit bien par être emprunté, il est nécessaire d'ignorer les chemins non équitables.

Deux nouveaux quantificateurs de chemins sont introduits en CTL' : A' signifie *pour tout chemin équitable* et E' signifie *il existe un chemin équitable*. Intuitivement, le fait de ne prendre en compte que les chemins équitables revient à se focaliser sur les composantes fortement connexes terminales. Autrement dit, CTL' est bien adaptée pour exprimer les propriétés asymptotiques des systèmes. Adrien Richard a d'ailleurs démontré, dans une note interne [39], qu'il existe pour toute formule CTL', notée ψ , une équivalence sémantique avec une formule CTL φ , c'est à dire :

$$\forall \psi \in \text{CTL}', \\ \llbracket \exists \varphi \in \text{CTL} \mid (\eta \models_{\text{CTL}} \varphi) \equiv (\eta \models_{\text{CTL}'} \psi) \rrbracket$$

Nu-SMV reste le model-checker de prédilection puisqu'il existe une traduction de CTL' vers CTL qui préserve la sémantique des formules. Seul CTL' permet de vérifier qu'un attracteur finira toujours par être atteint ($\text{AF}[\text{AG}(\varphi)]$) par les chemins équitables.

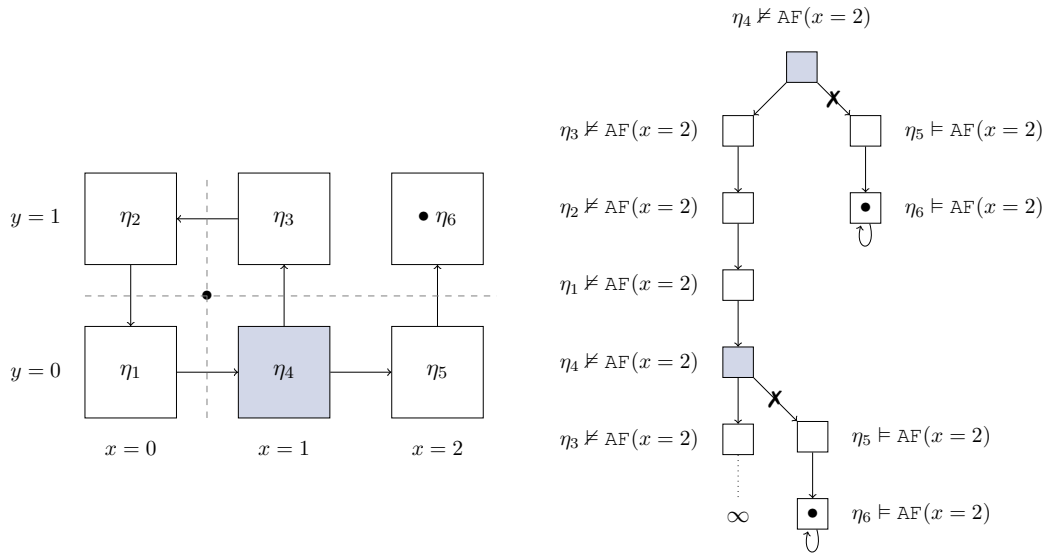


Figure 4.3 – Exemple de GTA pour lequel la formule $\text{AF}(x = 2)$ est satisfaite uniquement en CTL'. Si η_3 et η_5 peuvent être visités une infinité de fois alors les chemins permettant de les atteindre sont équitables. En CTL' tous les états du GTA sont étiquetés $\text{AF}(x = 2)$ alors qu'en CTL l'existence d'un chemin cyclique non équitable pour lequel aucun de ses états ne satisfait $x = 2$ implique $\sigma \not\models \text{AF}(x = 2)$.

4.2.5 Quelques exemples de propriétés formalisées en CTL (équitable)

Nous enchaînons avec quelques exemples de propriétés dynamiques classiquement étudiées dans la modélisation de réseaux de régulations biologiques. Étant donnés deux formules CTL φ et ψ , un graphe de transition asynchrone \mathcal{T} et trois états particuliers de \mathcal{T} notés η_1 , η_2 et η_3 caractérisés par φ_1 , φ_2 , et φ_3 :

1. $\varphi_1 \Rightarrow \text{EF}(\varphi_2)$ est une propriété d'atteignabilité qui vérifie qu'il existe au moins un chemin qui relie η_1 et η_2 .
2. $\varphi_1 \Rightarrow \text{EX}(\varphi_2 \wedge \text{EX}(\varphi_3))$ spécifie l'existence d'un chemin particulier de \mathcal{T} : $\eta_1 \rightarrow \eta_2 \rightarrow \eta_3$.
3. $[\neg \text{EF}(\varphi)] \equiv [\text{AG}(\neg \varphi)]$ est une propriété de sûreté qui vérifie que φ n'est satisfaite par aucun état de \mathcal{T} . Par exemple si φ exprime l'activation d'un gène alors cette propriété de *sûreté* vérifie que ce gène ne sera jamais activé.
4. $\text{AF}[\text{AG}(\varphi_2)]$ est une propriété de *vivacité* sur l'atteignabilité d'un état stable caractérisé par φ_2 . Une propriété de vivacité représente un comportement qui finira toujours par arriver, qu'importe l'état initial.
5. $\text{AF}(\text{AG}[\text{AF}(v = 1) \wedge \text{AF}(v = 2)])$ est une propriété de *vivacité* sur l'atteignabilité d'une oscillation stable de v , plus particulièrement v est ici multivaluée et peut prendre une infinité de fois les valeurs 1 et 2.
6. $\text{AG}(v = n) \Rightarrow \varphi$ se comprend en deux temps. La prémisse spécifie un environnement fixé. Un environnement est une variable qui est stabilisée à une valeur n telle que $0 \leq n \leq b_v$ dans la totalité de l'espace des états de \mathcal{T} . La propriété φ est une formule CTL satisfaite dans cet environnement.

7. $A[c \cup \varphi_1]$: la contrainte c est respectée avant d’atteindre l’état η_1 caractérisé par φ_1 , pour tout chemin de \mathcal{T} [22].
8. $\varphi_1 \rightarrow A[\neg\varphi \cup \psi]$ est la même propriété que la précédente mais formulée sous la forme d’une propriété de sûreté. À partir de l’état η_1 , la propriété φ ne sera jamais vraie avant d’atteindre un état qui satisfait ψ .

Les propriétés asymptotiques que constituent les formules 4 et 5 doivent être interprétées en CTL’. L’identification d’un attracteur nécessite d’ignorer les chemins non équitables, cf. figure 4.3. Par exemple, la sous-formule φ_2 de la formule 4 représente, lorsqu’elle est satisfaite, un ensemble d’états qui appartiennent à un attracteur. Autrement dit, un attracteur est détecté grâce à cette formule.

Les checkpoints peuvent être formalisés par des propriétés de sûreté sous réserve qu’une phase soit correctement formalisée. Par exemple, si on considère dans la formule 8 que φ_1 et φ_2 sont respectivement les états initial et final d’une phase π_i et que φ est l’union des domaines de toutes les autres phases du cycle cellulaire excepté π_i alors la propriété de sûreté vérifie, une fois qu’une phase a démarré, qu’aucune autre phase n’est visitée tant que la phase courante est en cours. L’interprétation de la propriété de sûreté qui pourrait définir un checkpoint dépend ici de la définition au préalable des domaines des différents phases du cycle cellulaire.

4.3 Logique de Hoare génétiquement modifiée

La logique de Hoare génétiquement modifiée permet également de vérifier l’existence de certains chemins au sein d’un graphe de transition asynchrone. Cette logique permet de vérifier la cohérence entre trois éléments : une précondition, une suite d’instructions et une postcondition. Elle s’inspire de la logique de Hoare introduite par Tony Hoare pour raisonner sur la correction d’un programme impératif [2]. Un triplet de Hoare de la forme $H \equiv \{P\} p \{Q\}$ formalise un programme impératif : P et Q caractérisent respectivement les conditions en entrée et en sortie du programme, p est une séquence d’instructions (incrémentations, décrémentation, `while`, `if`). Edsger Dijkstra a défini un algorithme de calcul de la plus faible précondition qui fournit une stratégie unique de preuve de correction de H . Cet algorithme prend en entrée une précondition Q et un programme impératif p et renvoie la plus faible précondition P_0 . L’exécution du programme p à partir de P peut avoir lieu et atteint la postcondition Q si et seulement si $P \models P_0$.

Cette logique a été adaptée pour le cadre de modélisation des réseaux de régulations biologiques avec multiplexes de René Thomas pour identifier si une trace biologique observée expérimentalement est correcte. Cette section vise à expliquer cette notion de *correction* transposée dans le monde de la modélisation de RRBM. Essentiellement, la logique de Hoare *génétiquement modifiée* permet de contraindre les paramétrisations d’un graphe de régulation biologique avec multiplexe (GRBM). Les instructions d’incrémentations et de décrémentation d’une variable, le `while` et le `if` permettent de formaliser subtilement des données d’expressions de gènes ou des conditions expérimentales comme par exemples le *knock-out* ou le *knock-in*, pour plus de détails nous référons le lecteur à [85]. Nous nous focaliserons uniquement sur les instructions d’incrémentations et de décrémentation puisqu’elles sont suffisantes pour les travaux de formalisation des phases du cycle cellulaire présentés dans cette thèse.

4.3.1 Syntaxe

La précondition P et la postcondition Q d'un triplet de Hoare $H \equiv \{P\} p \{Q\}$ sont définies à l'aide d'un langage qui étend celui des multiplexes : le langage des assertions. Les paramètres $K_{v,\omega}$ sont désormais des termes autorisés pour exprimer des contraintes sur leurs valeurs.

Définition 16: Langage des assertions

Étant donné à un GRBM $\mathcal{G} = (V, M, E)$ et la famille des paramètres \mathcal{K} associée à \mathcal{G} , le langage des assertions \mathcal{L}_A est défini inductivement par :

- les atomes sont de la forme $(t \geq t')$ où t et t' sont des termes biens formés à partir de V , de \mathcal{K} et de \mathbb{N} et des opérations binaires $+$ et $-$,
- si $\varphi \in \mathcal{L}_A$ et $\psi \in \mathcal{L}_A$ alors $\neg\varphi \in \mathcal{L}_A$, $(\varphi \wedge \psi) \in \mathcal{L}_A$ et $(\varphi \vee \psi) \in \mathcal{L}_A$.

Par exemple, considérons un GRBM et sa famille de paramètres \mathcal{K} associée en figure 4.4. Les expressions $(K_{x,\{m_1\}} \geq K_{x,\emptyset}) \wedge (K_{x,\{m_3\}} \geq K_{x,\emptyset})$ et $(K_{x,\{m_1\}} \geq 0) \wedge (K_{x,\{m_1\}} \leq b_x)$ sont deux exemples d'assertions de \mathcal{L}_A .

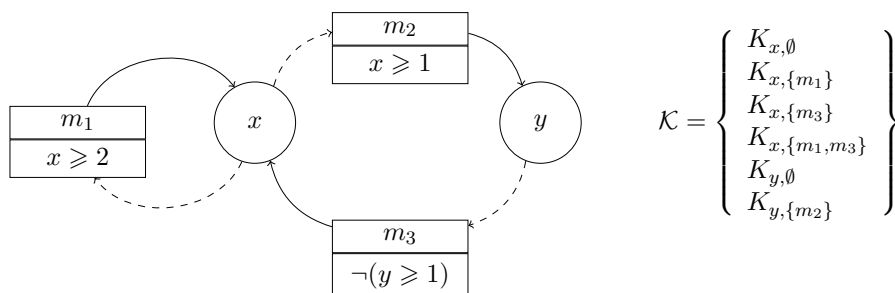


Figure 4.4 – Un exemple de GRBM associé à sa famille de paramètres \mathcal{K} pour illustrer l'enrichissement du langage des assertions par rapport au langage des multiplexes.

Un chemin est dans le contexte de la thèse limité à une séquence d'incrémentations ($v+$) ou de décrémentation ($v-$) d'une variable v . Le langage des chemins permet de représenter une séquence d'évènements discrets extraite de la littérature, directement d'expériences en laboratoire [76, 69] ou après analyse d'une série temporelle (figure 4.5).

Définition 17: Langage de chemin

Étant donné un GRBM $\mathcal{G} = (V, M, E)$, le langage des chemins \mathcal{L}_C est défini inductivement par :

- $\forall v \in V$, les expressions $v+$ et $v-$ appartiennent à \mathcal{L}_C et sont les atomes du langage des chemins,
- si $p_1 \in \mathcal{L}_C$ et $p_2 \in \mathcal{L}_C$ alors $(p_1; p_2) \in \mathcal{L}_C$,
- dans \mathcal{L}_C « $;$ » est associatif.

Les éléments de \mathcal{L}_C sont appelés les chemins de \mathcal{G} . Les atomes de \mathcal{L}_C sont appelés des évènements.

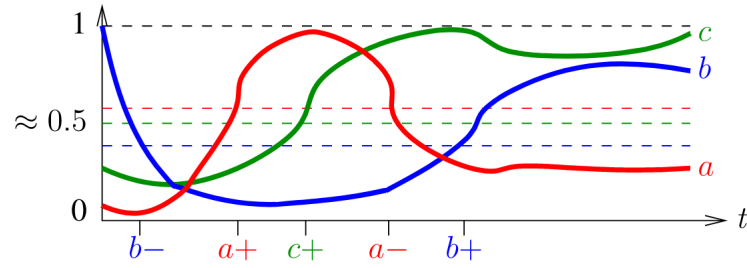


Figure 4.5 – **Un exemple de trace biologique observée expérimentalement.** Une trace est déduite d'une variation de l'activité de variables biologiques au cours du temps où chaque variable est munie d'au moins un seuil d'activation. En ordonnée est représentée la variation de l'activité d'une variable normalisée et la valeur 0.5 choisie pour l'exemple correspond à la moitié de l'activité maximale d'une variable. Figure extraite de [85].

Remarque 1. On restreint à dessein le langage \mathcal{L}_c défini initialement dans [85] au minimum suffisant dans le cadre de cette thèse. Par exemple, \mathcal{L}_c n'inclut ni les quantificateurs de chemins, ni la branche conditionnelle *if*, ni la boucle *while*.

Définition 18: Triplet de Hoare

Un triplet de Hoare associé à un GRBM \mathcal{G} est une expression de la forme $H \equiv \{P\} p \{Q\}$ tels que $P \in \mathcal{L}_A$, $Q \in \mathcal{L}_A$ et $p \in \mathcal{L}_C$.
 P et Q sont respectivement appelés la précondition et la postcondition de H .

4.3.2 Sémantique

La sémantique d'un triplet de Hoare est donnée par un graphe de transition asynchrone. Une trace observée expérimentalement et bien définie syntaxiquement n'a pas nécessairement d'équivalent sémantique. Un triplet de la forme $\{P\} v + \{Q\}$ ou $\{P\} v - \{Q\}$ a un équivalent sémantique si et seulement s'il réfère à une transition appartenant à un graphe de transition asynchrone.

Définition 19: Sémantique d'un chemin

Étant donné un RRBM $\mathcal{R} = (G, \sigma)$, la famille de relations binaires $\overset{p}{\rightsquigarrow}$ sur l'ensemble des états ζ , où p parcourt \mathcal{L}_C , est définie inductivement par :

- si p est le chemin vide alors trivialement $\eta \overset{p}{\rightsquigarrow} \eta$ pour tout $\eta \in \zeta$,
- si $p \equiv v+$: si $\eta \rightarrow \eta'$ est une transition de \mathcal{R} avec $\eta' = \eta[v \leftarrow (\eta(v) + 1)]$ alors $\eta \overset{p}{\rightsquigarrow} \eta'$,
- si $p \equiv v-$: si $\eta \rightarrow \eta'$ est une transition de \mathcal{R} avec $\eta' = \eta[v \leftarrow (\eta(v) - 1)]$ alors $\eta \overset{p}{\rightsquigarrow} \eta'$ et
- si $p \equiv (p_1; p_2)$: si $\eta \overset{p_1}{\rightsquigarrow} \eta'$ et $\eta' \overset{p_2}{\rightsquigarrow} \eta''$ alors $\eta \overset{p}{\rightsquigarrow} \eta''$.

La sémantique d'un chemin définie par la relation $\eta \overset{p}{\rightsquigarrow} \eta'$ peut se lire de la manière suivante : dans un système de transitions défini par un graphe de transition asynchrone, l'état η' est atteignable

depuis l'état η en passant par le chemin spécifié par p .

Définition 20: Sémantique d'un triplet de Hoare

Étant donné un RRBM $\mathcal{R} = (\mathcal{G}, \sigma)$ et ζ l'ensemble des états de son graphe de transition asynchrone. Un triplet de Hoare de la forme $\{P\} p \{Q\}$ est satisfait si et seulement si :

$$\forall \eta \in \zeta, (\eta \models P \implies \exists \eta' \mid (\eta \xrightarrow{p} \eta' \text{ et } \eta' \models Q))$$

4.3.3 Règles d'inférence

Pour vérifier qu'un triplet de Hoare défini syntaxiquement par $\{P\} p \{Q\}$ possède également une définition sémantique, chaque instruction élémentaire $v+$ ou $v-$ est associée à une règle d'inférence définissant une contrainte, qui lorsqu'elle est satisfaite, assure que la transition définie syntaxiquement par $v+$ ou $v-$ appartient au GTA considéré (définition 21).

Définition 21: Axiomes de l'incrémentement et de la décrémentation

Pour toute variable v , les axiomes de l'incrémentement sont définis par :

$$\frac{\{\Phi_v^+ \wedge Q[v \leftarrow v + 1]\}}{v+ \{Q\}}$$

$$\text{où } \Phi_v^+ = \bigwedge_{\omega \in E^{-1}(v)} (\Phi_v^\omega \implies K_{v,\omega} > v) \text{ et } \Phi_v^\omega \equiv \left(\bigwedge_{m \in \omega} \varphi_m \right) \wedge \left(\bigwedge_{m \in (E^{-1}(v) \setminus \omega)} \neg \varphi_m \right)$$

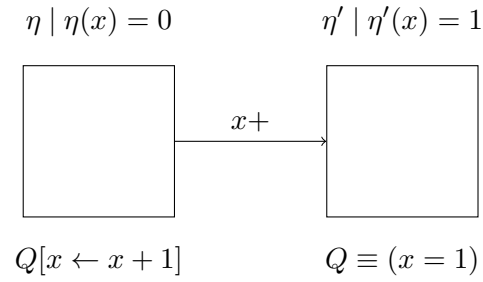
et les axiomes de la décrémentation sont définis par :

$$\frac{\{\Phi_v^- \wedge Q[v \leftarrow v - 1]\}}{v- \{Q\}}$$

$$\text{où } \Phi_v^- = \bigwedge_{\omega \in E^{-1}(v)} (\Phi_v^\omega \implies K_{v,\omega} < v)$$

Φ_v^ω est la formule qui exprime que l'ensemble ω est l'ensemble des ressources de v . Ce dernier est en effet nécessaire car les transitions du GTA sont contrôlées par les paramètres $K_{v,\omega}$, où les ressources représentent un concept crucial. On peut remarquer qu'il existe un unique ensemble ω dans un état η tel que $\eta \models \Phi_v^\omega$, cf. également à la définition d'un paramètre applicable au chapitre 2 page 29. Similairement Φ_v^+ (resp. Φ_v^-) formalise les conditions pour que v puisse augmenter (resp. diminuer). En effet, parmi toutes les formules en conjonction, il n'y en a qu'une dont la prémisse est vraie (celle qui met en jeu le paramètre effectif). Ainsi, la conjonction sera vraie si et seulement si le paramètre effectif fait augmenter (resp. diminuer) la valeur de la variable v , cf. figure 4.6.

D'autres règles sont introduites pour décomposer une séquence d'instructions élémentaires (définition 22) ou pour déduire la plus faible précondition (définition 23).



$$(a) \phi_v^\omega = \{m_3\}$$

$$(b) Q \equiv ((x = 0) \wedge K_{x, \{m_3\}} \geq 1)$$

Figure 4.6 – **Contrainte qui vérifie l'existence de la transition $\eta \rightarrow \eta'$ dans un GTA noté \mathcal{T} associé au GRBM de la figure 4.4.** On a $\eta \models \phi_v^+$ si $\eta \rightarrow \eta' \in \mathcal{T}$. Autrement dit, si la valeur du paramètre de x applicable en η (en l'occurrence $K_{x, \{m_3\}}$) est supérieure à sa valeur courante (en l'occurrence $\eta(x) = 0$), alors x peut s'activer ($x+$).

Définition 22: Règle d'inférence de la composition séquentielle

$$\frac{\{P_1\} p_1 \{P_2\} \quad \{P_2\} p_2 \{Q\}}{\{P_1\} p_1; p_2 \{Q\}}$$

Définition 23: Règle d'inférence du chemin vide

$$\frac{P \Rightarrow Q}{\{P\} \epsilon \{Q\}}$$

En outre, le formalisme de René Thomas contraint par définition les bornes des variables et des paramètres $K_{v, \omega}$. On en déduit l'axiome suivant :

Définition 24: Axiomes des bornes

Pour toute variable v et tout sous-ensemble ω de ses prédécesseurs :

$$(v \geq 0) \wedge (v \leq b_v) \wedge (K_{v, \omega} \geq 0) \wedge (K_{v, \omega} \leq b_v)$$

4.3.4 Calcul de la plus faible précondition

Edsger Dijkstra propose une stratégie de preuve qui consiste à décomposer récursivement le chemin p en instructions élémentaires en partant de la postcondition Q jusqu'à la précondition P , jusqu'à rencontrer la cas de base du chemin vide qui renvoie la plus faible précondition notée wp , [5].

Définition 25: La plus faible précondition wp

Étant donné un GRBM $\mathcal{G} = (V, M, E)$, sa famille des paramètres \mathcal{K} associée et un triplet de Hoare $\{P\} p \{Q\}$, la *plus faible précondition* notée wp est une fonction de p et Q définie inductivement par :

- si $p \equiv \epsilon$ alors $wp(\epsilon, Q) = Q$,
 - si $p \equiv v+$ alors $wp(v+, Q) = Q[v \leftarrow v + 1] \wedge \phi_v^+$,
 - si $p \equiv v-$ alors $wp(v-, Q) = Q[v \leftarrow v - 1] \wedge \phi_v^-$ et
 - si $p \equiv (p_1; e)$ alors $wp((p_1; e), Q) = wp(p_1, wp(e, Q))$,
- où v est une variable appartenant à V et e un évènement de la forme $v+$ ou $v-$.

Remarque 2. Les deuxième et troisième conditions définissent une contrainte sur le domaine des paramètres $K_{v,\omega}$, cf. les axiomes de l'incrémentatation et de la décrémentatation respectivement en définition 21. La dernière condition définit une propagation de la contrainte sur la valeur de $K_{v,\omega}$ de la postcondition Q vers la précondition P , basée sur l'axiome de la composition séquentielle, cf. définition 22.

Remarque 3. L'axiome des bornes n'est pas mentionnée dans la définition 25 puisqu'il est utilisé pour simplifier à la volée la plus faible précondition. En effet, la connaissance des bornes du domaine des paramètres $K_{v,\omega}$ et des variables v est une connaissance à prendre en compte à chaque étape de réécriture de la plus faible précondition.

La stratégie qu'emploie Dijkstra est qualifiée de *backward* et offre deux avantages. D'une part elle implique une unique preuve de correction de H et permet d'autre part de réaliser un simple jeu de réécriture de la postcondition. Un triplet de Hoare est satisfait s'il existe une paramétrisation σ d'un GRBM \mathcal{G} qui satisfait la wp . Ce problème de satisfaction de contrainte une fois résolu permet de réduire l'ensemble des paramétrisations possibles. Il est décidable puisque la complétude de la logique de Hoare génétiquement modifiée a été prouvée dans l'article fondateur [85].

4.4 Gestion de l'exhaustivité des modèles avec TotemBioNet

Cette section a fait l'objet d'un *tool paper* à la conférence internationale CMSB (*Computational Methods in Systems Biology*) en 2020 [96].

TotemBioNet est une plate-forme logicielle qui intègre deux outils : SMBioNet et HoareFol et constitue un outil efficace d'aide à la paramétrisation d'un GRBM. SMBioNet est un outil développé par Adrien Richard qui prend en entrée un GRBM \mathcal{G} et un ensemble de formules CTL Φ , procède à une énumération de toutes les paramétrisations de \mathcal{G} et sélectionne celles qui satisfont la conjonction des éléments de Φ . HoareFol est un outil développé par Maxime Folschette [89] qui implémente en OCaml le calcul de la plus faible précondition déduite d'un triplet de Hoare H . Il génère un programme en Answer Set Programming (ASP) qui est envoyé au solveur Clingo [73] pour calculer les paramétrisations satisfaisant H . TotemBioNet se distingue en combinant ces deux approches de manière à ce que les modélisateurs convergent rapidement vers l'ensemble exhaustif des paramétrisations satisfaisant toutes les connaissances biologiques formalisées.

4.4.1 Entrées et sorties

Les entrées de TotemBioNet sont (représentées en gris en figure 4.7) :

- un GRBM \mathcal{G} ,
- toute connaissance *a priori* sur les valeurs des paramètres de \mathcal{G} ,
- et des propriétés sur la dynamique du système exprimées en CTL, en CTL', ou par un triplet de Hoare.

Toutes les paramétrisations σ de \mathcal{G} qui satisfont les propriétés données en entrée sont renvoyées par TotemBioNet. Nommons l'ensemble de ces paramétrisations Ψ .

Souvent, l'objectif d'une étude de modélisation est de répondre à une question que se pose un biologiste. Mais comment vérifier avec méthode que cette question est bien posée ? Pour paraphraser une citation de Jean-Marc Jancovici : « la vie se résume à des problèmes mal posés qui admettent plusieurs solutions ».

La question essentiellement traitée dans le cadre de modélisation de René Thomas est *de savoir si l'ensemble des régulations formalisées par un GRBM est cohérent avec certaines propriétés dynamiques*. Le GRBM et les propriétés dynamiques encodées en CTL ou à l'aide de la logique de Hoare peuvent inclure des hypothèses à tester. Si l'ensemble des paramétrisations cohérentes biologiquement Ψ est vide alors les hypothèses sont fausses ou d'autres connaissances plus solides bibliographiquement sont remises en question. Si la taille de Ψ est trop grande alors les hypothèses testées revêtent peu d'intérêt puisqu'elles apportent peu de connaissances supplémentaires. Par contre si le sous-ensemble de paramétrisations qui satisfait les hypothèses est relativement plus petit que celui d'origine alors :

1. les hypothèses formalisées par le biologiste méritent d'être testées expérimentalement puisqu'elles ont montré leur pouvoir de réfutation dans un cadre formel,
2. l'apprentissage des valeurs de paramètres renseigne le biologiste sur les règles logiques d'activation d'un ensemble de variables impliquées dans les hypothèses testées formellement.

In fine, TotemBioNet permet de raffiner des questions à tester expérimentalement de sorte de converger plus rapidement vers la mise en évidence de nouvelles connaissances sur le système qui intéresse une équipe de biologistes, donc vers des questions pertinentes d'un point de vue biologique.

4.4.2 Workflow

Si un triplet de Hoare est défini, TotemBioNet appelle HoareFol pour calculer la plus faible précondition. La logique de Hoare génétiquement introduit les quantificateurs existentiels lorsque les conditions expérimentales ne sont pas assez précises pour savoir quelle variable passe son seuil en premier. Par exemple $\exists[(v_1+; v_2-), (v_2-; v_1+)]$ signifie que v_1 s'active à un seuil donné et que v_2 se désactive à un seuil donné mais que l'ordre de ces événements n'est pas connu. En outre, les quantificateurs universels permettent d'abstraire une collection d'observations expérimentales qui doivent toutes être satisfaites (e.g $\forall[(v_1+; v_2-), (v_2-; v_1+)]$). Dans le cas des quantificateurs existentiels, la plus faible précondition calculée est sous une forme disjonctive et est évaluée à la volée, modèle par modèle. Dans le cas des quantificateurs universels, la plus faible précondition calculée est sous une forme conjonctive et sert à réduire de surcroît l'espace des paramétrisations.

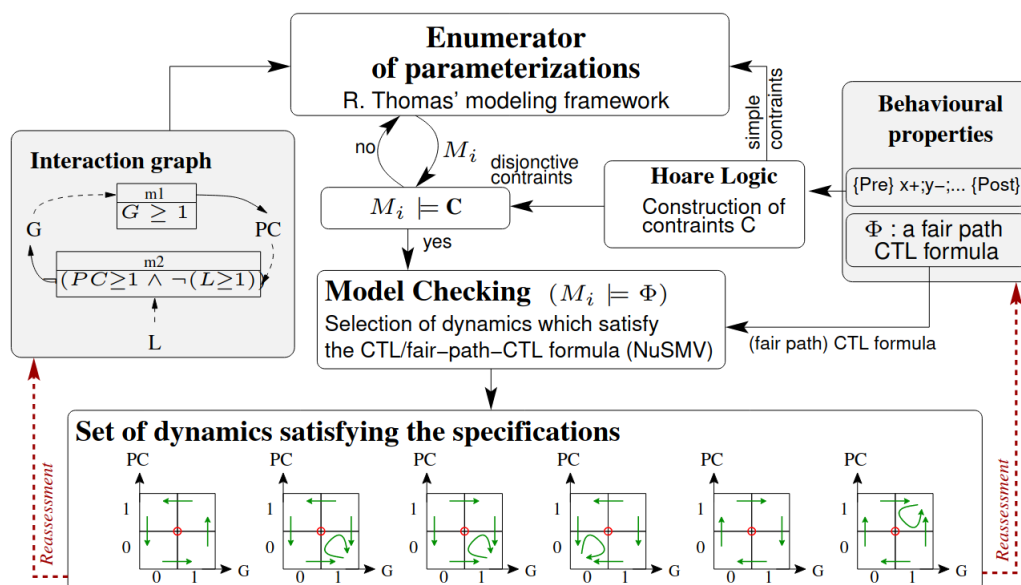


Figure 4.7 – **Fonctionnement de TotemBioNet**. Les entrées sont représentées en gris. L'ensemble des dynamiques satisfaisant les spécifications en entrée sont les sorties de TotemBioNet. Elles conduisent à remettre en question (flèches en pointillés rouge) la formalisation des entrées : le GRBM ou les spécifications des comportements.

TotemBioNet énumère les paramétrisations qui satisfont la condition de Snoussi et la plus faible précondition (sous une forme conjonctive). À partir de ce sous-ensemble de paramétrisations, TotemBioNet construit à la volée le système de transitions, correspondant au GTA déduit du GRBM et d'une paramétrisation en entrée, qu'il envoie à son tour en entrée du model-checker Nu-SMV. L'explosion combinatoire du nombre de paramétrisations déduites de \mathcal{G} mentionnée en page 30 du chapitre 2 est gérée par la résolution de :

- la contrainte de Snoussi (définition 8 page 31),
- la plus faible précondition (définition 25 page 74).

mais aussi par :

- l'exploitation des auto-influences [20],
- et l'expérience de pensée introduite en section 2.4.2.4 page 31.

Si plusieurs formules CTL (équitables) ou plusieurs triplet de Hoare sont définis, la conjonction des spécifications est vérifiée. Enfin, un type particulier de variable peut être déclaré dans le fichier d'entrée de TotemBioNet : *la variable d'environnement*. Elle n'est soumise à aucune influence systémique et ne possède donc aucun prédécesseur. On lui associe un unique paramètre dont la valeur est fixée directement par le modélisateur. Lors de l'exécution d'un algorithme de model-checking, les états qui ne satisfont pas la valeur fixée pour la variable d'environnement sont tout simplement ignorés. Des expériences de perte ou de gain de fonctions peuvent être réalisées en fixant la valeur de la variable d'environnement, respectivement à 0 ou à $n \in \llbracket 1, b_v \rrbracket$, typiquement pour étudier l'impact sur des propriétés d'atteignabilité. Notons par ailleurs que ce type d'expérience *in silico* existe également dans le formalisme des réseaux booléens, voir par exemple le cas d'étude de [97].

La complexité de l'algorithme de model-checking de Nu-SMV est donnée par $\Theta(|\varphi| \times (|\zeta| + |R|))$ où $|\varphi|$ est la taille de la formule CTL testée, $|\zeta|$ le nombre d'états d'un GTA et $|R|$ son nombre de transitions. Toutes ces méthodes permettent de réduire l'espace de recherche des paramétrisations à considérer pour la vérification de modèle.

Le code source de TotemBioNet, maintenu par Hélène Collavizza, ainsi que plusieurs exemples de cas d'utilisation sont disponibles à l'adresse suivante : <https://gitlab.com/totembionet/totembionet>. TotemBioNet s'appuie sur l'éditeur de graphes yEd pour construire le graphe de régulation biologique avec multiplexe. Cet éditeur est disponible à l'adresse suivante : <https://www.yworks.com/products/yed>. Une session typique consiste à construire le graphe d'influence à l'aide de yEd, à générer automatiquement le fichier texte correspondant, puis à ajouter des propriétés temporelles et des triplets de Hoare dans une syntaxe concrète. TotemBioNet génère un fichier de sortie qui contient toutes les paramétrisations, étiquetées avec `OK-model` quand les propriétés dynamiques sont vérifiées, `KO-model` sinon avec toutes les propriétés qui ne sont pas satisfaites [96].

Contributions de la thèse

Les trois chapitres qui suivent font l'objet d'un article en préparation (à soumettre en 2022) :
« *Une formalisation purement discrète et la vérification formelle des checkpoints intrinsèques du cycle cellulaire : une preuve de concept.* ».

-
Déborah Boyenval, Gilles Bernot, Jean-Paul Comet et Franck Delaunay.

CHAPITRE 5

Notre ensemble de modèles pour l'étude des phases du cycle cellulaire

| | | |
|------------|---|------------|
| 5.1 | Abstraction et question de modélisation | 83 |
| 5.2 | Les variables et leurs seuils discrets | 84 |
| 5.3 | Nos conditions initiales | 86 |
| 5.4 | Formalisation des régulations impliquées dans la progression du cycle cellulaire | 88 |
| 5.4.1 | Notre graphe d'influence conçu pour la définition d'un cycle cellulaire canonique | 88 |
| 5.4.2 | Une séquence canonique d'évènements de régulation au cours des phases du cycle cellulaire | 92 |
| 5.4.2.1 | Les évènements requis en G1 | 93 |
| 5.4.2.2 | Les évènements requis en S | 95 |
| 5.4.2.3 | Les évènements requis en G2 | 95 |
| 5.4.2.4 | Les évènements requis en M | 96 |
| 5.4.2.5 | La séquence canonique résultante | 97 |
| 5.4.3 | Définition des multiplexes | 98 |
| 5.4.3.1 | Multiplexes et abstraction mécanistique | 98 |
| 5.4.3.2 | Multiplexes et abstraction phénoménologique | 99 |
| 5.5 | Première identification des valeurs des paramètres | 102 |
| 5.5.1 | Vérification de la stabilité de la phase de quiescence | 103 |
| 5.5.2 | Formalisation d'un cycle cellulaire canonique comme un attracteur cyclique | 104 |
| 5.5.3 | Se poser les bonnes questions sur les checkpoints à l'aide de TotemBioNet | 104 |
| 5.5.4 | Identification manuelle de certains paramètres | 106 |

5.1 Abstraction et question de modélisation

Une étape de modélisation est nécessaire au préalable pour valider une formalisation des checkpoints intrinsèques du cycle cellulaire en tant que concept. La preuve de concept se fait sur un ensemble non vide de modèles qui formalisent les principales voies de régulations impliquées dans le contrôle du cycle cellulaire. Un GRBM est d’abord construit pour représenter ces voies (sections 5.2 à 5.4). Plusieurs propriétés classiques du cycle cellulaire sont ensuite définies pour identifier les paramétrisations de ce GRBM cohérentes biologiquement, qualificatif défini formellement dans le chapitre 4. On se focalise en section 5.5 sur deux propriétés : un certain ordonnancement dans la mise à jour des variables du GRBM et un contrôle du cycle cellulaire par les facteurs de croissance. Si l’ensemble des modèles (où modèle est un couple formé par un GRBM et une de ses paramétrisations) qui satisfont ces propriétés est non vide alors la preuve de concept est réalisée avec succès. Cet ensemble de modèles est identifié en combinant de manière adéquate CTL et la logique de Hoare génétiquement modifiée, comme expliqué à la fin du chapitre précédent.

Les composants clés de la régulation du cycle cellulaire sont les complexes *cyclin/cyclin-dependant kinases* (abrégiés *cyc/cdk*) et leurs principaux inhibiteurs. Une question de modélisation doit être définie pour que notre GRBM soit minimal, au sens où aucune variable et aucun seuil n’est inutilement défini. *Quelles sont les évènements requis lors de chaque phase et quels sont les régulations qui permettent à ces évènements de se réaliser dans un certains ordre ?* Si une variable ou un seuil n’intervient pas dans la caractérisation de ces évènements alors elle est inutile.

John Tyson et Bela Novak proposèrent en 2008 un modèle générique capable de reproduire l’organisation temporelle du cycle cellulaire [40]. Jonathan Behaegel enrichit ce modèle en 2016 et définit un triplet de Hoare caractéristique du cycle cellulaire, de la forme $\{precondition\} \text{ chemin } \{postcondition\}$ [69]. Le *chemin* est une séquence d’évènements de la logique de Hoare génétiquement modifiée (abrégiée LHGM). Des informations supplémentaires sont données au chapitre 3 page 50. Cette séquence est découpée en sous-séquences propres à chaque phase qui mettent en exergue les évènements requis lors d’une phase donnée.

Nous procédons dans ce chapitre à une mise à jour conséquente du modèle de [69] vis à vis, d’une part, des connaissances et du recul apportés depuis 2016 et, d’autre part, d’une spécialisation pour les mammifères. Cinq variables, trois multivaluées et deux booléennes, suffisent à représenter une séquence des évènements au cours du cycle. Une fois cet ensemble minimal de variables munies de leur seuil défini, l’ensemble exhaustif des voies de régulations entre ces variables est recensé et formalisé sous la forme d’un GRBM. Le lecteur pourra se reporter à la figure 5.18 pour avoir à tout moment une vision globale qui permet de situer les différents détails présentés le long du chapitre.

48 paramètres gouvernent sa dynamique et un ensemble de plus de 10^{20} modèles est admis par notre GRBM, ce qui met pour l’instant hors de portée d’un point de vue calculatoire nos preuves de concept des chapitres 6 et 7. Dans le chapitre précédent, nous avons démontré que TotemBioNet a la capacité de réduire l’explosion combinatoire du nombre de modèles en exploitant des connaissances *a priori* sur la dynamique globale du système étudié. Ces connaissances ne manquent pas tant les travaux gravitant autour du cycle cellulaire et de sa régulation s’accumulent ; la section 5.5 conclut que seulement 16 modèles ont de l’intérêt pour l’étude des checkpoints.

5.2 Les variables et leurs seuils discrets

Les six variables de notre GRBM représentent une molécule, ou un groupe de molécules qui agissent de la même manière sur leurs cibles ou bien des voies moléculaires à part entières.

| Variable | Groupe de molécules | Entité biologique | Domaine |
|-----------|------------------------|----------------------------------|------------|
| <i>sk</i> | « Starting kinase » | cycE/cdk2, CAK | $[[0, 2]]$ |
| <i>a</i> | - | cycA/cdk1, cycA/cdk2, CAK | $[[0, 2]]$ |
| <i>b</i> | - | cycB/cdk1, CAK | $[[0, 2]]$ |
| <i>en</i> | « Enemies » | APC-cdh1, P21, P27, PP1/2A | $[[0, 1]]$ |
| <i>ep</i> | « Exit protein » | APC-cdc20 | $[[0, 1]]$ |
| <i>gf</i> | Facteurs de croissance | EGF, FGF, PDGF, IGF, TGF β | $[[0, 1]]$ |

Table 5.1 – Les variables de notre GRBM et leur domaine. Toutes nos variables, excepté *gf*, s'inspirent des notations de [40].

Le cycle cellulaire est une succession d'évènements cellulaires, *e.g* à gros grain : le passage du point de restriction, l'activation de la machinerie de réplication ou encore les protéines impliquées dans la séparation des chromatides soeurs au cours de la mitose. Des décennies de travaux expérimentaux [14] mettent en évidence que ces évènements sont sous le contrôle des *cyc/cdk*. Les cdk sont des enzymes qui phosphorylent les substrats déterminants dans la réalisation des différents évènements de la réplication et de la mitose. D'après [80], les cdk 2 et 1 initient respectivement les processus de la phase S et M. Les cdk 4 et 6 régulent l'entrée d'une cellule dans le cycle cellulaire depuis la phase de quiescence en réponse aux facteurs de croissances du milieu extracellulaire.

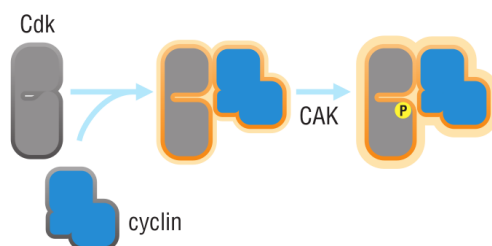


Figure 5.1 – L'activation complète des complexes *cyc/cdk* nécessite une complexation avec les CAK. Les variables *sk*, *a* et *b* incluent donc les CAK. Figure extraite de [32].

Les cdk 1,2,4 et 6 ne sont actives que lorsqu'elles sont liées à deux sous-unités activatrices : les *cyclines* et les *cyclin-activating-kinases*, abrégées CAK, *cf.* figure 5.1. Les CAK correspondent aux complexes *cycH/cdk7* et phosphorylent les cdk 4,6,2 et 1 ce qui leur permet d'adopter une conformation qui améliore la fixation à leurs substrats. Les cyclines ne sont pas simplement des activateurs de leur cdk associée, elles aident à la diriger vers des substrats spécifiques, soit en liant directement le substrat, soit en l'amenant vers un compartiment sub-cellulaire où se trouve le substrat.

Les complexes *cycD/cdk4* et *cycD/cdk6* sont impliqués dans la transition entre la phase de

quiescence et la phase G1, elles ne sont pas explicitement représentés dans le modèle original. Rendre explicite ces complexe sous-entendrait que la question de modélisation se focalise au moins en partie sur les déterminants moléculaires de l'entrée dans le cycle cellulaire, ce qui n'est pas notre cas.

Le complexe *cycE/cdk2* est un déterminant du passage du point de restriction et est impliqué dans l'initiation du processus de réplication (phase S). Il est activé en phase G1 et affiche un pic d'activité juste avant la transition G1/S [101]. Ce complexe est judicieusement nommé *starting kinase* (variable *sk*) par [40] et nous garderons sa nomenclature le long de la thèse.

Les complexes *cycA/cdk2* et *cycA/cdk1* sont respectivement activés en phase S et G2. Ils sont regroupés sous le nom de variable « *a* ». Le complexe *cycB/cdk1* (« *b* » dans la table 5.1) est préalablement activé en fin de G2 et son activité augmente au moment de l'entrée en phase M.

La variable booléenne *gf* sort du lot : on la considère comme une variable d'*environnement* sous le contrôle total du modélisateur, à l'image de l'expérimentaliste qui contrôle les conditions de son expérience. Les facteurs de croissance permettront de vérifier le contrôle de l'entrée dans le cycle cellulaire : une cellule doit être sensible aux facteurs de croissance pour être arrêtable en phase de quiescence.

Les complexes *cyc/cdk* sont maintenus inactifs par une ubiquitine ligase (APC-cdh1), des phosphatases (PP1/2A) ou des protéines qui forment une complexation inhibitrice (p21 et p27). Ces inhibiteurs sont regroupés sous le terme de *enemies* en table 5.1. APC-cdc20 est également une ubiquitine ligase qui marque pour la dégradation certains complexes *cyc/cdk* mais est défini par une variable à part entière, étant donné son rôle prépondérant dans la sortie de mitose (*exit protein* en table 5.1). *en* et *ep* sont des variables booléennes.

La variable *sk* possède deux seuils : $sk = 0$ signifie que *sk* agit sur aucune cible. $sk = 1$ abstrait l'intervalle de concentration de *sk* correspondant à une première accumulation du complexe *cycE/cdk2* qui devient capable d'activer indirectement sa propre transcription ($sk \rightarrow sk$) et d'inhiber APC-cdh1, p21 et p27 ($sk \rightarrow en$). Lorsque $sk \geq 2$, *cycE/cdk2* est pleinement activé et marque la transition G1/S. *sk* cible également à ce seuil le complexe *cycA/cdk1-2* ($sk \rightarrow a$).

La variable *a* abstrait à la fois les complexes *cycA/cdk2* et *cycA/cdk1* : $a = 0$ est le cas trivial où ces complexes n'agissent sur aucune cible. $a = 1$ représente spécifiquement l'action du complexe *cycA/cdk2* en phase S. À ce seuil, *a* active *sk* et inhibe APC-cdh1, p21 et p27 (*en*). $a = 2$ représente en plus l'action du complexe *cycA/cdk1*, qui s'active plus tardivement que le complexe *cycA/cdk2*, en phase G2. Il réprime également l'activité de *en* mais active en plus le complexe *cycB/cdk1* (*b*). Comme $a = 2$ représente la présence des deux complexes *cycA/cdk2* et *cycA/cdk1* à la fois, cette abstraction ne peut décrire l'action de *cycA/cdk1* sans décrire simultanément l'action de *cycA/cdk2*, alors que $a = 1$ décrit uniquement l'action de *cycA/cdk2* (figure 5.2). Notre abstraction est accompagnée de l'hypothèse de modélisation suivante : la façon dont est définie *a* implique que *cycA/cdk1* active *sk* alors qu'à notre connaissance aucune interaction directe ou indirecte a été recensée entre *cycA/cdk1* et *cycE/cdk2*. Quelles qu'en soient les raisons, si la variable *a* venait à être scindée pour différencier les actions de *cycA/cdk2* et *cycA/cdk1*, l'arc $cycA/cdk1 \rightarrow sk$ doit être considéré avec la plus grande prudence.

La variable *b* est associée à trois niveaux de concentration. Lorsque $b = 1$, le complexe *cycB/cdk1* est partiellement activé et n'est capable d'agir indirectement que sur sa propre activation ($b \rightarrow b$). À ce seuil, *b* est capable de prendre le relais de *sk* et *a* pour l'inhibition de APC-cdh1, p21 et p27 ($b \rightarrow en$). Le second seuil de *b* correspond à l'accumulation de la cycline B suffisante pour activer l'initiation de la mitose et APC-cdc20 ($b \rightarrow ep$), une protéine quant à elle impliquée dans la sortie

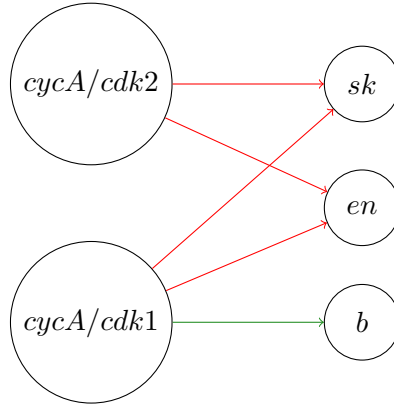


Figure 5.2 – Abstraction de *cycA/cdk2* et *cycA/cdk1* au sein d'une unique variable.

de mitose.

En section 5.4.2, on démontre que tous ces seuils permettent de définir l'action de chacune de ces variables sur des évènements caractéristiques des phases du cycle cellulaire et par extension les évènements de régulation requis en G1, S, G2 et M.

5.3 Nos conditions initiales

Les définitions des variables et seuils de notre GRBM sont suffisantes pour caractériser les conditions initiales de notre système. Avant même de définir les régulations entre nos variables, la définition de ces conditions initiales met déjà en évidence que toutes les propriétés attendues seront observables :

1. Une propriété de quiescence stable quand les facteurs de croissance sont retirés stablement du milieu extracellulaire.
2. Un engagement perpétuel dans un cycle cellulaire lorsque les facteurs de croissance sont constamment actifs.

Une cellule en phase de quiescence (notée G0) est décrite canoniquement par une inhibition des complexes moléculaires nécessaires à l'initiation de la réplication et de la mitose (à savoir *sk*, *a* et *b*) par APC-cdh1 (*en*) [99]. La protéine APC-cdc20 (*ep*), cruciale pour la sortie de mitose, est inactive en G0. On obtient une première condition initiale caractérisée par :

$$sk = 0 \wedge a = 0 \wedge b = 0 \wedge en = 1 \wedge ep = 0 \quad (5.1)$$

L'ensemble de nos variables admet une *variable d'environnement* : les facteurs de croissance. *gf* est une variable booléenne sous le contrôle total du modélisateur donc deux conditions initiales du cycle cellulaire sont à considérer. Soit les facteurs de croissance sont inactifs (*gf* = 0) et la cellule est en phase de quiescence.

$$G0 \equiv sk = 0 \wedge a = 0 \wedge b = 0 \wedge en = 1 \wedge ep = 0 \wedge gf = 0 \quad (5.2)$$

Soit la concentration en facteurs de croissance est suffisante pour potentiellement activer le complexe *cycE/cdk2* (*sk*) au travers de la voie MEK/ERK ($gf = 1$) et la cellule est en G1 :

$$G1_{init} \equiv sk = 0 \wedge a = 0 \wedge b = 0 \wedge en = 1 \wedge ep = 0 \wedge gf = 1 \quad (5.3)$$

Plus précisément, $gf = 1$ signifie que les complexes *cycD/cdk4-6* peuvent activer *sk* en réprimant les effets inhibiteurs des membres de la famille pRB, et fournissent par la même occasion une potentielle ressource pour les gènes nécessaires à la réplication et la mitose. Plus de détails seront donnés en section 5.4.2. Puisque ces formules décrivent l'état de chacune de nos six variables, elles caractérisent un état au niveau systémique (définition 5 page 29). $G1_{init}$ décrit l'état le plus précoce de G1 que nous sommes en mesure de définir.

Grâce à gf , l'espace des états de notre GRBM est partitionné en deux parties et chacune des parties admet un unique attracteur. La première partie unit les états qui satisfont la formule

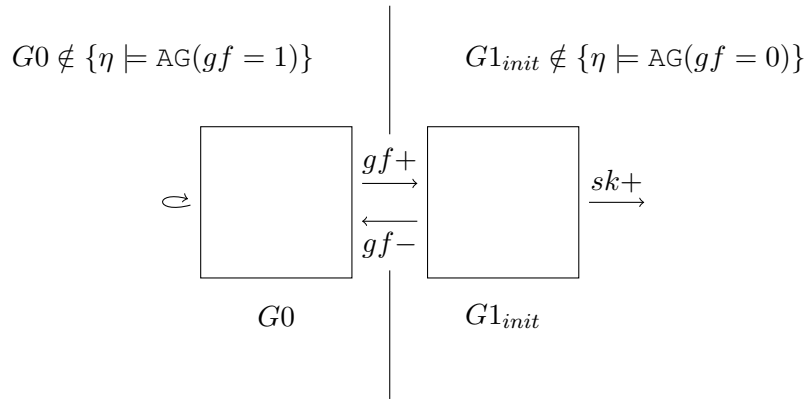


Figure 5.3 – **Deux contextes environnementaux donc deux états initiaux.** Chaque état initial est supposé appartenir à un unique attracteur.

CTL $AG(gf = 0)$, qui signifie littéralement que gf est constamment inactif. Cette partie doit admettre un unique attracteur : l'état $G0$ stable. En effet, tant que le milieu extracellulaire n'est pas exposé à un niveau de facteurs de croissance suffisant, la cellule doit restée bloquée en phase de quiescence. Par ailleurs en pratique, les cellules en cultures *in vitro* sont synchronisables en $G0$ (celle de notre définition) par déprivation de sérum bovin foetal, qui est une solution fortement concentrée en facteurs de croissance (*cf.* la liste en table 5.1). La seconde partie unit cette fois les états qui satisfont la formule CTL $AG(gf = 1)$. $G1_{init}$ est l'état initial du cycle cellulaire qui est formellement considéré comme un attracteur complexe, à la manière de [87, 86]. Il est même considéré comme l'unique attracteur de cette partition où gf est constitutivement actif.

Une cellule qui n'est pas en quiescence est une cellule en prolifération : elle s'engage à effectuer tous les évènements requis au cours du cycle cellulaire. En figure 5.3, $sk+$ marque le premier évènement du cycle cellulaire que l'on est capable de formaliser et marque surtout une transition quiescence/prolifération, communément nommé le *point de restriction*. On explique en section 5.4.2 ce qui permet à sk de s'activer en $G1_{init}$ et pourquoi $sk+$ est annoté comme le passage du point de restriction. [23, 80] décrivent le point de restriction comme un point de non-retour du cycle cellulaire qui dépend de la concentration en facteurs de croissance. Si les

facteurs de croissance sont retirés du milieu extracellulaire avant le point de restriction, la cellule s'arrête immédiatement en G0 ou en début de G1. S'ils sont retirés après le point de restriction, la cellule passe par G1, S, G2, M et s'arrête en phase G0 ou début de G1 du cycle cellulaire suivant. La formule 5.2 se distingue de la formule 5.3 uniquement par la valeur de gf . Le passage de $G0/G1_{init}$ dépend de gf et est complètement réversible (pas d'hystérésis), contrairement au passage de la quiescence à la prolifération qui est irréversible (hystérésis).

5.4 Formalisation des régulations impliquées dans la progression du cycle cellulaire

La section s'organise comme suit :

- Les influences qu'exercent nos variables entre elles sont d'abord décrites au plus haut niveau d'abstraction à l'aide d'un graphe d'influence où un noeud est une variable et un arc représente une action positive (étiqueté +) ou négative (étiqueté -) sur le taux de synthèse de la variable cible (sous-section 5.4.1).
- Une séquence canonique des événements de régulations impliquant nos variables est ensuite définie à l'aide des conditions initiales et du graphe d'influence (sous-section 5.4.2).
- Un graphe de régulation avec multiplexe est ensuite déduit du graphe d'influence et multiplexes phénoménologiques (sous-section 5.4.3).

5.4.1 Notre graphe d'influence conçu pour la définition d'un cycle cellulaire canonique

La régulation de la progression du cycle cellulaire est décrite par un réseau complexe de multiples boucles de rétroactions positives et négatives. Dans le chapitre 3, nous avons mis l'accent sur certaines études qui sont remarquables dans leur démarche d'abstraction et qui ont réussi à identifier des modules caractéristiques de certaines phases du cycle cellulaire ou de certaines transitions de phases [80, 82, 72]. Ces réseaux ont une structure particulière : ils contiennent une boucle de rétroaction positive jouant le rôle de switch bistable. Nous décrivons en chapitre 3 le rôle crucial des switches multi-stables dans la séparation temporelle des phases et l'unidirectionnalité du cycle cellulaire. Nous décrivons notre graphe d'influence de la progression du cycle cellulaire de façon modulaire (introduit en figure 5.12). Soit le module est impliqué dans une phase donnée et représente donc une boucle positive, soit il est impliqué dans une transition de phases et dans ce cas est représenté structurellement par une boucle de rétroaction négative.

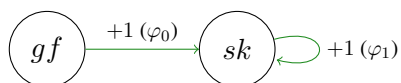


Figure 5.4 – **Boucle positive impliquée dans le passage du point de restriction qui est situé canoniquement en G1.** La boucle sur sk permet au signal d'activation de gf de subsister même si les facteurs de croissance sont retirés du milieu extra-cellulaire.

L'arc activateur $gf \rightarrow sk$ abstrait la chaîne activatrice partant des facteurs de croissance (nommés *mitogens* dans la figure 5.5) jusqu'au complexe *cycE/cdk2*. La formule $\varphi_0 \equiv gf \geq 1$ est un

littéral du langage des multiplexes qui représente bien plus que la présence de gf , elle sous-entend également l'activation de la voie MEK/ERK, la présence du complexe $cycD/cdk4-6$, l'absence de Rb et la présence de E2F. Les facteurs de croissance activent, au travers de la voie MEK/ERK,

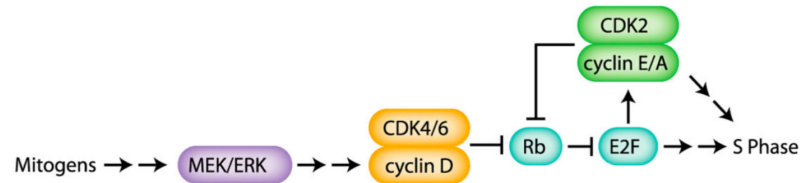


Figure 5.5 – **Schéma simplifié du paradigme dominant autour de la régulation du passage du point de restriction.** Figure adaptée de [62].

la transcription de la cycline D qui, une fois associée aux $cdk4-6$, lève indirectement l'inhibition du facteur de transcription EF2 par la protéine Rb. Le complexe $cycD/cdk4-6$ phosphoryle la protéine Rb ce qui dissocie partiellement E2F du complexe qu'il formait avec Rb. E2F active la transcription du gène codant pour Cdc25, une phosphatase activatrice du complexe $cycE/cdk2$, et de la cycline E. Dans cette configuration, E2F est en quantité suffisante pour activer le complexe $cycE/cdk2$ qui hyperphosphoryle en retour la protéine Rb, créant ainsi une boucle d'activation de $cycE/cdk2$ de la figure 5.4. Lorsque la boucle positive $cycE/cdk2 \rightarrow Rb \rightarrow E2F \rightarrow cycE/cdk2$ est effective, une cellule devient insensible à la privation de facteurs de croissance étant donné que cette boucle est suffisante pour maintenir le complexe $cycE/cdk2$ actif (au seuil 1) [101]. Le passage du point de restriction est donc associé à l'activation du complexe $cycE/cdk2$ au seuil 1. $gf \rightarrow sk$ est un exemple typique d'abstraction d'une voie de régulation qui se base sur la relation de transitivité :

$$\frac{\frac{gf \rightarrow MEK/ERK \quad MEK/ERK \rightarrow cycD/cdk4-6}{gf \rightarrow cycD/cdk4-6} \quad \frac{cycD/cdk4-6 \rightarrow Rb}{gf \rightarrow Rb} \quad \frac{Rb \rightarrow E2F}{gf \rightarrow E2F} \quad \frac{E2F \rightarrow sk}{gf \rightarrow sk}}$$

Cet arbre explicite des règles qui font référence à des interactions positives (en vert) ou négatives (en rouge) qui doivent être corroborées par des travaux expérimentaux. D'autres voies se greffent à celles de la figure 5.5, on renvoie particulièrement à [37] qui propose une carte détaillée et curée de la régulation du cycle cellulaire par RB/E2F. La cohérence de notre abstraction a été vérifiée à la volée en consultant une des bases de données les plus documentées, nettoyées et appliquées aux modèles mammifères : Reactome*. L'idée est de vérifier à la main qu'aucune régulation spécifiée dans la base de donnée entre les variables de notre modèle ne vient réfuter notre abstraction des arcs activateurs ou inhibiteurs. Par exemple, si un arc activateur a été défini entre x et y , nous vérifions qu'il existe bien un nombre pair (resp. impair) d'arcs inhibiteurs abstraits dans $x \xrightarrow{+} y$ (resp. $x \xrightarrow{-} y$).

La fin de la phase G1 est marquée par la levée de l'inhibition de $cycE/cdk2$ (sk) par APC-cdh1, ce qui nous donne l'occasion de décrire un des mécanismes qui sous-tendent les arcs allant de en vers sk , a ou b , figure 5.6. Les cyclines E, A et B (par extension les variables sk , a et b) sont ubiquitinées par APC-cdh1 ce qui entraîne leur dégradation par le protéasome. Ce mécanisme de

*. <https://reactome.org/>

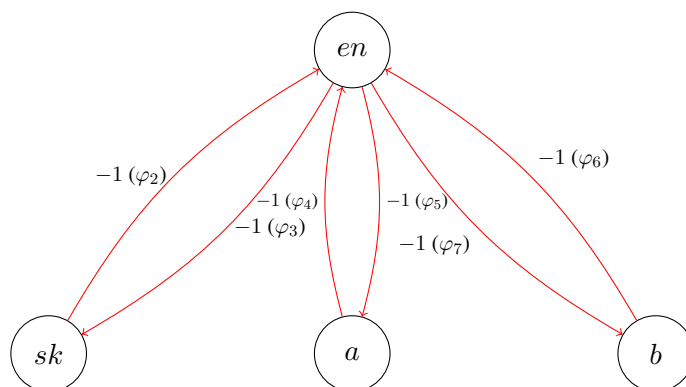


Figure 5.6 – **Inhibitions réciproques entre les cyclines E, A et B et les protéines APC-cdh1, p21 et p27.**

dégradation, que l'on retrouvera plus bas, est appelé la destruction protéolytique. L'activité inhibitrice d'APC-cdh1 est réprimée par la phosphorylation de cdh1 qui est catalysée par sk , a et b ce qui crée une boucle de rétroaction positive entre les complexes cyc/cdk et leurs inhibiteurs en . Plus précisément, cette phosphorylation de cdh1 provoque la dégradation protéolytique de APC-cdh1 prise en charge par l'ubiquitine ligase SCF (« l'ubiquitine ubiquitinée »). La variable en abstrait également les *Cdk Inhibitor Proteins* (CKI) qui ont la particularité d'inhiber le site actif des complexes cyc/cdk en se liant directement à eux. Parmi eux figurent les protéines p27 et p21. Elles sont également la cible des complexes cyc/cdk de notre modèle, et s'en suit le même processus de dégradation que celui d'APC-cdh1. Ces protéines sont déterminantes également dans le processus de blocage du cycle cellulaire en réponse à des lésions de l'ADN, processus (nommé checkpoint extrinsèque dans le chapitre 3) qui n'entre cependant pas dans notre question de modélisation [36].

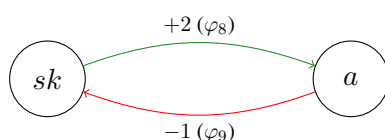


Figure 5.7 – **Boucle de rétroaction négative impliquée dans la transition G1/S.**

Une fois le complexe $cycE/cdk2$ pleinement activé ($sk = 2$), il active sa propre destruction via le complexe $cycA/cdk2$ (a). sk inhibe au seuil 2 la protéine pRb et de ce fait libère une version de la protéine E2F qui active spécifiquement la transcription de la cycline A ($sk \rightarrow a$) [19, 58] (figure 5.4). Elle s'associe ensuite à la cdk2 pour former le complexe $cycA/cdk2$ et peut également se lier à cdk1 pour former le complexe $cycA/cdk1$. Le complexe $cdk2/cycA$ catalyse la phosphorylation inhibitrice de E2F ce qui a pour conséquence d'inhiber par extension la transcription de la cycline E et de cdk2 ($a \rightarrow sk$). L'inactivation de sk permet d'inhiber un évènement requis en G1 : la formation des complexes de pré-réplication aux origines de réplication qui est décrite en sous-section 5.4.2.

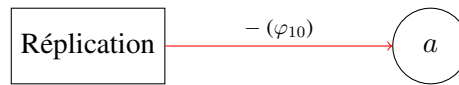


Figure 5.8 – **La transition S/G2 est contrôlée par l’activité répressive de la voie ETAA1-ATR-CHK1 sur l’axe cycline A-CDK1-FOXM1 jusqu’à la fin de la réplication.** La variable a est sensible au stress endogène provoqué par le processus de réplication.

Le processus de réplication influence négativement un régulateur du cycle cellulaire : le complexe *cycA/cdk1*. Cette inhibition passe par la protéine ATR. Elle est connue principalement pour son rôle déterminant dans la réponse aux dommages à l’ADN. Néanmoins, une étude récente conclut à un rôle dans la coordination de la réplication et de la mitose en absence de dommages à l’ADN [84]. Les auteurs regroupent d’ailleurs ces mécanismes impliqués dans cette coordination sous le terme de checkpoint intrinsèque S/G2. ATR joue le rôle de senseur de terminaison de la réplication et assure que la cellule n’active pas des gènes impliqués dans la phase G2 ou M avant la fin de la réplication. En section 5.4.2, on montre justement que *cycA/cdk1* active l’expression de gènes impliqués en mitose. Lorsque la réplication débute mais tant qu’elle n’est pas terminée, l’ADN simple brin est recouvert par la protéine RPA qui recrute ATR et inhibe subséquemment l’activité du complexe *cycA/cdk1* par phosphorylation. Cette inhibition passe par l’activation de Chk1, un inhibiteur du complexe *cycA/cdk1*. ATR inhibe des facteurs pro-mitotiques suivant deux voies. ATR inhibe FOXM1 qui est pour la phase M ce que E2F est pour la phase S. FOXM1 active des facteurs pro-mitotiques comme Plk1 et Aurora B, des protéines déterminantes dans le *Spindle Assembly Checkpoint* (SAC), et active également la transcription de la cycline B [24]. L’activité d’ATR chute brutalement au moment de la transition S/G2, c’est à dire à la fin de la réplication. Il en résulte la levée de l’inhibition du complexe *cycA/cdk1* médiée par ATR.

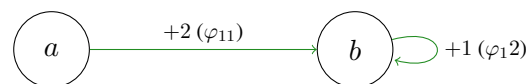


Figure 5.9 – **Une boucle de rétroaction positive impliquée dans l’activation de b au tout début de la phase M.**

Le complexe *cycA/cdk1* hyperphosphoryle la protéine FOXM1, comme décrit précédemment, ce qui a pour conséquence d’activer la transcription de la cycline B, Plk1 et des différents isoformes de la phosphatase Cdc25 (Cdc25A, Cdc25B et Cdc25C). Une fois le complexe *cycB/cdk1* formé, il renforce lui-même son activation ($b \rightarrow b$) au travers de deux mécanismes, décrits de manière concise dans la revue [43]. Brièvement, *cycB/cdk1* active la protéine Plk1 qui à son tour active FOXM1 entraînant une accumulation de la cycline B au niveau du centrosome, l’endroit où se trouve les substrats mitotiques de b . Le complexe *cycB/cdk1* est maintenu partiellement activé par la phosphorylation inhibitrice de Wee1 et Myt1, deux protéines non représentées explicitement dans notre modèle sachant que l’on suppose que l’activation abrupte de b suffit à décrire l’activation de la mitose, cf. la sous-section 5.4.2.4. Le complexe *cycB/cdk1* lève son inhibition par Wee1 et Myt1 en les phosphorylant directement et en activant les protéines Cdc25, les phosphatases antagonistes de Wee1 et Myt1.

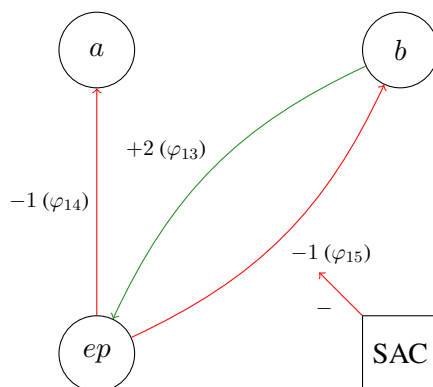


Figure 5.10 – APC-cdc20 (*en*) et la dégradation des cyclines A et B impliquées dans le processus de mitose. La dégradation de la cycline B est dépendante du SAC. La variable *ep* est sensible au SAC.

La protéine Cdc20 ne peut se lier à APC que lorsque cette dernière a préalablement été phosphorylée par le complexe Cdk1/CyclineB ($b \rightarrow ep$) [48]. APC-cdc20 active l'anaphase, c'est à dire la séparation des chromosomes en chromatides soeurs, voir figure 5.11. Une fois le processus de mitose enclenché, le *Spindle Assembly Checkpoint* (SAC) bloque l'activité d' APC-cdc20 tant que les chromosomes ne sont pas tous attachés correctement au fuseau mitotique au niveau des kinétochores. APC-cdc20 compte parmi ses cibles la sécurine, une protéine qui inhibe la protéine responsable de la séparation des chromosomes en chromatides soeurs. Le rôle du SAC est d'inhiber APC-cdc20 (donc l'anaphase) tant qu'il existe au moins un chromosome non attaché au fuseau. Le SAC inhibe l'action de *ep* sur *b* et la sécurine. En bref, Mad2 se fixe au kinétochores libres pour former un complexe qui séquestre Cdc20. Une fois que les kinétochores sont tous attachés aux microtubules, la protéine Mad2 libère Cdc20 et peut s'associer à APC pour agir sur la cycline B ($ep \rightarrow b$) et la sécurine. La cycline A échappe quant à elle à la répression de sa destruction par *ep*, ce qui explique que le SAC ne cible pas l'arc inhibiteur $ep \rightarrow a$. Les justifications sont données au moment de la définition de la trace canonique du cycle cellulaire où la destruction de *a* précède celle de *b* en phase M, cf. page 97.

APC-cdc20 est également une ubiquitine ligase qui compte parmi ses substrats la cycline A ($ep \rightarrow a$), la cycline B ($ep \rightarrow b$), mais pas la cycline E [17] (absence de l'arc $ep \rightarrow sk$). APC-cdh1 cible cdc20 et exerce donc une action inhibitrice sur le complexe APC-cdc20 ($en \rightarrow ep$).

5.4.2 Une séquence canonique d'évènements de régulation au cours des phases du cycle cellulaire

Nos cinq variables systémiques régulent les processus de la réplication et de la mitose. Ces influences ne sont pas représentées dans notre graphe de la figure 2.3. Une description des régulations unidirectionnelles allant de *sk*, *a*, *b*, *en* et *ep* vers les G1, S, G2 et M permet d'associer à chaque phase un ensemble des évènements requis. Ces évènements sont représentés par les atomes du langage des assertions (définition 16 page 70). Ensuite, une séquence canonique de ces évènements est définie à partir de notre état initial du cycle cellulaire $G1_{init}$.

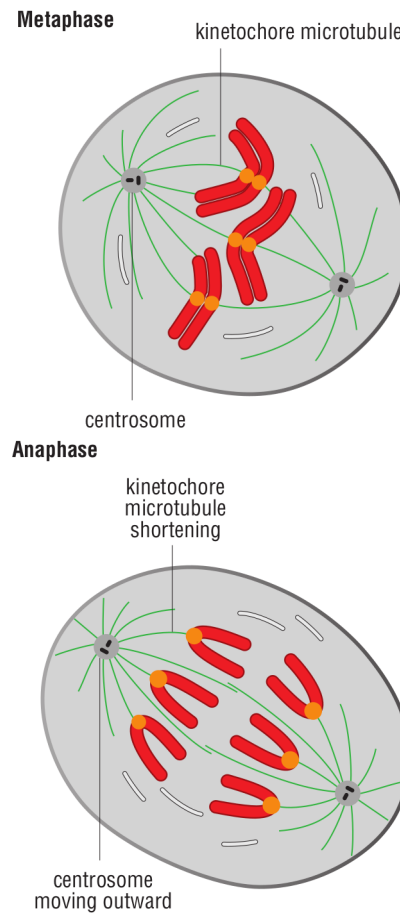


Figure 5.11 – **Métaphase et anaphase.** Au cours de la mitose, le cytosquelette d'une cellule eucaryote est structuré sous la forme d'un fuseau mitotique composé de microtubules (en vert). Ils sont capables de se lier à chaque chromatide soeur d'un chromosome (en rouge) au niveau d'une structure appelée un kinétochore (en orange) au cours de la métaphase. La transition métaphase/anaphase est marquée par la séparation des chromatides soeurs qui à la fin de l'anaphase migrent vers des pôles opposés de la cellule. Figure adaptée de [32].

5.4.2.1 Les évènements requis en G1

L'évènement $gf+$: Les facteurs de croissance inhibent indirectement la protéine Rb (figure 5.5), une protéine qui réprime E2F donc, par extension, les gènes nécessaires à la progression en phase S. En effet, E2F active la transcription des gènes impliqués dans la formation du complexe de pré-réplication (*e.g.* ceux codant pour *cdc6* et *cdt1*) [41].

Le premier évènement $sk+$ ($sk = 0 \rightarrow sk = 1$) : Il est annoté par le passage du point de restriction en se basant principalement sur le modèle canonique proposé par [101] (figure 5.5). sk complète la phosphorylation de Rb, initialement réalisée par le complexe *cycD/cdk4-6*, ce qui entraîne une levée d'inhibition drastique des gènes de la phase S sous contrôle de E2F [21].

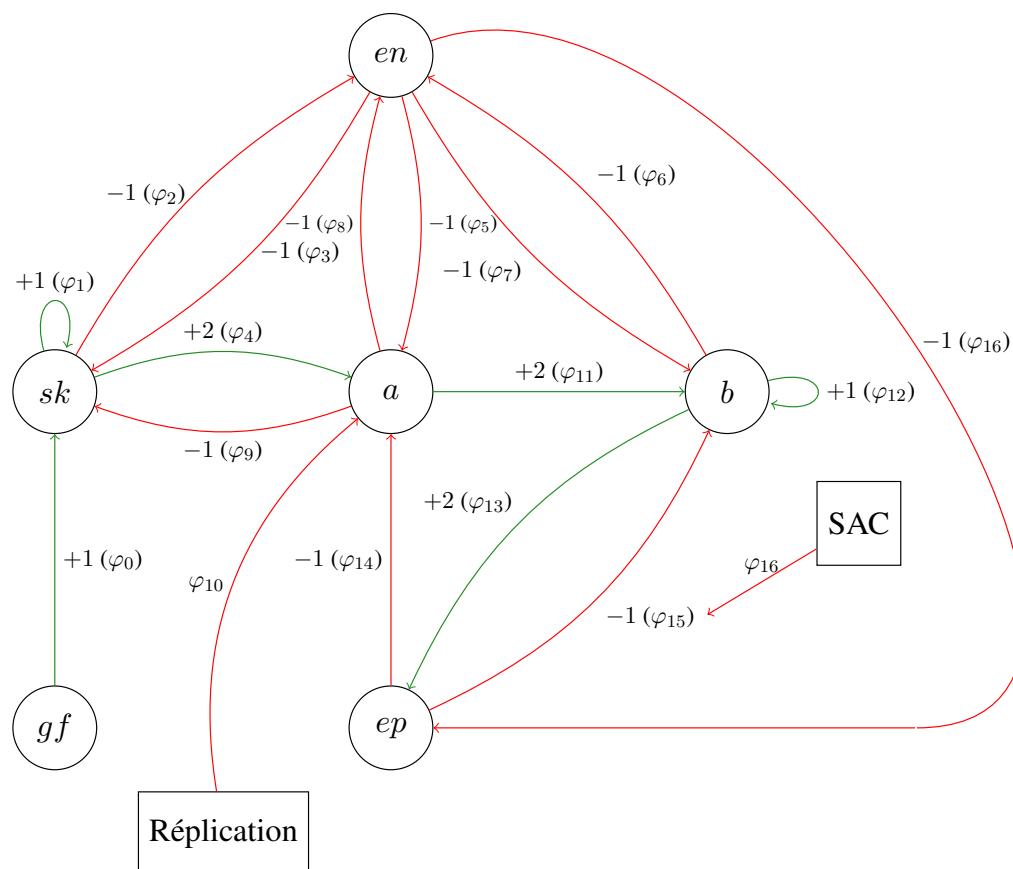


Figure 5.12 – Révision du graphe d'influence de [69]. La variable gf n'a aucun prédécesseur et est appelée par conséquent une variable d'environnement puisqu'elle est sous le contrôle total du modélisateur.

L'évènement en- : La protéine p21 se lie à une protéine capable d'interagir avec les ADN polymérase et inhibe son activité de catalyse de la processivité pour l'ADN [32].

Le second évènement $sk +$ ($sk = 1 \rightarrow sk = 2$) : Le complexe $cycE/cdk2$ (sk) est pleinement actif dans un état hyperphosphorylé. Il participe à l'activation du complexe de pré-réplication (abrégé CPR) aux origines de répliation (figure 5.13). Ce complexe moléculaire est formé de plusieurs sous-unités protéiques dont certaines sont phosphorylées directement par sk . Par exemple la phosphorylation de $cdt1$ et $cdc6$ par sk participe à l'assemblage du CPR aux origines de répliation [41]. En outre, il a été démontré que sk phosphoryle $Sld2$ et $Sld3$ ce qui confère au CPR sa capacité à dérouler l'ADN et permet le recrutement de l'ADN polymérase [56]. L'inhibition de la $cdk2$ dans des cellules humaines inhibe l'initiation de la phase S [26]. Le choix de nos variables implique l'hypothèse de modélisation suivante : une fois que sk a atteint son second seuil de régulation, on suppose que toutes les origines de répliations sont liées à CPR actif. Autrement dit, on suppose aucune mutation liée à un défaut d'assemblage ou d'activation d'un CPR si tous

les évènements de régulation du cycle cellulaire se déroulent dans l'ordre prévu (*cf. H_c page 97*). En outre, le complexe *cycE/cdk2* active indirectement la duplication de la chromatine, en activant le facteur de transcription des gènes codant pour les principales histones, indispensables à la formation du nucléosome.

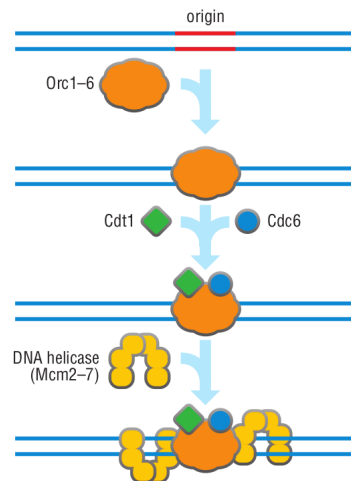


Figure 5.13 – **Assemblage du complexe de pré-réplication.** Figure extraite de [34].

5.4.2.2 Les évènements requis en S

Le premier évènement $a+$ ($a = 0 \rightarrow a = 1$) : Il est associé à l'inactivation des protéines qui forment le CPR. De fait, le complexe *cycA/cdk2* est impliqué dans la prévention de la multi-réplication au cours d'un même cycle. Pour limiter la réplication du génome à une fois par division cellulaire, les cellules ne doivent pas être en mesure de répliquer l'ADN dans les régions qui ont déjà été répliquées avant la fin de la division cellulaire. La variable a inhibe au seuil 1 par phosphorylation les protéines formant le CPR. Comme décrit plus haut, a inhibe sk au seuil 1 *via* une réaction de phosphorylation inhibitrice de E2F. Il s'avère que l'inhibition de E2F a également pour effet de réprimer l'expression des gènes nécessaires à la réplication de l'ADN, ainsi que de plusieurs protéines essentielles à l'assemblage des CPR. La prévention de l'initiation de plusieurs cycles de réplication lors d'une même phase S évite des blocages des fourches de réplication et des cassures des brins d'ADN [56].

Les deux évènements $sk-$: Ils représentent l'inactivation des complexes *cycE/cdk2* consécutive à l'évènement $a+$. Nous supposons qu'ils témoignent d'une inhibition effective des complexes de pré-réplication.

5.4.2.3 Les évènements requis en G2

Les cyclines A et B sont les deux régulateurs majeurs de l'entrée en mitose.

Le second évènement $a+$ ($a = 1 \rightarrow a = 2$) : Le complexe *cycA/cdk1* inhibe également l'assemblage du CPR en phosphorylant de surcroît Orc1 et Cdc6 [56]. Ce complexe modélisé

uniquement au seuil 2 de a est impliqué dans l'activation de FOXM1, un facteur de transcription des gènes codant pour des protéines essentielles pour l'initiation de la mitose, parmi eux : Cdc25, la cycline B, Plk1 et Aurora B [24]. Plk1 et Aurora A, deux protéines nécessaires à la maturation du centrosome et la formation du fuseau mitotique.

Le premier évènement $b+$ ($b = 0 \rightarrow b = 1$) : L'accumulation de cycline B a lieu brutalement en début de mitose. Elle nécessite une activation préalable du complexe cycB/cdk1.

5.4.2.4 Les évènements requis en M

L'évènement $b+$: En réponse à l'augmentation abrupte de la concentration en cycline B, le complexe cycB/cdk1 s'active pleinement et est ensuite exporté vers le noyau où il commence à phosphoryler les substrats nucléaires nécessaires à l'initiation de la mitose. Par exemple, b active à ce seuil Plk1 et Aurora A, deux protéines nécessaires à la maturation et la séparation du centrosome (le centre organisateur des microtubules) et la formation du fuseau mitotique (illustré en figure 5.11). Un autre exemple est l'initiation d'un évènement de la prophase : la destruction de l'enveloppe nucléaire [47].

L'évènement $ep+$: APC-cdc20 est impliquée dans la sortie de mitose marquée par le début de l'anaphase, phase de la mitose au cours de laquelle les chromatides soeurs sont abruptement séparées. APC-cdc20 dégrade indirectement la Sécurine, une protéine responsable de la cohésion des chromatides soeurs. APC-cdc20 joue un rôle essentiel la répartition équitable des chromosomes entre les deux cellules filles, indispensable pour protéger la cellule contre la perte d'un chromosome (aneuploïdie) [32].

Les deux évènements $a-$ et les deux $b-$: Ils représentent la dégradation des cyclines A et B consécutives à l'évènement $ep+$. La dégradation de a et b marque une levée de l'activation de leurs substrats impliqués dans la condensation des chromosomes et la dissociation de l'enveloppe nucléaire. [30] mentionnent que ces quatre évènements (*i.e* la dégradation des cyc/cdk mitotiques) induisent la terminaison de la phase M caractérisée notamment par la décondensation des chromosomes et l'assemblage de l'enveloppe nucléaire. La dégradation de a réprime son action inhibitrice du complexe de pré-réplication, ce qui participe à la préparation des cellules filles à la réplication du prochain cycle.

L'évènement $en+$: Les phosphatases PP2A et PP1 déphosphorylent les substrats mitotiques activés plus tôt dans le cycle par cycA/cdk1 et cycB/cdk1. L'association de cet évènement et des quatre précédents sont suffisants pour déclencher la terminaison de la mitose à la manière d'un switch [34]. [30] émettent l'hypothèse que l'association d'APC avec Cdh1 en fin de mitose maintient l'activité ubiquitineuse d'APC pour le cycle suivant.

L'évènement $ep-$: Il n'est pas prouvé qu'il soit requis en M, d'après nos connaissances actuelles. Nous supposons qu'il a lieu en M car est nécessaire pour boucler sur l'état initial de notre phase G1, nous référons le lecteur au triplet de Hoare du cycle cellulaire canonique en équation 5.4.

5.4.2.5 La séquence canonique résultante

In fine, le multi-ensemble des évènements requis lors de chaque phase vient d'être défini à l'aide des régulateurs clés de la progression du cycle cellulaire :

$$\begin{aligned} G1 &: \{sk+ : 2, en- : 1\} \\ S &: \{a+ : 1, sk- : 2\} \\ G2 &: \{a+ : 1, b+ : 1\} \\ M &: \{b+ : 1, ep+ : 1, a- : 2, b- : 2, en- : 1, ep- : 1\} \end{aligned}$$

La condition initiale du cycle cellulaire canonique est donnée par la formule $G1_{init}$ en section 5.3. Elle formalise l'état initial de G1 dans lequel gf est activé, et est pour rappel sous contrôle du modélisateur. Le passage à 1 de gf a donc déjà eu lieu dans notre état initial du cycle cellulaire, ce qui implique nous ne l'incluons pas dans le multi-ensemble des évènements requis en G1. Nous avons déduit de la littérature une séquence canonique du cycle cellulaire à partir de $G1_{init}$, découpée en sous-séquences caractéristiques de G1, S, G2 et M. Cette séquence et ce découpage préliminaire en phase représentent un consensus autour des dépendances entre les évènements requis lors de chaque phase du cycle cellulaire. Voici le triplet de Hoare canonique résultant :

$$H_c \equiv \{G1_{init}\} \underbrace{\underbrace{sk+; en-; sk+; a+; sk-; sk-}_{G1}; \underbrace{a+; b+}_{G2}; \underbrace{b+; ep+; a-; a-; b-; b-; ep-; en+}_{M}}_{p_c} \{G1_{init}\} \quad (5.4)$$

L'ordre des évènements requis au sein d'une phase n'est pas une question déterminante sachant que la thèse se focalise sur les checkpoints du cycle cellulaire. Néanmoins il s'avère qu'une exploration de la littérature nous offre des connaissances sur des dépendances entre les évènements requis pour chaque phase :

- **En phase G1** : Partant de $G1_{init}$, sk ne peut être pleinement activé (second $sk+$) que si en , au sens d'APC-cdh1, est préalablement inhibé par sk . Le premier $sk+$ précède $en-$ [70].
- **En phase S** : une fois les évènements requis de G1 réalisés en partant de $G1_{init}$, la dégradation de sk (deux $sk-$) est induite une fois le complexe $cycA/cdk2$ activé (premier $a+$) [19].
- **En phase G2** : b est initialement activé par le complexe $cycA/cdk1$. L'évènement $a+$ a lieu avant $b-$ [56].
- **En phase M** :
 1. APC-cdc20 (ep) ne peut agir sur ses cibles que lorsque b est totalement activé ($b = 2$) et APC-cdh1 (en) inactif [48].
 2. APC-cd20 (ep) cible à la fois la cycline A et la cycline B pour la dégradation protéolytique et l'on observe que l'inhibition de a a lieu avant celle de b . Lorsque APC-cdc20 est activée entre la prophase et la prométaphase, son action sur b est inhibée tant que le SAC est activé, c'est à dire jusqu'à la transition métaphase/anaphase Son action sur a est quant à elle indépendante du SAC [46]. Le premier $a-$ précède donc le premier $b-$. On admet que les deux $a-$ ont lieu avant les deux $b-$ dans le simple but d'avoir une séquence canonique unique.

3. La dégradation des cyclines dépend d'APC-cdc20 en phase M [30]. L'évènement $ep+$ a lieu avant $a-; a-; b-; b-$.
4. Une fois seulement lorsque les cyclines A et B dégradées ($a-; a-; b-; b-$), on suppose (toujours dans le but d'avoir une unique séquence) qu'une fraction libre d'APC est capable de se lier à sa sous-unité Cdh1 pour former APC-cdh1 ($en+$). L'évènement $en+$ suit $a-; a-; b-; b-$.
5. On admet que en induit ensuite la dégradation de Cdc20 ($ep-$) pour boucler la boucle. L'évènement $en+$ précède $ep-$ dans la séquence canonique.

Les connaissances actuelles nous permettent aussi d'expliquer pourquoi certains évènements appartenant chacun à une phase différente doivent avoir lieu dans un certain ordre. L'évènement qui clôt une phase ne doit pas avoir lieu avant celui qui initie la phase suivante. Pour rappel, cela est un garant de l'intégrité de l'ADN au cours du cycle cellulaire. Voici deux exemples caractéristiques :

- **G1/S** : le second $sk+$ est requis avant le premier $a+$. Cette chronologie est cruciale car si $a+$ a lieu avant $sk+$ l'activation des CPR ne peuvent avoir lieu étant donné que $cycE/cdk2$ est un activateur des CPR et $cycA/cdk2$ un de leur inhibiteur [56].
- **S/G2** : l'activité du complexe $cycA/cdk1$ en phase G2 est réprimée tant que la réplication n'est pas terminée. D'après nos hypothèses de modélisation, la réplication est admise terminée après réalisation de tous les évènements requis en S. Le premier évènement $a+$ est, d'après la sémantique de René Thomas, naturellement requis avant le second $a+$, mais les deux évènements $sk-$ sont requis avant $a+$.

5.4.3 Définition des multiplexes

5.4.3.1 Multiplexes et abstraction mécanistique

Les arcs $\varphi_{[1-9]}$ et $\varphi_{[11-15]}$ du graphe d'influence de la figure 2.3 formalisent des littéraux du langage des multiplexes (voir figure 2.5 page 27). Si la variable x active (resp. inhibe) au seuil n la variable y alors on définit $x \geq 1$ (resp. $\neg(x \geq 1)$) pour formaliser cette régulation. Chacune de ces formules est associée à une étiquette pour former des multiplexes. Voici un exemple d'atome qui représente la présence de sk en concentration suffisante pour s'auto-activer :

$$\varphi_1 \equiv (sk \geq 1)$$

On associe à φ_1 l'étiquette rp (pour *restriction point*) dans le GRBM présenté en figure 5.18. Voici désormais un exemple de littéral négatif qui formalise une concentration de sk insuffisante pour induire une répression de l'activité de en :

$$\varphi_2 \equiv \neg(sk \geq 1)$$

On associe à φ_2 l'étiquette no_sk dans le GRBM de la figure 5.18. Un multiplexe représente conceptuellement une potentielle ressource. Lorsqu'un état satisfait la formule d'un multiplexe, il constitue une ressource (l'ensemble des ressources d'une variable est défini formellement par ϕ_v^ω en page 72).

Les mécanismes d'inhibition caractérisés par les formules $\varphi_5, \varphi_7, \varphi_{14}$ et φ_{15} du graphe d'influence de la figure 2.3 sont regroupés en un unique multiplexe : celui étiqueté no_CI en figure 5.18 et de formule $\neg(en \geq 1 \wedge ep \geq 1)$. Il s'agit d'un cas typique de regroupement de régulations de mêmes

natures pour minimiser le nombre de paramétrisations, nous renvoyons le lecteur au chapitre 2 page 28. Cette formule signifie que l'absence des deux inhibiteurs *en* ou *ep* est nécessaire pour constituer une potentielle ressource des variables cibles (*a* et *b*). Autrement dit, la présence de l'un de ces inhibiteurs ne peut en aucun cas constituer une ressource.

5.4.3.2 Multiplexes et abstraction phénoménologique

L'activation ou l'inhibition de nos variables a permis de formaliser les différents phénomènes qui ont lieu au cours de la réplication et de la mitose. Certains de ces phénomènes du cycle cellulaire *agissent en retour* sur nos variables. Deux variables de notre modèle sont concernées : *a* et *b*. Le processus de réplication dans un état inachevé inhibe le complexe *cycA/cdk2-1*, autrement dit, la fin de la réplication est un phénomène qui représente une ressource pour la variable *a* (figure 5.8). Un cas similaire est retrouvé en phase M. L'action de la protéine APC-cd20 sur le complexe *cycB/cdk1* est inhibée par le SAC (figure 5.10). La réplication et le SAC seraient deux processus complexes et coûteux à modéliser au niveau moléculaire tant les espèces moléculaires impliquées sont nombreuses et les mécanismes de régulation de nature variée. Au moment de l'écriture de cette thèse, 3029 gènes sont annotés avec le terme « réplication » [106], 446 pour « Spindle Assembly Checkpoint » [107] pour *Homo sapiens*. Toujours dans l'optique de suivre la philosophie du Rasoir d'Ockham, nous choisissons de nous contenter de nos six variables *sk*, *a*, *b*, *en*, *ep* et *gf* pour représenter ces processus sachant qu'elles nous ont permis de représenter une séquence plausible des phases du cycle cellulaire (triplet 5.4).

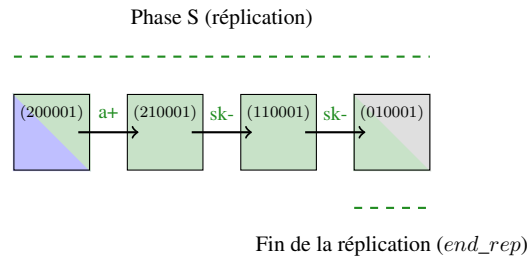


Figure 5.14 – Le multiplexe *end_rep* est défini à l'aide de la séquence canonique du cycle cellulaire. Ici est représentée la séquence canonique de la phase S à partir d'un état caractérisé par la formule $sk = 2 \wedge a = 0 \wedge b = 0 \wedge en = 0 \wedge ep = 0 \wedge gf = 1$ (abrégié 200001).

L'état abrégé 200001 est le marqueur de la phase S le plus tardif dans notre séquence canonique. Seules les variables *a* et *sk* interviennent dans les évènements de régulation de cette représentation de la phase S. Par conséquent, seules ces variables seront utilisées pour définir la fin de la réplication dans notre modèle. On définit le multiplexe *end_rep* qui a pour formule φ_{10} (figure 5.8) :

$$\varphi_{10} \equiv (a \geq 1 \wedge \neg(sk = 1))$$

L'activation du SAC est observable dans notre modèle lorsque *ep* est activé mais tant qu'il ne peut pas agir sur la dégradation de la cycline B (*b-*). En revanche même si le SAC est actif, *a* peut toujours commencer à se dégrader [46]. Cette fenêtre temporelle est caractérisée de nouveau à l'aide de notre triplet de Hoare canonique du cycle cellulaire : H_c . L'état abrégé (022011) correspond à la description la plus proche du SAC que l'on peut être amenés à réaliser à l'aide de nos variables

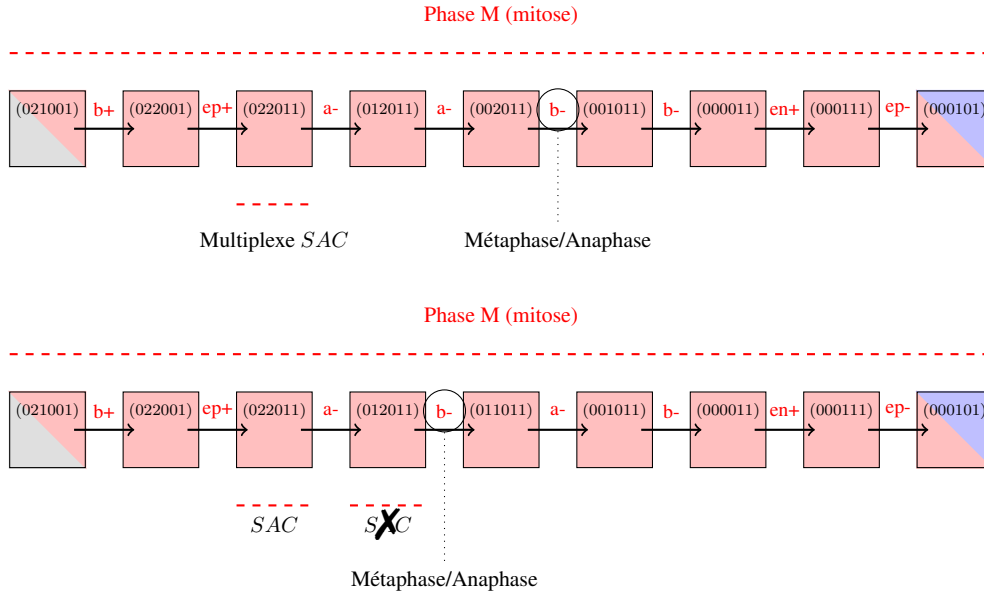


Figure 5.15 – Le multiplexe SAC est défini à l'aide d'une trace alternative de la séquence canonique du cycle cellulaire. La séquence canonique de M est représentée à partir de l'état (021001) dans la partie supérieure de la figure. Aucune observation n'empêche de représenter la dégradation de b par la trace alternative représentée dans la partie inférieure de la figure. La transition métaphase/anaphase est marquée par la levée de la répression de la dégradation de b (premier $b-$) par le SAC. On notera désormais cet évènement SAC_{off} .

du cycle cellulaire (partie supérieure de la figure 5.15). Le SAC est une ressource de b tant que sa dégradation n'est pas observée. Une étude des traces alternatives du chemin de H_c (corroborée en chapitre 6) montre qu'une fois que a commence à se dégrader (*i.e* $a = 1$ après le premier $a-$), la dégradation de b peut avoir lieu ($b-$), *cf.* la partie inférieure de la figure 5.15. L'état (012011) n'est donc pas adapté à la représentation du SAC qui est supposé être une ressource de b l'empêchant de se dégrader. En somme, le SAC est une ressource de b tant que ep est actif et tant que a n'est pas encore dégradé. On définit le multiplexe SAC en figure 5.18 de formule :

$$\varphi_{16} \equiv (ep \geq 1 \wedge a \geq 2)$$

Enfin, $restrict_sk$ est un autre multiplexe phénoménologique mais qui est issu de la seule observation de la séquence canonique de la phase S. La formule caractéristique de l'état abrégé (110001) en figure 5.16 est très proche d'un état de la phase G1 abrégé (100001), voir figure 5.19 partie supérieure. En se restreignant aux variables requises en phase S, le multiplexe $restrict_sk$ doit représenter une ressource que a doit posséder dans cet état particulier de la phase S. Cette ressource empêche a de réaliser la transition qui ramène le système en G1 :

$$\varphi_{17} \equiv (sk \geq 2 \wedge a \geq 1)$$

Pour finir, le passage du point de restriction (pour rappel abrégé PPR) est associé en sous-section 5.4.2.1 au moment du premier $sk+$, soit en phase G1 d'après H_c . Or, des observations expérimentales montrent que le passage du point de restriction, associé spécifiquement à cet évènement

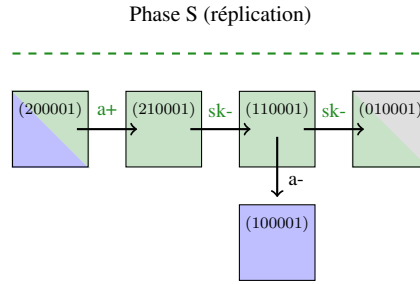


Figure 5.16 – Le multiplexe *restrict_sk* est également défini à l’aide de la séquence canonique du cycle cellulaire. Introduire ce multiplexe empêche un retour à la phase G1 en cours (en bleu) à partir d’un état de S (en vert).

sk+, peut aussi avoir lieu en fin de phase M [62]. Le problème est donc de définir un découpage raisonnable de la phase M canonique en une portion précoce et une portion tardive.

Une fois ce problème réglé, il est nécessaire de limiter le PPR uniquement à la portion tardive de la phase M. L’action des facteurs de croissance est prouvée déterminante dans le passage du point de restriction en sous-section 5.4.2.1. L’action de *gf* sur *sk* est restreinte à la portion tardive de M caractérisée à l’aide du premier évènement *b-* caractéristique de la transition métaphase/anaphase, figure 5.17. Plus précisément, on ajoute une condition nécessaire à *gf* pour activer sur *sk*. Exprimé en terme de ressource, on obtient le multiplexe *restrict_gf* suivant :

$$\varphi_0 \equiv (gf \geq 1 \wedge \neg(b \geq 2))$$

qui signifie que *sk* possède *gf* comme ressource uniquement lorsque *b* n’est pas saturé, c’est à dire après le SAC en M (portion tardive donc).

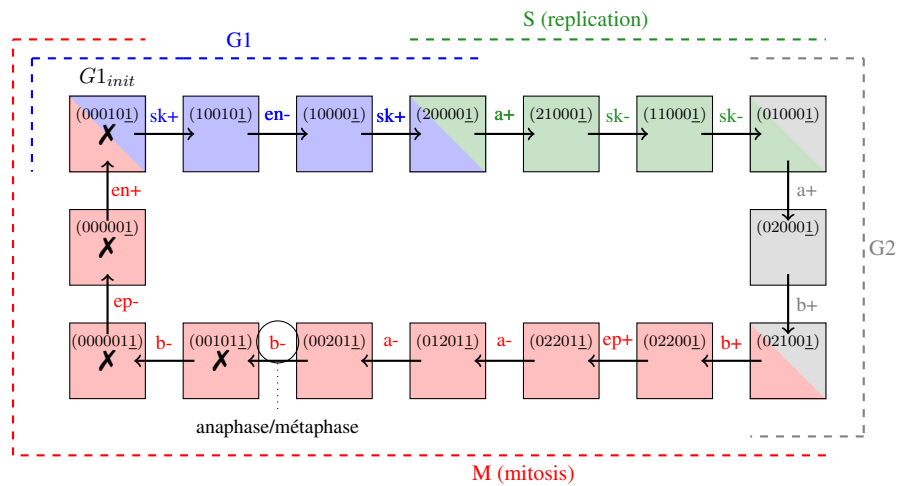


Figure 5.17 – La fenêtre temporelle du cycle cellulaire au cours de laquelle le passage du point de restriction (le premier *sk+*) est autorisé. Un exemple avec notre triplet H_c . *sk* possède la ressource nécessaire (*restrict_sk*) pour réaliser le PPR dans les états marqués d’une croix.

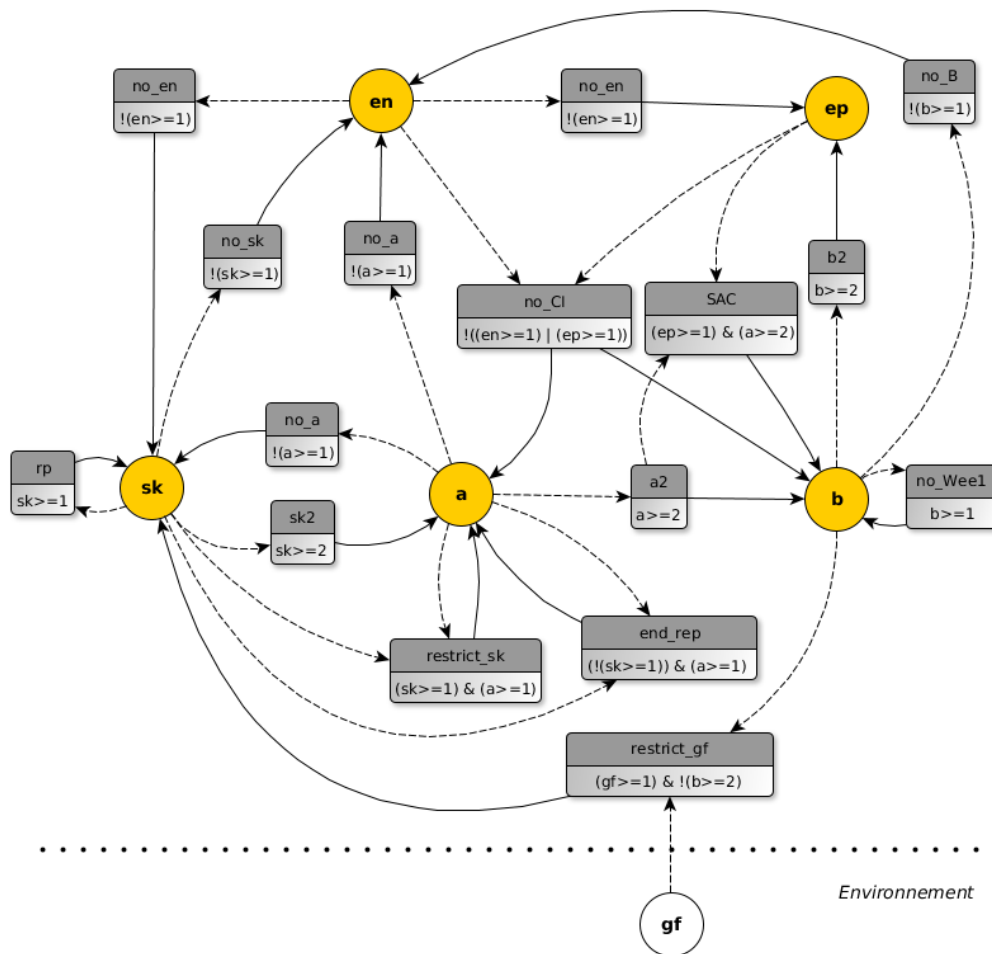


Figure 5.18 – Le GRBM du cycle cellulaire des mammifères. Les multiplexes sont représentés par des rectangles en deux parties : l'étiquette et la formule associée. Leur formule caractérise soit un littéral du langage des multiplexes, soit des coopérations connues entre certaines variables pour *potentiellement* activer leur cible. Les coopérations entre les prédécesseurs d'une variable ne sont pas connues à ce stade. L'instanciation de tous les éléments de la famille des paramètres \mathcal{K} permettra d'identifier ces règles de coopération. Ce graphe s'inspire initialement de l'abstraction proposée dans le modèle de [40] et est représenté à l'aide du logiciel yEd. Nous renvoyons le lecteur vers une démonstration de l'utilisation de TotemBioNet pour plus d'informations [96].

5.5 Première identification des valeurs des paramètres

Lorsque la cellule ne prolifère pas, elle est dans une phase de quiescence. Notre GRBM doit engendrer uniquement deux bassins d'attraction : un attracteur cyclique qui inclut notre cycle cellulaire canonique (H_c) et un état stable correspondant à notre phase de quiescence (G_0). La transition d'un bassin d'attraction vers un autre est contrôlée par les facteurs de croissance du milieu extracellulaire, c'est à dire par le modélisateur lui même qui se projette dans l'un ou dans l'autre des bassins en fixant la valeur de gf qui est une variable d'environnement. Nous y reviendrons en

détail plus bas. La figure 5.19 illustre les deux attracteurs formalisés à partir des variables et des seuils de notre GRBM.

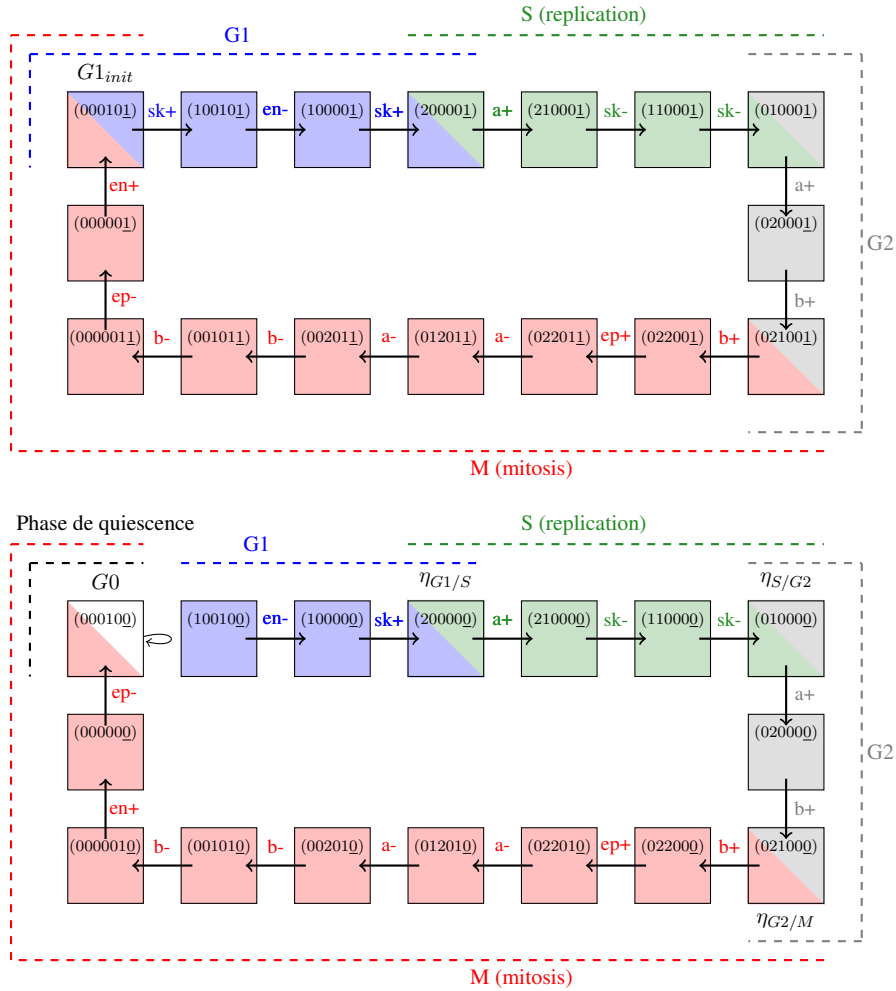


Figure 5.19 – **Deux attracteurs associés à notre GRBM.** TotemBioNet permettra de détecter la présence d’un unique attracteur dans le contexte $AG(gf = 1)$ (resp. $AG(gf = 0)$) à l’aide de la formule 5.8 (resp. formule 5.5). Le cycle cellulaire canonique H_c appartient au premier attracteur (sous-figure du haut). L’état $G0$ est stable et caractérise le second attracteur (sous-figure du bas). Une fois le passage de restriction passé (premier $sk+$), les phases G1, S, G2 et M ont lieu qu’importe la concentration en facteurs de croissance.

5.5.1 Vérification de la stabilité de la phase de quiescence

La phase de quiescence est formalisée par la formule $G0$ en section 5.3 page 86. Toutes les variables sont instanciées donc cette formule spécifie un état. La stabilité de cet état est aisément formalisable en CTL, cf. chapitre 4 page 68. La stabilité de $G0$ est définie à l’aide des modalités de CTL équitable :

$$AG(gf = 0) \Rightarrow AF(AG(G0)) \tag{5.5}$$

Cette formule signifie qu'à partir de n'importe quel état associé à notre GRBM et appartenant à l'hyperplan $gf = 0$, tous les chemins mèneront à la phase de quiescence.

Au moins l'un de ces chemins doit vérifier la succession des phases G1, S, G2 et M. En effet, une fois le passage du point de restriction atteint, une cellule privée de facteurs de croissance réalise les événements la réplication et la mitose du cycle courant. Le passage du point de restriction est ici associé à l'évènement $sk+$ à partir de $G1_{init}$. Si l'on retire les facteurs de croissance après le passage du point de restriction, l'état initial de G1 projeté dans l'hyperplan $gf = 0$ devient :

$$G1'_{init} \equiv sk = 0 \wedge a = 0 \wedge b = 0 \wedge en = 1 \wedge ep = 0 \wedge gf = 1 \quad (5.6)$$

et le cycle cellulaire projeté dans l'hyperplan $gf = 0$ devient :

$$H'_c \equiv \{G1'_{init}\} \underbrace{\begin{array}{c} \text{en-; sk+; a+; sk-; sk-; a+; b+; b+; ep+; a-; a-; b-; b-; ep-; en+} \\ \text{G1} \quad \text{S} \quad \text{G2} \quad \text{M} \end{array}}_{p'_c} \{G0\} \quad (5.7)$$

5.5.2 Formalisation d'un cycle cellulaire canonique comme un attracteur cyclique

Notre cycle cellulaire canonique (H_c) est supposé faire partie d'un unique attracteur appartenant à l'hyperplan $gf = 1$. Les états parcourus le long du chemin p_c depuis l'état $G1_{init}$ doivent donc être parcourus une infinité de fois. On définit la formule suivante à l'aide des modalités de CTL équitable :

$$AG(gf = 1) \Rightarrow AG[AF(G1_{init})] \quad (5.8)$$

5.5.3 Se poser les bonnes questions sur les checkpoints à l'aide de TotemBioNet

Sans aucune exploitation des valeurs des paramètres contraints manuellement, TotemBioNet réduit l'espace de recherche des modèles à 649 600 modèles, grâce à la résolution quasi-instantanée de :

- la condition de Snoussi (définition 8 page 31),
- la plus faible précondition calculée à partir de H_c (voir plus bas $wp(H_c)$),
- celle calculée à partir de H'_c (voir plus bas $wp(H'_c)$).

L'algorithme de model-checking de TotemBioNet nécessite 4h30min[†] pour identifier les paramétrisations de notre GRBM qui satisfont la propriété de blocage en phase de quiescence d'une cellule constamment privée de facteurs de croissance (formule 5.5) et celle d'une division infinie d'une cellule constamment infusée en facteurs de croissance (formule 5.8). Notons par ailleurs que dans les cellules souches embryonnaires, E2F et le complexe cycE/cdk2 (cibles de gf) sont continuellement actifs de manière indépendante du cycle cellulaire [27]. Ceci leur permet de réaliser théoriquement à l'infini le cycle cellulaire, si l'on exclut par exemple les phénomènes de sénescence. L'efficacité de TotemBioNet permet de converger encore plus rapidement sur la question de modélisation des checkpoints en exploitant additionnellement des contraintes sur les

†. Les temps CPU ont été mesurés sur un processeur Intel Core i7-8650U 1.90GHz × 8 sous Ubuntu 20.04.4 LTS.

valeurs des paramètres définies manuellement dans la section suivante. Toutes les contraintes sur les paramètres de notre GRBM sont recensées en annexe A page 159. Ces dernières permettent d'identifier en moins d'une seconde (858 ms précisément), un ensemble de **16 paramétrisations** qui satisfont toutes les propriétés dynamiques devenues incontournables dans la validation des modèles du cycle cellulaire (table 5.2).

La précondition calculée à partir du GRBM et du triplet de Hoare H_c est :

$$\begin{aligned}
& wp(H_c) \\
& \equiv \\
& \underbrace{sk = 0 \wedge a = 0 \wedge b = 0 \wedge ep = 0 \wedge en = 1 \wedge gf = 1}_{G1_{init}} \wedge \\
& \underbrace{K_{sk, \{no_a, restrict_gf\}} > 0}_{sk+} \wedge \underbrace{K_{en, \{no_a, no_b\}} < 1}_{en-} \wedge \underbrace{K_{sk, \{no_a, restrict_gf, rp, no_en\}} > 1}_{sk+} \wedge \\
& \underbrace{K_{a, \{sk2, no_CI\}} > 0}_{a+} \wedge \underbrace{K_{sk, \{rp, no_en, restrict_gf\}} < 1}_{sk- ; sk-} \wedge \\
& \underbrace{K_{a, \{end_rep, no_CI\}} > 1}_{a+} \wedge \underbrace{K_{b, \{a2, no_CI\}} > 0}_{b+} \wedge \\
& \underbrace{K_{b, \{a2, no_Wee1, no_CI\}} > 1}_{b+} \wedge \underbrace{K_{ep, \{b2, no_en\}} > 0}_{ep+} \wedge \underbrace{K_{a, \{end_rep\}} < 1}_{a- ; a-} \wedge \\
& \underbrace{K_{b, \{no_Wee1\}} < 1}_{b- ; b-} \wedge \underbrace{K_{en, \{no_a, no_sk, no_b\}} > 0}_{en+} \wedge \underbrace{K_{ep, \emptyset} < 1}_{ep-}
\end{aligned}$$

Celle calculée à partir de H'_c est :

$$\begin{aligned}
& wp(H'_c) \\
& \equiv \\
& \underbrace{sk = 1 \wedge a = 0 \wedge b = 0 \wedge ep = 0 \wedge en = 1 \wedge gf = 0}_{G1'_{init}} \wedge \\
& \underbrace{K_{en, \{no_a, no_b\}} < 1}_{en-} \wedge \underbrace{K_{sk, \{no_a, rp, no_en\}} > 1}_{sk+} \wedge \\
& \underbrace{K_{a, \{sk2, no_CI\}} > 0}_{a+} \wedge \underbrace{K_{sk, \{rp, no_en\}} < 1}_{sk- ; sk-} \wedge \\
& \underbrace{K_{a, \{end_rep, no_CI\}} > 1}_{a+} \wedge \underbrace{K_{b, \{a2, no_CI\}} > 0}_{b+} \wedge \\
& \underbrace{K_{b, \{a2, no_Wee1, no_CI\}} > 1}_{b+} \wedge \underbrace{K_{ep, \{b2, no_en\}} > 0}_{ep+} \wedge \underbrace{K_{a, \{end_rep\}} < 1}_{a- ; a-} \wedge \\
& \underbrace{K_{b, \{no_Wee1\}} < 1}_{b- ; b-} \wedge \underbrace{K_{en, \{no_a, no_sk, no_b\}} > 0}_{en+} \wedge \underbrace{K_{ep, \emptyset} < 1}_{ep-}
\end{aligned}$$

Tous les résultats exposés sont disponibles à l'adresse suivante : <https://gitlab.com/deborahboyenval/temporal-separation-prototype/-/tree/master/ProofOfConceptCheckpoint>.

Pour être plausible biologiquement, notre ensemble de 16 modèles doit impérativement vérifier la préservation de l'intégrité de l'ADN le long du cycle cellulaire. Cette propriété de préservation

représente un défi de formalisation puisqu'elle nécessite au préalable une modélisation convaincante des phases du cycle cellulaire. Notre triplet H_c capture l'ensemble des observations qui forment un consensus sur la séquence des activations et inactivation des principaux régulateurs de la réplication et la mitose. En revanche, la question de *la formalisation des phases* n'est pas résolue puisque ce choix de séquence pose trop *a priori* sur l'ordre des événements au sein d'une phase. Pour autant, TotemBioNet est très efficace pour détecter l'insatisfiabilité d'une trace canonique du cycle cellulaire et cela suffirait à remettre en cause, à ce stade de modélisation, la définition des événements requis au sein de chaque phase. On ne peut en revanche pas conclure à une formalisation des phases mais à une première vérification de la cohérence biologique de notre GRBM.

| Propriété dynamique | Formalisation | Logique |
|--|--|----------|
| Validation du cycle cellulaire canonique | $\{G1_{init}\} p_c \{G1_{init}\}$ - (5.4) | LHGM |
| Une cellule qui a passé son point de restriction termine le cycle cellulaire même privée de facteurs de croissance | $\{G1'_{init}\} p'_c \{G0\}$ - (5.7) | LHGM |
| Blocage en phase de quiescence | $AG(gf = 0) \Rightarrow AF[AG(G0)]$ - (5.5) | FAIR-CTL |
| Le cycle cellulaire canonique appartient à un attracteur cyclique | $AG(gf = 1) \Rightarrow AG[AF(G1_{init})]$ - (5.8) | FAIR-CTL |

Table 5.2 – Les propriétés du cycle cellulaire vérifiées avant d'aborder la question de la modélisation des checkpoints. LHGM pour « Logique de Hoare Génétiquement Modifiée » et FAIR-CTL pour la version équitable de CTL.

L'approche de vérification formelle de modèles *à la Thomas*, présentée au chapitre 4 et illustrée dans ce chapitre, part du principe que l'on sait formaliser les hypothèses d'étude des biologistes. La formalisation des phases se heurte à ce principe, comme expliqué au chapitre 3. L'ensemble des 16 modèles résultants dans ce chapitre représente un matériel de départ solide pour tester une formalisation du domaine des phases du cycle cellulaire à partir des événements requis (chapitre 6), ainsi que pour tester une formalisation des checkpoints intrinsèques qui contraignent les événements qui peuvent avoir lieu au sein de chaque domaine d'une phase (chapitre 7).

5.5.4 Identification manuelle de certains paramètres

Les paramètres $K_{v,\omega}$ définissent un contexte où un ensemble ω de ressources d'une variable biologique v est fixé. Nous avons transposé les connaissances accumulées sur le cycle cellulaire en contraintes sur la valeurs de certains paramètres $K_{v,\omega}$. Ces contraintes sont obtenues à partir d'une expérience de pensée (section 2.4.2.4 page 31) et peuvent être dans le meilleur des cas corroborées par une ou plusieurs observations expérimentales non contradictoires.

La variable en : APC-cdh1 est actif uniquement lorsque cdk1 et cdk2 sont inactives, c'est à dire lorsque sk , a et b sont inactives [56]. En reprenant les termes du formalismes booléens, on pourrait définir la fonction d'activation de en suivante :

$$f_{en} : sk = 0 \wedge a = 0 \wedge b = 0$$

Cette règle d'activation de en permet d'identifier tous les paramètres de en qui s'active uniquement lorsqu'il possède ses trois ressources, à savoir l'absence de l'inhibition de sk (no_sk), de a (no_a) et de b (no_b).

$$\begin{aligned}
K_{en,\emptyset} &= 0 \\
K_{en,\{no_sk\}} &= 0 \\
K_{en,\{no_a\}} &= 0 \\
K_{en,\{no_b\}} &= 0 \\
K_{en,\{no_sk,no_a\}} &= 0 \\
K_{en,\{no_sk,no_b\}} &= 0 \\
K_{en,\{no_a,no_b\}} &= 0 \\
K_{en,\{no_sk,no_a,no_b\}} &= 1
\end{aligned}$$

Ce jeu de paramètres n'est autre que la table de vérité de la conjonction.

La variable sk : L'activation du complexe $cycA/cdk1$ ($a = 2$) empêche sk d'agir sur l'activation des complexes de pré-réplication ($sk = 2$) [56]. On suppose en plus que l'activation de a au seuil 1 est déterminante dans la répression de l'activité de sk au point où l'absence de cette ressource empêche sk d'agir sur ses cibles, qu'importe la valeur de la variable d'environnement gf [101].

$$\forall \omega \mid \{no_a\} \notin \omega, K_{sk,\omega} = 0$$

et :

$$\begin{aligned}
K_{sk,\emptyset} &= 0 \\
K_{sk,\{restrict_gf\}} &= 0 \\
K_{sk,\{rp\}} &= 0 \\
K_{sk,\{no_en\}} &= 0 \\
K_{sk,\{restrict_gf,rp\}} &= 0 \\
K_{sk,\{restrict_gf,no_en\}} &= 0 \\
K_{sk,\{rp,no_en\}} &= 0 \\
K_{sk,\{rp,restrict_gf,no_en\}} &= 0
\end{aligned}$$

Il existe une autre configuration où sk est incapable d'agir sur ses cibles : lorsqu'aucun de ses activateurs n'est actif. sk est activé par les facteurs de croissance (ressource $restrict_gf$) et par lui même au seuil 1 (ressource rp) :

$$\forall \omega \mid \{restrict_gf, rp\} \notin \omega, K_{sk,\omega} = 0$$

et :

$$\begin{aligned}
K_{sk,\emptyset} &= 0 \\
K_{sk,\{no_en\}} &= 0 \\
K_{sk,\{no_a\}} &= 0 \\
K_{sk,\{no_en,no_a\}} &= 0
\end{aligned}$$

Ensuite, sk ne peut totalement s'activer tant que la protéine $p21$ est active ($en = 1$) :

$$\forall \omega \mid \{no_en\} \notin \omega, K_{sk,\omega} < 2$$

Par conséquent :

$$\begin{aligned}
K_{sk,\emptyset} &< 2 \\
K_{sk,\{no_a\}} &< 2 \\
K_{sk,\{rp\}} &< 2 \\
K_{sk,\{restrict_gf\}} &< 2 \\
K_{sk,\{no_a,rp\}} &< 2 \\
K_{sk,\{no_a,restrict_gf\}} &< 2 \\
K_{sk,\{rp,restrict_gf\}} &< 2 \\
K_{sk,\{no_a,rp,restrict_gf\}} &< 2
\end{aligned}$$

La concentration de sk ne peut atteindre sa saturation alors que le passage du point de restriction ($sk+$) n'a pas eu lieu [101] :

$$\forall \omega \mid \{rp\} \notin \omega, \quad K_{sk,\omega} < 2$$

$$\begin{aligned}
K_{sk,\emptyset} &= < 2 \\
K_{sk,\{restrict_gf\}} &< 2 \\
K_{sk,\{no_en\}} &< 2 \\
K_{sk,\{no_a\}} &< 2 \\
K_{sk,\{restrict_gf,no_en\}} &< 2 \\
K_{sk,\{restrict_gf,no_a\}} &< 2 \\
K_{sk,\{no_en,no_a\}} &< 2 \\
K_{sk,\{restrict_gf,no_en,no_a\}} &< 2
\end{aligned}$$

La variable a : La fin de la réplication, représentée par le multiplexe end_rep (figure 5.14), lève la répression de l'activité du complexe $cycA/cdk1$. La ressource end_rep est nécessaire à l'activation du complexe $cycA/cdk1$ ($a = 2$), autrement dit l'absence de cette ressource ne permet pas à a se s'activer au seuil 2 :

$$\begin{aligned}
K_{a,\emptyset} &< 2 \\
K_{a,\{restrict_sk\}} &< 2 \\
K_{a,\{no_CI\}} &< 2 \\
K_{a,\{sk2\}} &< 2 \\
K_{a,\{restrict_sk,no_CI\}} &< 2 \\
K_{a,\{restrict_sk,sk2\}} &< 2 \\
K_{a,\{no_CI,sk2\}} &< 2 \\
K_{a,\{restrict_sk,no_CI,sk2\}} &< 2
\end{aligned}$$

La variable b : L'activation partielle de b par a est *nécessaire* à l'auto-activation de b [43], ce qui est interprété par l'attrance de b vers la valeur 1 dans cette configuration de ressources :

$$\forall \omega \mid a2 \notin \omega, \quad K_{b,\omega} < 1$$

Par conséquent :

$$\begin{aligned}
K_{b,\emptyset} &= 0 \\
K_{b,\{no_Wee1\}} &= 0 \\
K_{b,\{no_CI\}} &= 0 \\
K_{b,\{no_Wee1,no_CI\}} &= 0
\end{aligned}$$

L'activation partielle de b par a est *suffisante* à l'auto-activation de b [43] :

$$\forall \omega \mid a2 \in \omega, K_{b,\omega} \geq 1$$

Ainsi :

$$\begin{aligned} K_{b,\{a2,no_Wee1\}} &\geq 1 \\ K_{b,\{a2,no_CI\}} &\geq 1 \\ K_{b,\{a2,SAC\}} &\geq 1 \\ K_{b,\{a2,no_Wee1,no_CI\}} &\geq 1 \\ K_{b,\{a2,no_Wee1,SAC\}} &\geq 1 \end{aligned}$$

Ensuite, PPA et PP2A (*en*) sont des phosphatases qui contrecarrent l'action de b sur les substrats mitotiques, modélisée lorsque b franchit son second seuil de modélisation [30].

$$K_{b,\omega} < 2, \forall K_{b,\omega} \setminus K_{b,\{a2,SAC,no_Wee1\}} \mid no_CI \in \omega$$

Donc :

$$\begin{aligned} K_{b,\emptyset} &< 2 \\ K_{b,\{a2\}} &< 2 \\ K_{b,\{no_Wee1\}} &< 2 \\ K_{b,\{a2,no_Wee1\}} &< 2 \\ K_{b,\{a2,SAC\}} &< 2 \end{aligned}$$

L'exception du paramètre $K_{b,\{a2,SAC,no_Wee1\}}$ vient du fait que le SAC empêche la dégradation de b avant celle de a lors de la mitose. Lorsque le SAC est déclenché (multiplexe *SAC* satisfait), que a est totalement actif (multiplexe *a2* satisfait) et que b est saturé (multiplexe *no_Wee1* satisfait), la dégradation de b est bloquée. En conséquence :

$$K_{b,\{a2,SAC,no_Wee1\}} = 2;$$

La variable ep : La protéine APC ne peut se lier à *cdc20* que si elle a été préalablement phosphorylée par le complexe *cycB/cdk1*. APC-*cdc20* (ep) ne peut agir sur ses cibles que (condition nécessaire) lorsque b est totalement activé ($b = 2$) [48]. Par conséquence :

$$\forall \omega \mid b2 \notin \omega, K_{ep,\omega} < 1$$

Ce qui nous permet de contraindre les paramètres suivants :

$$\begin{aligned} K_{ep,\emptyset} &= 0 \\ K_{ep,\{no_en\}} &= 0 \end{aligned}$$

Les variables a et b : Dès lors que la variable APC-*cdc20* est activée (ep), l'activité des complexes a et b est totalement réprimée ce qui fait tendre la valeur de ces deux variable vers zéro [30]. Dès que ep s'active, le multiplexe *no_CI* devient faux et n'appartient pas aux ressources des variables concernées :

$$\forall \omega_a \mid no_CI \notin \omega_a, K_{a,\omega_a} = 0 \quad \text{et} \quad \forall K_{b,\omega_b} \mid \{no_CI, a2\} \notin \omega_b, K_{b,\omega_b} = 0$$

Cette contrainte implique que l'activation de en , c'est à dire APC-cdh1, p21, p27, PP1 et PP2A, cause également l'inactivation de a et b . Cette hypothèse est admise pour la variable a car ne contredit pas les contraintes déjà définies sur la valeur de ses paramètres, par conséquent :

$$\begin{aligned} K_{a,\emptyset} &= 0 \\ K_{a,\{end_rep\}} &= 0 \\ K_{a,\{sk2\}} &= 0 \\ K_{a,\{restrict_sk\}} &= 0 \\ K_{a,\{sk2,restrict_sk\}} &= 0 \end{aligned}$$

Cette hypothèse n'est pas admise pour les paramètres de b incluant la ressource $a2$, cf. le paragraphe traitant la contrainte des valeurs de b page 109, ainsi :

$$\begin{aligned} K_{b,\emptyset} &= 0 \\ K_{b,\{SAC\}} &= 0 \\ K_{b,\{no_Wee1\}} &= 0 \\ K_{b,\{no_Wee1,SAC\}} &= 0 \end{aligned}$$

CHAPITRE 6

Formalisation des phases du cycle cellulaire

| | | |
|------------|---|------------|
| 6.1 | Préambule | 113 |
| 6.2 | Définition syntaxique d'une phase canonique | 114 |
| 6.3 | Formalisation d'une phase à partir de sa définition canonique | 115 |
| 6.3.1 | Définition de l'hyperpavé d'un triplet de Hoare | 115 |
| 6.3.2 | Vers une formalisation d'une phase sous la forme d'un prédicat du premier ordre | 117 |
| 6.4 | Application à notre modèle du cycle cellulaire | 118 |
| 6.4.1 | Les phases canoniques de notre modèle | 118 |
| 6.4.2 | Les hyperpavés de nos phases canoniques | 120 |
| 6.4.3 | Les phases de notre modèle | 123 |
| 6.4.3.1 | Implémentation du prédicat <i>estAdmis_σ</i> en Prolog | 123 |
| 6.4.3.2 | Ordre des événements totalement imposé pour les phases G1, S et G2 | 124 |
| 6.4.3.3 | Le cas remarquable de la phase M | 125 |

6.1 Préambule

Dans le chapitre 5, la satisfiabilité des propriétés temporelles classiques du cycle cellulaire a été prouvée. Elles expriment des propriétés d’atteignabilité et de stabilité qu’il est possible de formaliser en CTL ou en utilisant la logique de Hoare génétiquement modifiée (LHGM) et de les vérifier formellement sans difficulté. En revanche, la vérification formelle des checkpoints du cycle cellulaire ne peut pas suivre la même démarche puisqu’elle présuppose que l’on sait *préalablement* formaliser les connaissances biologiques sous-jacentes telles que les *phases* du cycle cellulaire.

En introduction, cinq checkpoints du cycle cellulaire ont été décrits biologiquement : le passage du point de restriction, le contrôle de l’initiation de la réplication (G1/S), de sa terminaison (S/G2), de l’entrée en mitose (G2/M) et l’assemblage du fuseau mitotique au cours de la phase M (SAC). Même si ces cinq checkpoints diffèrent de par la nature des événements moléculaires qui ont lieu et le moment du cycle cellulaire où ils ont lieu, un mécanisme générique se dégage : *un checkpoint empêche tout événement qui initie une phase donnée de se réaliser tant que tous les événements de la phase précédente ne sont pas correctement terminés*. Cet énoncé met explicitement en évidence le besoin de définir ce qu’est une phase préalablement à la formalisation d’un checkpoint. Le cycle cellulaire a par exemple été formalisé par un triplet de Hoare dans le chapitre précédent, ou par une composante fortement connexe terminale, pour ne citer que [28, 87]. La définition du cycle cellulaire est ici typiquement consécutive à celle d’un modèle donné et n’a pas été abordée d’un point de vue générique indépendant de sa formalisation spécifique. Ce chapitre est donc dédié à une première formalisation générique d’une phase qui sépare le concept d’une part, de la preuve de concept *via* l’instanciation d’un modèle d’autre part.

Dans un premier temps, une phase est définie de manière canonique par un triplet de Hoare de la forme $\{P\} p \{Q\}$ où p est une séquence d’événements moléculaires supposée complète, *cf.* section 6.2. L’*unique* séquence choisie correspond soit à un ordre canonique des événements, c’est à dire celui qui fait le plus consensus dans la littérature du domaine, soit à une hypothèse raisonnable formulée en vue répondre à la question suivante : *quels sont les événements qui peuvent définir une phase du cycle cellulaire ?* La seconde étape consiste à relâcher totalement la contrainte sur l’ordre des événements au sein d’une phase en considérant toutes les permutations de p . Il en découle une définition *par excès* d’une phase en section 6.3.1. La dernière étape vise à identifier les permutations de p admises au sein d’un RRBM *via* le calcul de la plus faible précondition de la logique de Hoare génétiquement modifiée, *cf.* section 6.3.2. Enfin la preuve de notre concept de phase est réalisée en partant de :

- l’ensemble des 16 modèles solutions obtenus à la fin du chapitre 5 (section 5.5.3 page 5.5.3)
- la spécification du cycle cellulaire canonique H_c (chapitre 5 page 97).

6.2 Définition syntaxique d'une phase canonique

Le cycle cellulaire est décrit comme une séquence d'événements cellulaires subdivisée en phases G1, S, G2 et M. Dans le chapitre 5.4, nous avons défini un GRBM dédié à la représentation des phases du cycle cellulaire. Ensuite, on a choisi de voir le cycle cellulaire, et par extension une phase, comme une séquence canonique d'évènements de régulation observée à partir d'un état initial. L'hypothèse *a priori* qui sous-tend ce choix est qu'une fois que les évènements de régulation associés à une phase se sont réalisés, les évènements cellulaires caractéristiques de la phase ont nécessairement eu lieu. En effet, tous les seuils de nos variables décrivent l'action des régulateurs du cycle cellulaire sur les évènements cellulaires caractéristiques d'une phase, cf. section 5.4.2 page 92. Toujours dans le chapitre 5, la syntaxe de logique de Hoare génétiquement modifiée s'est avérée adaptée à la représentation de ces évènements de régulation. On rappelle qu'un évènement de la LHGM est de la forme $v+$ ou $v-$ et signifie qu'une variable franchit l'un de ses seuils de régulation. Nous proposons de définir une phase canonique comme un cas particulier de la syntaxe de logique de Hoare génétiquement modifiée LHGM, introduite par [64], que l'on nomme *triplet de Hoare élémentaire*.

Définition 26: Chemin syntaxique élémentaire

Étant donné un GRBM \mathcal{G} , un *chemin syntaxique élémentaire* p est une expression de la forme $p = e_1; \dots; e_n$, où les e_i sont des évènements de \mathcal{G} , et on note $|p| = n$ sa longueur.

Définition 27: Triplet de Hoare élémentaire

Étant donné un GRBM \mathcal{G} , un *triplet élémentaire* $H = \{P\} p \{Q\}$ est un triplet de Hoare tel que :

- P définit un unique état associé à \mathcal{G} ,
- p est un chemin syntaxique élémentaire pour \mathcal{G} ,
- Q est la post-condition obtenue de manière unique à partir de P en suivant le chemin p .

Notation 1: Phase canonique

Une *phase canonique* π est caractérisée syntaxiquement par le triplet de Hoare élémentaire $H_\pi = \{P_\pi\} p_\pi \{Q_\pi\}$, où $\pi \in \{G1, S, G2, M\}$ et p_π est le chemin canonique de la phase, c'est à dire la séquence d'évènements moléculaires qui fait le plus consensus dans la littérature ou dans le cadre d'une hypothèse à tester.

Remarque 4. Pour passer d'une définition syntaxique d'une phase canonique de la forme $H_\pi \equiv \{P_\pi\} p_\pi \{Q_\pi\}$ à sa définition sémantique, H_π doit d'abord être satisfiable :

$$\exists(\mathcal{G}, \sigma) \mid \forall \eta_{P_\pi} \models P_\pi, \sigma, \eta_{P_\pi} \models_{wp} (p_\pi, Q_\pi)$$

Une phase canonique est ensuite caractérisée d'un point de vue sémantique par l'union des états parcourus par p_π depuis son état initial η_{P_π} et jusqu'à son état final. Dans ce cas, la succession des états qui caractérisent π sémantiquement existe dans le GTA déduit du modèle (\mathcal{G}, σ) , cf. 6.4.1 pour illustration.

Notation 2: État checkpoint

Considérons pour finir deux phases successives du cycle cellulaire, notées respectivement π_i et π_{i+1} , avec $H_{\pi_i} \equiv \{P_{\pi_i}\} p_{\pi_i} \{Q_{\pi_i}\}$ et $H_{\pi_{i+1}} \equiv \{P_{\pi_{i+1}}\} p_{\pi_{i+1}} \{Q_{\pi_{i+1}}\}$. L’état final de π_i qui satisfait Q_{π_i} satisfait également $P_{\pi_{i+1}}$, et est noté $\eta_{\pi_i/\pi_{i+1}}$. Il est qualifié d’état *checkpoint* étant donné qu’il satisfait la réalisation de tous les événements du chemin p à partir de l’état satisfaisant P_{π_i} , si H_{π} est satisfiable.

Remarque 5. Lorsque chaque phase $\pi \in \{G1, S, G2, M\}$ est définie par un triplet de Hoare élémentaire, on en déduit une définition syntaxique et sémantique du cycle cellulaire canonique, noté H_C . Son état initial est caractérisé par la précondition de $G1$ notée P_{G1} . S’agissant d’un cycle, sa postcondition est équivalente à sa précondition et le chemin canonique est défini de sorte que se succèdent dans l’ordre les quatre phases à partir de P_{G1} . Le triplet de Hoare élémentaire résultant est de la forme : $H_C \equiv \{P_{G1}\} p_{G1}; p_S; p_{G2}; p_M \{P_{G1}\}$. Prouver la validité de chacune des phases canoniques suffit à prouver la validité du cycle cellulaire canonique H_C . En effet, si H_{G1} , H_S , H_{G2} et H_M sont valides alors H_C est valide car $p_{G1}; p_S; p_{G2}; p_M$ est une composition séquentielle de la logique de Hoare modifiée, cf. définition 22.

Les définitions syntaxiques de nos *phases canoniques* pour le graphe \mathcal{G}_c sont reportées en section 6.4.1 et inaugure notre preuve de concept.

6.3 Formalisation d’une phase à partir de sa définition canonique

6.3.1 Définition de l’hyperpavé d’un triplet de Hoare

La définition d’une phase canonique permet d’exhiber un multi-ensemble des événements qui doivent avoir lieu au cours d’une phase π avant le début de la phase suivante π_{i+1} . Ce multi-ensemble associe à chaque événement possible de la forme $v+$ ou $v-$ son nombre d’occurrence dans π . Dans le cas où aucune contrainte sur l’ordre des événements au sein de π n’est connue, nous proposons de définir *par excès* une phase par l’hyperpavé associé à son triplet de Hoare élémentaire.

Définition 28: Hyperpavé d’un triplet de Hoare élémentaire

Étant donné un triplet de Hoare élémentaire $H \equiv \{P\} p \{Q\}$ dans un GRBM $\mathcal{G} = (V, M, E)$ et une variable v de \mathcal{G} , la borne inférieure (resp. supérieure) de v pour H est définie comme suit :

- soit $\#_v^-$ (resp. $\#_v^+$), le nombre d’occurrence de l’événement $v-$ (resp. $v+$) dans p ,
- soit η l’état initial caractérisé par la formule P , la borne inférieure (resp. supérieure) de v pour H est par définition $\min_v^H = \max(0, \eta(v) - \#_v^-)$ (resp. $\max_v^H = \min(b_v, \eta(v) + \#_v^+)$).

L’hyperpavé de H est alors le sous-ensemble de ζ correspondant au produit cartésien $h(H) = \prod_{v \in V} [\min_v^H, \max_v^H]$ et on note $\psi_H \equiv \bigwedge_{v \in V} (v \geq \min_v^H \wedge v \leq \max_v^H)$ sa propriété caractéristique.

Considérons à titre d'exemple le triplet de Hoare élémentaire suivant :

$$H_\pi \equiv \{x = 0 \wedge y = 0 \wedge z = 0\} x+; y+; x-; y+ \{x = 0 \wedge y = 2 \wedge z = 0\}$$

qui définit une phase π à partir du graphe d'influence jouet de la figure 6.1. Le comptage des

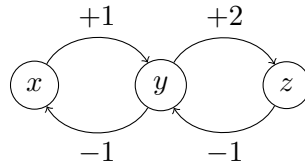
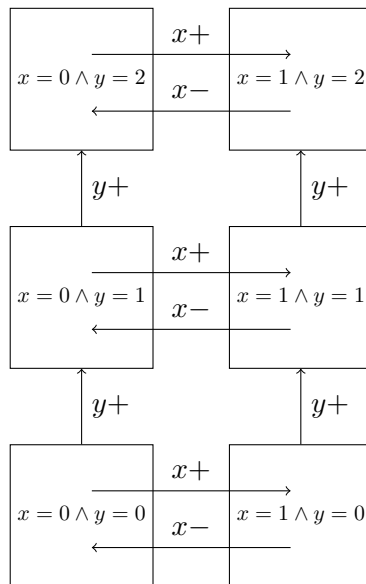


Figure 6.1 – Un graphe d'influence pour représenter les évènements d'une phase π . Seules les deux variables x et y importent pour représenter π et l'on considère que z importe pour une autre phase.

évènements appartenant au chemin de H_π permet de déduire le multi-ensemble associé à π : $\{x+ : 1, x- : 1, y+ : 2\}$. La variable z n'importe pas pour définir le domaine de π puisqu'aucun évènement impliquant z n'est observé dans p_π . Ainsi $\min_z^{H_\pi} = \max_z^{H_\pi}$. Les variables x et y varient respectivement de $\min_x^{H_\pi} = 0$ à $\max_x^{H_\pi} = 1$ et de $\min_y^{H_\pi} = 0$ à $\max_y^{H_\pi} = 2$ donc la formule caractéristique du domaine de π est $\psi_H = (x \geq 0 \wedge x \leq 1) \wedge (y \geq 0 \wedge y \leq 2) \wedge (z = 0)$. L'hyperpavé est ici inclus dans l'hyperplan ($z = 0$). L'hyperpavé $h(H)$ exhibe toutes les permutations de p :



où tous les évènements sont piochés sans remise dans le multi-ensemble associé à π . Les définitions par excès des phases G1, S, G2 et M sont appliquées à notre modèle du cycle cellulaire en section 6.4.2.

Au bilan : étant donné une phase du cycle cellulaire π , on considère classiquement que seul

l’ensemble des événements qui la constituent importe. Dans ce cadre, toutes les permutations de ses événements qui ne sortent pas des bornes sont à considérer *a priori*. Par conséquent $h(H_\pi)$ est l’ensemble des états que cette phase pourrait *potentiellement* couvrir. L’hyperpavé d’une phase peut être représenté par un GTA où les états appartenant à l’ensemble ζ sont reliés par des transitions correspondant aux événements qui caractérisent π . Pour illustrer l’approche, considérons un exemple jouet dans lequel notre phase G2 canonique, définie dans le chapitre précédent, montre une oscillation de la variable en :

$$\begin{aligned} & \{sk = 0 \wedge a = 1 \wedge b = 0 \wedge ep = 0 \wedge en = 1 \wedge gf = 1\} \\ & \quad a+; \mathbf{en-}; b+; \mathbf{en+} \\ & \{sk = 0 \wedge a = 2 \wedge b = 1 \wedge ep = 0 \wedge en = 1 \wedge gf = 1\} \end{aligned}$$

Dans ce cas, on remarque qu’il existe des permutations d’évènements qui induisent une sortie de borne de en : il s’agit de toute permutation telle que $en+$ précède $en-$ étant donné que $en = 1$ (sa valeur maximale) en précondition. Il s’agit d’une contradiction de l’axiome des bornes de la logique de Hoare génétiquement modifiée. Aucune sortie de borne n’est admise par un hyperpavé puisqu’une variable doit rester dans son domaine défini.

6.3.2 Vers une formalisation d’une phase sous la forme d’un prédicat du premier ordre

L’hyperpavé $h(H_\pi)$ permet de définir la phase π avec le minimum de contraintes. Autrement dit, aucune connaissance biologique *a priori* n’a été exploitée pour vérifier la faisabilité des permutations de p_π . Or, nous savons déjà qu’il existe des connaissances biologiques qui n’admettent pas toutes les permutations possibles. Certaines d’entre elles ne permettront pas d’atteindre Q_π à partir de P_π car certaines transitions ne sont pas admises par un paramétrage σ . Par exemple, le chapitre précédent démontre que notre ensemble solution de modèles du cycle cellulaire n’admet que $sk+$ comme événement initiateur de la phase G1, sachant que sa phase canonique est définie par le triplet :

$$\begin{aligned} & \{sk = 0 \wedge a = 0 \wedge b = 0 \wedge ep = 0 \wedge en = 1 \wedge gf = 1\} \\ & \quad sk+; en-; sk+ \\ & \{sk = 2 \wedge a = 0 \wedge b = 0 \wedge ep = 0 \wedge en = 0 \wedge gf = 1\} \end{aligned}$$

Toute permutation du chemin canonique avec $en-$ comme premier événement n’est pas admise biologiquement. À l’inverse il existe des permutations du chemin canonique d’une phase qui doivent être obligatoirement possibles au sein de notre modèle mathématique, ne serait-ce que le chemin canonique lui-même.

Pour que la formalisation d’une phase ne soit ni trop fortement ni trop faiblement contrainte, elle ne doit ni être définie par une phase canonique $H_\pi \equiv \{P_\pi\}p_\pi\{Q_\pi\}$ ni par son hyperpavé $h(H_\pi)$, mais par la partie *admissible biologiquement* de son hyperpavé et au vu du graphe de régulation. Le problème se résume à vérifier formellement la satisfiabilité de toutes les permutations de p_π exhibées par l’hyperpavé H_π .

Définition 29: Prédicat *estAdmis*

Étant donné un modèle $\mathcal{R} = (\mathcal{G}, \sigma)$, une phase canonique définie par un triplet élémentaire H de la forme $\{P\} p \{Q\}$ où P (resp. Q) est sa précondition (resp. postcondition) et p son chemin canonique, tout chemin p' est admis par \mathcal{R} si :

$$p' \in \text{permutations}(p) \quad \text{et} \quad \mathcal{I}(wp(p', Q)) \quad \text{où}$$

- $\text{permutations}(p)$ est l'ensemble des permutations p ,
- $wp(p', Q)$ est la fonction qui calcule la plus faible précondition de la logique de Hoare génétiquement modifiée associée à p' et Q
- \mathcal{I} est l'interprétation de cette plus faible précondition grâce à σ et η_P :

$$\mathcal{I} : \mathcal{K} \cup \zeta \rightarrow \mathbb{N}, \quad \mathcal{I}|_{\mathcal{K}} = \sigma \quad \text{et} \quad \mathcal{I}|_{\zeta} = \eta_P$$

Le prédicat $\text{estAdmis}_\sigma(p', H)$ est introduit : il est vrai si p' est admis par \mathcal{R} à partir de l'état caractérisé par P dans H .

De toute évidence, on ne peut considérer un RRBM $\mathcal{R} = (\mathcal{G}, \sigma)$ comme valide que si le chemin canonique de chaque phase fait partie de l'ensemble des permutations admises par \mathcal{R} . Ce dernier abstrait un ensemble de connaissances que nous pouvons *interroger* formellement, à l'aide de notre prédicat $\text{estAdmis}_\sigma(p', H)$, pour répondre à la requête suivante : *quelles sont les permutations p' du chemins canonique de H qui n'entrent pas en contradiction avec les paramétrages σ ?* Rappelons que σ désigne toute solution d'un ensemble de propriétés biologiques dynamiques encodées, dans le contexte de la thèse, soit avec CTL soit avec la logique de Hoare génétiquement modifiée. On considère que le qualificatif *admissible biologiquement* est un équivalent d'*admissible par ensemble de modèle* (\mathcal{G}, σ) .

Remarque 6. Grâce aux deux définitions précédentes, on pourra mieux préciser le domaine d'une phase au sein de l'espace des états. Il s'agit de l'union des états atteignables par les permutations de p_π , admises par \mathcal{R} . Naturellement le domaine d'une phase dépend du paramétrage σ et rappelons que l'ensemble de modèles solutions de propriétés biologiques, lorsqu'il est non vide, est rarement réduit à un singleton. On pourra in fine définir un sur-ensemble a priori plus précis que l'hyperpavé en prenant l'union des domaines de la phase π pour tous les paramétrages à considérer. L'ensemble résultant est appelée la partie admise biologiquement de l'hyperpavé de π .

Les parties admises biologiquement par les 16 paramétrages σ des hyperpavés de nos quatre phases du cycle cellulaire sont calculées en section 6.4.3.

6.4 Application à notre modèle du cycle cellulaire

6.4.1 Les phases canoniques de notre modèle

Une séquence canonique des événements du cycle cellulaire a été définie à partir d'un état d'une cellule en début de G1. On rappelle à la page 97 :

$$G1_{init} \equiv (sk = 0 \wedge a = 0 \wedge b = 0 \wedge ep = 0 \wedge en = 1 \wedge gf = 1)$$

L'état initial du cycle cellulaire est l'état initial de G1 et $G1_{init}$ est renommé P_{G1} dans le contexte de formalisation des *phases*. Un triplet de Hoare élémentaire a servi à définir le cycle cellulaire canonique duquel on déduit des phases canoniques.

$$H_c \equiv \{P_{G1}\} \underbrace{sk+; en-; sk+; a+; sk-; sk-; a+; b+; b+; ep+; a-; a-; b-; b-; ep-; en+}_{p_c} \{P_{G1}\}$$

Compte tenu des évènements requis dans chaque phase, le chemin p_C a été décomposé en quatre chemins élémentaires : $p_C = p_{G1}; p_S; p_{G2}; p_M$. L'expression de p_{G1} , p_S , p_{G2} et p_M est donnée en table 6.1. Ces quatre chemins sont chacun une sous-séquence du chemin canonique p_c donc sont également canoniques. L'état initial de S, G2 et M est déductible de P_{G1} ainsi que de p_{G1} , p_S et p_{G2} .

| | |
|-------------|--|
| P_{π_i} | $P_{G1} \equiv (sk = 0 \wedge a = 0 \wedge b = 0 \wedge ep = 0 \wedge en = 1 \wedge gf = 1)$ $P_S \equiv (sk = 2 \wedge a = 0 \wedge b = 0 \wedge ep = 0 \wedge en = 0 \wedge gf = 1)$ $P_{G2} \equiv (sk = 0 \wedge a = 1 \wedge b = 0 \wedge ep = 0 \wedge en = 0 \wedge gf = 1)$ $P_M \equiv (sk = 0 \wedge a = 2 \wedge b = 1 \wedge ep = 0 \wedge en = 0 \wedge gf = 1)$ |
| p_{π_i} | $p_{G1} \equiv sk+; en-; sk+$ $p_S \equiv a+; sk-; sk-$ $p_{G2} \equiv a+; b+$ $p_M \equiv b+; ep+; a-; a-; b-; b-; en+; ep-$ |
| H_{π_i} | $H_{G1} : \{P_{G1}\} p_{G1} \{P_S\}$ $H_S : \{P_S\} p_S \{P_{G2}\}$ $H_{G2} : \{P_{G2}\} p_{G2} \{P_M\}$ $H_M : \{P_M\} p_M \{P_{G1}\}$ |

Table 6.1 – Les phases canoniques définies à l'aide du GRBM \mathcal{G}_c en figure 5.18 et de H_c .

On notera désormais $\eta_{M/G1}, \eta_{G1/S}, \eta_{S/G2}$ et $\eta_{G2/M}$ les états initiaux des phases caractérisés respectivement par P_{G1}, P_S, P_{G2} et P_M . Enfin, les triplets de Hoare H_{G1}, H_S, H_{G2} et H_M définissent respectivement les phases canoniques G1, S, G1 et M associées à notre GRBM que l'on notera \mathcal{G}_c . Chaque phase $\pi_i \in \{G1, S, G2, M\}$ est pour l'instant définie par un triplet de Hoare élémentaire de la forme $\{H_{\pi_i} \equiv P_{\pi_i}\} p_{\pi_i} \{Q_{\pi_i}\}$. La postcondition Q_{π_i} est égale à la précondition de la phase suivante notée $P_{\pi_{i+1}}$ ce qui permet de raccorder les phases entre elle à l'aide d'un état checkpoint au sens de la notation 2.

Les phases canoniques ont été définies uniquement en terme de syntaxe jusque-là. Le triplet H_c à l'origine de ces définitions syntaxiques a été prouvé valide dans le chapitre précédent. Pour définir sémantiquement nos phases canoniques, il est nécessaire de prouver la validité de H_{π_i} pour tout $\pi_i \in \{G1, S, G2, M\}$.

Proposition 1. *La correction de H_C implique la correction des 4 triplets H_{G1}, H_S, H_{G2} et H_M .*

Démonstration.

Si H_C est correct alors tout préfixe de p_C est un chemin faisable pour \mathcal{R}_C à partir de η_{G1} , en particulier p_{G1} . Ainsi en partant de η_{G1} et en suivant p_{G1} on aboutit forcément à η_S , ce qui

implique par définition que le triplet $\{P_{G1}\} p_{G1} \{P_S\}$ est correct. De même le triplet $H' = \{P_S\} p_S; p_{G2}; p_M \{P_{G1}\}$ est correct. Le même argument s'applique à partir de H' et prouve la correction de H_S ainsi que la correction de $H'' = \{P_{G2}\} p_{G2}; p_M \{P_{G1}\}$. Enfin une itération immédiate donne la correction de H_{G2} et H_M .

□

Remarque 7. En section 5, nous mentionnions que si H_{G1} , H_S , H_{G2} et H_M sont corrects alors, en appliquant la règle de composition séquentielle de la logique de Hoare, H_C est prouvé correct. La correction de H_C est donc équivalente à la correction des 4 triplets H_{G1} , H_S , H_{G2} et H_M .

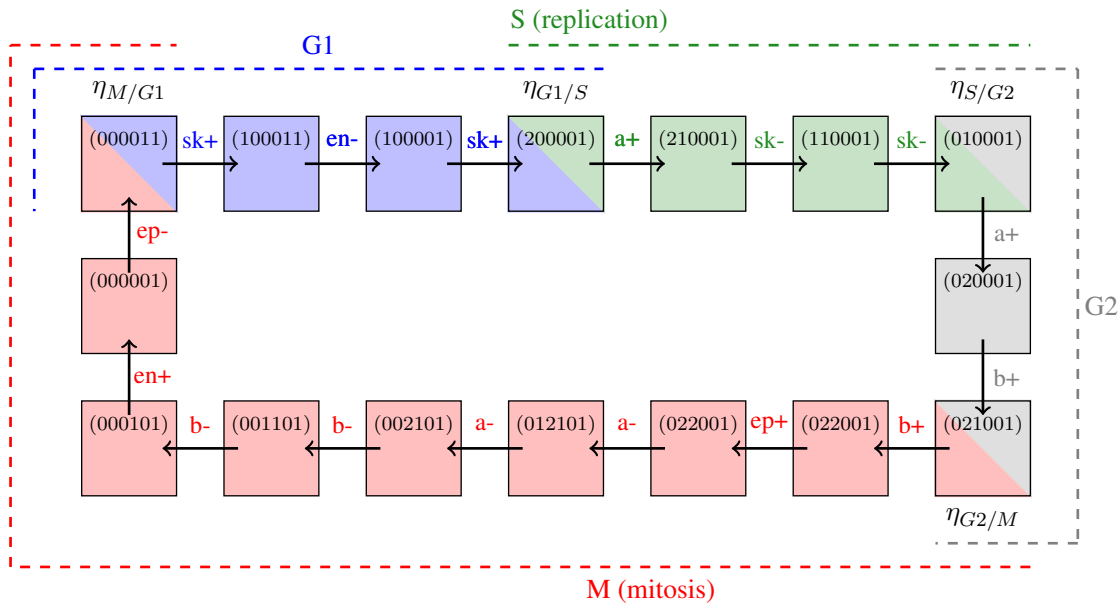


Figure 6.2 – Les phases canoniques H_π associées à \mathcal{G}_C pour tout $\pi \in \{G1, S, G2, M\}$. Pour des raisons de visibilité, un état est caractérisé ici par un vecteur de la forme : $(\eta(sk) \ \eta(a) \ \eta(b) \ \eta(ep) \ \eta(en) \ \eta(gf))$. Par exemple, (000011) est équivalent à P_{G1} . Les états doublement colorés sont les états checkpoint de notre cycle cellulaire au sens de la notation 2.

6.4.2 Les hyperpavés de nos phases canoniques

Les hyperpavés de nos phases canoniques sont les suivants :

La phase G1. L'hyperpavé $h(H_{G1})$ est inclus dans l'intersection des hyperplans $(a = 0 \wedge b = 0 \wedge ep = 0)$ et $(gf = 1)$, tandis que sk varie de 0 à 2 et en de 1 à 0. L'état initial de G1 est le coin $(sk = 0 \wedge en = 1)$ de $h(H_{G1})$ et son état final $\eta_{G1/S}$ correspond au coin opposé $(sk = 2 \wedge en = 0)$. La propriété caractéristique de $h(H_{G1})$ est donc définie par :

$$\psi_{H_{G1}} \equiv (sk \geq 0 \wedge sk \leq 2) \wedge (en \geq 0 \wedge en \leq 1) \wedge (a = 0 \wedge b = 0 \wedge ep = 0 \wedge gf = 1)$$

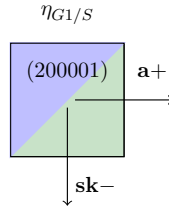


Figure 6.3 – **L'état checkpoint de notre phase canonique G1.** L'état final de la phase G1 (coloré en bleu) est également l'état initial de la phase S (coloré en vert) et est noté $\eta_{G1/S}$. Cet état vérifie que tous les événements de $G1$ ont bien eu lieu à partir de son état initial, à savoir dans ce cas deux $sk+$ et un $en-$. Deux transitions sont *a priori* possibles pour initier la phase S, d'après $h(H_S)$.

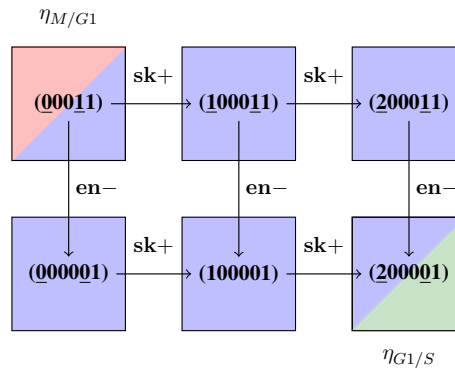


Figure 6.4 – **L'hyperpavé de la phase G1 définie canoniquement par H_{G1} .** De manière similaire à la figure 6.2, un état est représenté par le vecteur d'états discrets : $(\eta(sk) \eta(a) \eta(b) \eta(ep) \eta(en) \eta(gf))$.

Toutes les permutations de p_{G1} sont à considérer puisqu'aucune n'induit de sortie des bornes des variables. Les états appartenant à l'hyperpavé $h(H_{G1})$ et les transitions entre ces états sont représentés graphiquement en figure 6.4. L'état checkpoint $\eta_{G1/S}$ est représenté en figure 6.3.

La phase S. L'hyperpavé $h(H_S)$ est inclus dans l'intersection des hyperplans $(b = 0 \wedge ep = 0 \wedge en = 0)$ et $(gf = 1)$, alors que sk varie cette fois de 2 à 0 et a de 0 à 1. L'état initial de S est le coin $(sk = 2 \wedge a = 0)$ de $h(H_S)$ et son état final $\eta_{S/G2}$ correspond au coin opposé $(sk = 0 \wedge a = 1)$. Toutes les permutations de p_S sont exhibées par l'hyperpavé $h(H_S)$ en figure 6.5. La propriété caractéristique de $h(H_S)$ est donc définie par :

$$\psi_{H_S} \equiv (a \geq 0 \wedge a \leq 1) \wedge (sk \geq 0 \wedge sk \leq 2) \wedge (b = 0 \wedge ep = 0 \wedge en = 0 \wedge gf = 1)$$

Remarque 8. Les domaines de nos phases canoniques ne sont pas mutuellement exclusifs puisque que l'état final d'une phase π est également l'état initial de la phase π_{i+1} . Ce cas exclu, leurs domaines seraient tous disjoints, cf. figure 6.2. Naturellement les états appartenant à l'intersection des deux hyperpavés incluent l'état qui caractérise à la fois l'état final de G1 et l'état initial de S. L'intersection des deux hyperpavés inclue également l'état caractérisé par (000001) dans les figures 6.4 et 6.5.

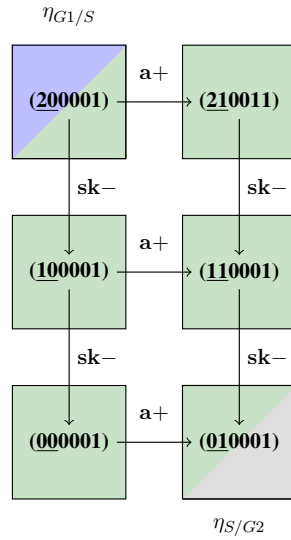


Figure 6.5 – L'hyperpavé de la phase S défini canoniquement par le triplet H_S (cf. table 6.1).

La phase G2. L'hyperpavé $h(H_{G2})$ est inclus dans l'intersection des hyperplans ($sk = 0 \wedge ep = 0 \wedge en = 0$) et ($gf = 1$), tandis que seules les variables a et b varient : a de 1 à 2 et b de 0 à 1. L'état initial η_{G2} est le coin ($a = 1, b = 0$) de $h(H_{G2})$ et l'état final $\eta_{G2/M}$ est le coin opposé ($a = 2, b = 1$). Toutes les permutations de p_{G2} sont à considérer dans l'hyperpavé $h(H_{G2})$ représenté en figure 6.6. La propriété caractéristique de $h(H_{G2})$ est définie par :

$$\psi_{H_{G2}} \equiv (a \geq 1 \wedge a \leq 2) \wedge (b \geq 0 \wedge b \leq 1) \wedge (sk = 0 \wedge ep = 0 \wedge en = 0 \wedge gf = 1)$$

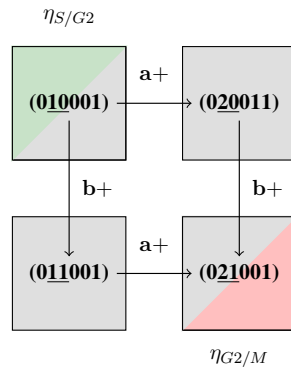


Figure 6.6 – L'hyperpavé $h(H_{G2})$ de la phase G2 défini canoniquement par le triplet H_{G2} (cf. table 6.1).

La phase M : L'hyperpavé $h(H_M)$ est inclus dans l'intersection des hyperplans ($sk = 0, gf = 1$) et ($gf = 1$), tandis que a varie de 2 à 0, ep varie de 0 à 1 puis de 1 à 0, en varie de 0 à 1, et b varie potentiellement entre 0 et 2 partant de $b = 1$. L'état initial η_M , où $b = 1$, n'est donc pas un

coin de $h(H_M)$. Selon la permutation des événements, b peut éventuellement ne varier qu'entre 0 et 1. C'est le cas de toute permutation incluant $(b-; b+; b-)$, peu importe l'intercalément des autres événements de M . L'état final est un coin de $h(H_M)$ ($a = 0, b = 0, ep = 0, en = 1$) qui est η_{G1} , l'état initial de la phase G1. Toutes les permutations de p_M ne sont pas à considérer. Celles qui autorisent une sortie de borne des variables b et ep ne doivent pas être prises en compte. À partir de l'état initial caractérisé par (021001) (cf. figure 6.2), $b+$ soit avoir lieu avant le second $b-$ et $ep+$ avant $ep-$. La propriété caractéristique de $h(H_M)$ est définie par :

$$\psi_{H_M} \equiv (a \leq 2 \wedge a \geq 0) \wedge (b \leq 2 \wedge b \geq 0) \wedge (en \leq 1 \wedge b \geq 0) \wedge (ep \leq 2 \wedge b \geq 0) \wedge (sk = 0 \wedge gf = 1)$$

6.4.3 Les phases de notre modèle

6.4.3.1 Implémentation du prédicat $estAdmis_\sigma$ en Prolog

La définition de la partie admissible de $h(H_{G1})$, $h(H_S)$, $h(H_{G2})$ et $h(H_M)$ par les modèles (G_c, σ) revient en pratique à identifier les solutions du prédicat $estAdmis_\sigma(p'_\pi, H_\pi)$ pour tout $\pi \in \{G1, S, G2, M\}$. Le langage logique Prolog a été choisi car $estAdmis$ repose sur la déclaration d'une clause de Horn qui traduit directement la formule de la définition 29. Prolog prévoit dans son architecture une stratégie de résolution pour les clauses de Horn. La déclaration du problème est la seule préoccupation dans la conception d'un programme Prolog.

La formule suivante en logique propositionnelle est une clause de Horn :

$$p' \in permutations(p) \text{ et } \mathcal{I}(wp(p', Q)) \implies estAdmis_\sigma(p', H)$$

où l'on rappelle que H est un triplet de Hoare élémentaire formé par P , Q et p respectivement une précondition, une postcondition et un chemin canonique. Cette formule est ensuite réécrite directement sous la forme d'un prédicat Prolog :

$$estAdmis(H, \mathcal{G}, \Sigma, p') : - permutation(p, p'), wp(p', Q, \Sigma).$$

Ce prédicat prend en entrée H , un GRBM noté \mathcal{G}^* et Σ , l'ensemble des paramétrisations σ qui satisfont des contraintes biologiques ($\phi_{K_{v,\omega}}$ au chapitre 2 page 34 pour ne citer qu'un exemple). On note que σ est un terme caché dans \mathcal{I} (cf. définition 29). Les solutions p' sont calculées à l'aide de l'instruction `findall` de SWI-Prolog [104] :

$$findall([p'], estAdmis(H, \mathcal{G}, \Sigma, p')).$$

La résolution de la plus faible précondition (wp) est déléguée à l'outil HoareFol [89] intégré dans TotemBioNet [96]. Le triplet $\{P\} p' \{Q\}$ est prouvé correct pour la paramétrisation σ si :

$$\forall \eta_P \models P, (\sigma, \eta_P) \models wp(p', Q)$$

Plus précisément, le prédicat wp est formé :

- d'un prédicat faisant un appel système de TotemBioNet avec pour argument son fichier d'entrée formé par :
 1. un bloc où les variables d'environnement sont instanciées[†],

*. Implicitement admis par le contexte de la définition 29 page 118.

†. Pour rappel, elles ne possèdent aucun prédécesseur et sont donc uniquement sous l'influence du modélisateur.

2. un bloc où les variables systémiques et leurs domaines sont définis,
 3. un bloc où les formules des multiplexes et les arcs de régulations sont déclarés,
 4. un bloc où les contraintes sur les valeurs de paramètres sont déclarées (la conjonction de ces contraintes caractérisent Σ),
 5. et un bloc où le triplet $\{P\} p' \{Q\}$ est déclaré.
- d'un prédicat qui parse le fichier de sortie de TotemBioNet en vue d'identifier le nombre de OK-models (cf. figure 4.7 page 76) noté n .
 - d'un prédicat qui vérifie que $n > 0$.

Nous ne donnerons pas plus de détails sur les types de données en entrée et autres subtilités d'implémentation. Pour en savoir plus, le lecteur se référera au lien suivant : <https://gitlab.com/totembionet/totembionet>. En Prolog, prouver la satisfiabilité du prédicat *estAdmis* revient à évaluer toutes les valeurs possibles de p' . Si le chemin canonique p contient n évènements alors la complexité en temps est $\Theta(n!)$. Il est donc indispensable de ne pas prendre en compte les répétitions d'un même évènement. Pour y remédier, le prédicat *permutation*(p) est remplacé par un nouveau prédicat : *permutationNR*(p). Ce dernier élimine les répétitions. La complexité devient $\Theta(\frac{n!}{k_1! \dots k_n!})$, où les k_i représentent le nombre d'occurrence d'un évènement. Les phases G1, S et G2 sont associées à très peu d'évènements, avec répétition de surcroît. Par conséquent, les résultats arrivent quasi instantanément (quelques secondes en moyenne). Pour la phase M, il faut compter 1h12min[‡] en moyenne.

6.4.3.2 Ordre des évènements totalement imposé pour les phases G1, S et G2

Cette section se focalise uniquement sur une application du prédicat *estAdmis* $_{\sigma}(p'_{\pi}, H_{\pi})$ pour $\pi \in \{G1, S, G2\}$. La phase M présente des résultats qui se démarquent des autres et sera par conséquent traitée dans une section dédiée. Les valeurs du paramètre H_{π} pour nos phases canoniques ont été définies plus haut (section 6.1). Pour ces trois phases, les connaissances biologiques accumulées dans les 16 occurrences de (\mathcal{G}_c, σ) n'admettent qu'une unique permutation du chemin canonique : la permutation identité. Cette dernière est nécessairement admise d'après la proposition 1. Aucune autre permutation n'est satisfiable car au moins un de leurs évènements n'est pas réalisable. Ces permutations ne reflètent donc pas une séquence d'évènements crédible d'un point de vue biologique.

Nous prenons la phase G1 comme exemple mais le même raisonnement s'applique aux deux suivantes. Les chemins $p'_{G1} = en-; sk+; sk+$ et $p''_{G1} = sk+; sk+; en-$ sont non admis biologiquement car $\mathcal{I}(wp(p'_{G1}, P_S)) \vee \mathcal{I}(wp(p''_{G1}, P_S))$ est insatisfiable. Le premier évènement de p'_{G1} ($en-$), est réalisable à partir de la précondition P_{G1} si et seulement si $\sigma(K_{en, \{no_sk, no_a, no_b\}}) < 1$, puisque la variable en doit être attirée par une valeur strictement plus petite que sa valeur courante. Cette contrainte entre en contradiction avec le paramétrage $\sigma(K_{en, \{no_sk, no_a, no_b\}}) = 1$, commun aux 16 paramétrages σ considérés.

In fine, nos phases canoniques H_{G1} , H_S et H_{G2} de la section 6.4.1 définissent aussi les phases G1, S et G2 dans nos modèles du cycle cellulaire. La figure 6.7 illustre la partie admise biologiquement de l'hyperpavé $h(H_{G1})$. Ces résultats montrent que la mise à jour de la composante discrète du modèle du cycle cellulaire de [69] a permis de contraindre totalement l'ordre des évènements

‡. Les temps CPU ont été mesurés sur un processeur Intel Core i7-8650U 1.90GHz \times 8 sous Ubuntu 20.04.4 LTS.

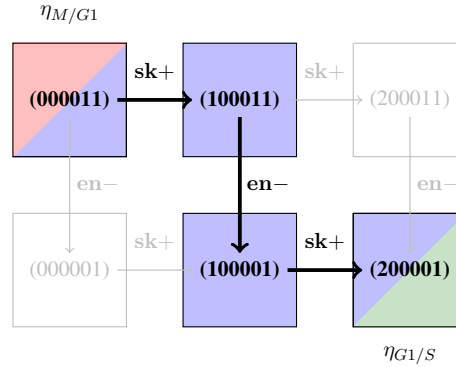


Figure 6.7 – La partie admise biologiquement de l'hyperpavé $h(H_{G1})$ est représentée en couleur. La partie non faisable est grisée. Cet exemple montre que la définition de la phase G1 se réduit ici à sa spécification canonique.

moléculaires au sein des phases G1, S et G2. Aucune information chronométrique a été nécessaire ce qui suggère que l'ordre des évènements au sein de ces phases se résume bien à une question de *chronologie* et de dépendance des évènements les uns aux autres. En revanche, il n'en est pas de même pour la phase M, plus particulièrement les connaissances biologiques actuelles n'imposent pas de contrainte forte sur la transition $M/G1$.

6.4.3.3 Le cas remarquable de la phase M

Nos 16 modèles (\mathcal{G}_c, σ) admettent plusieurs permutations du chemin canonique de la phase M. Plus exactement la figure 6.8 montre que 8 ordonnancements de p_M sont solutions du prédicat $estAdmis_\sigma(p'_M, P_{G1})$.

$$p'_M \in \left\{ \begin{array}{l} b+; ep+; a-; a-; b-; b-; ep-; en+ \\ b+; ep+; a-; a-; b-; b-; en+; ep- \\ b+; ep+; a-; a-; b-; ep-; b-; en+ \\ b+; ep+; a-; b-; a-; b-; en+; ep- \\ b+; ep+; a-; b-; a-; b-; ep-; en+ \\ b+; ep+; a-; b-; a-; ep-; b-; en+ \\ b+; ep+; a-; b-; b-; a-; ep-; en+ \\ b+; ep+; a-; b-; b-; a-; en+; ep- \end{array} \right\}$$

Figure 6.8 – Ensemble des permutations exhibées par $h(H_M)$ qui sont admises biologiquement.

D'après la proposition 1, le chemin canonique de M (en rouge) est nécessairement admis. Les connaissances accumulées au chapitre 5.4 page 5.4.2 ne permettent d'identifier un ordre total des évènements de M, en revanche nous avons identifié un ordre partiel qui doit être satisfiable :

1. La seule séquence qui peut initier la phase M est $b+; ep+$ à partir de son état initial.

2. La dégradation de la cycline A est initiée avant celle de la cycline B ce qui signifie que $a-$ doit toujours précéder $b-$.

Pour qu'une paramétrisation soit validée, elle doit invalider tout triplet $\{P_M\} p'_M \{P_{G1}\}$ tel que p'_M ne satisfait pas ces deux contraintes.

Toute permutation de p_M solution du prédicat $estAdmiss_\sigma$ possède bien le préfixe $b+; ep+$. Si l'on se réfère à la figure 6.9, la seule transition possible à partir de $\eta_{G2/M}$ (l'état initial de M) est celle qui incrémente la variable b . Le paramètre applicable de b en $\eta_{G2/M}$ est $K_{b,\{a2,no_Wee1,no_CI\}} = 2$. Pour toutes les autres variables v sauf la variable b , la valeur du paramètres applicable en $\eta_{G2/M}$ est égale à la valeur courante. La même justification s'applique à la transition $ep+$. Ensuite, nous avons prouvé qu'il existe bien un dépendance entre les évènements $a-$ et $b-$. Pour rappel, nous proposons l'explication mécaniste suivante : la destruction de la cycline A ($a-$) par APC-cdc20 (ep) débute avant celle de la cycline b ($b-$) car la première est indépendante du SAC alors que la seconde est *SAC-dépendante*. Le mécanisme de régulation biologique sous-jacent est encodé par le multiplexe phénoménologique nommé SAC (figure 5.18 page 102). Dès que l'action de APC-cdc20 est effective, la cycline A est dégradée ($a-$) mais pas la B. Au sein de notre phase M, le processus de dégradation de cette dernière ($b-$) est reportée au moment où $a \geq 2$ devient faux, c'est à dire après le premier $a-$.

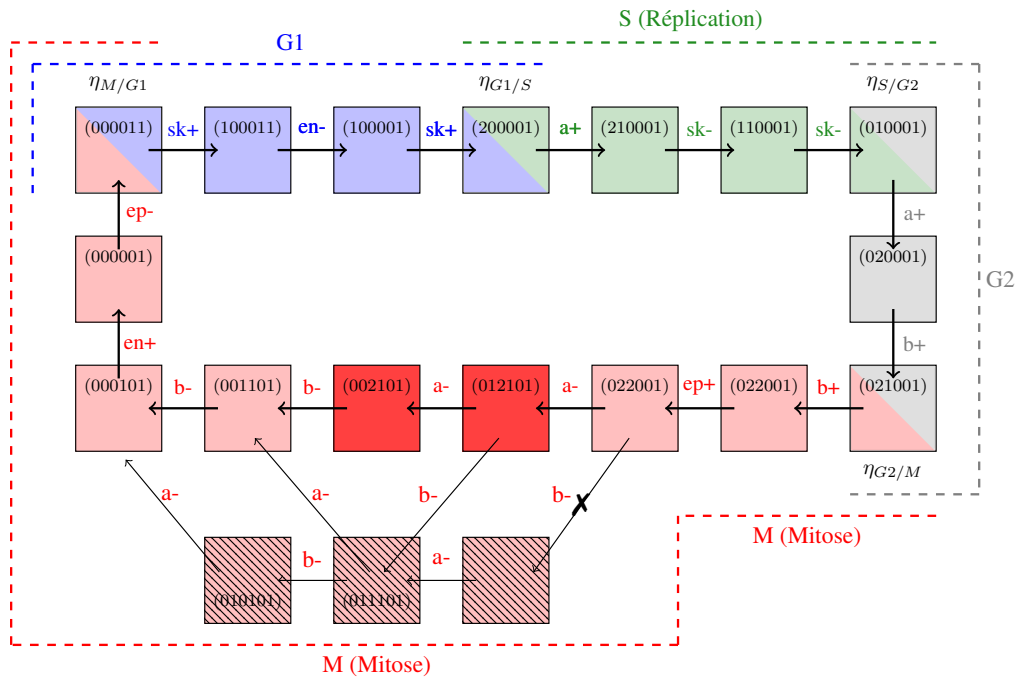


Figure 6.9 – **La dégradation asynchrone des cyclines en phase M et l'inactivation du SAC.** Le premier évènement $a-$ doit avoir lieu avant le premier $b-$ donc $b-$ ne doit pas être faisable à partir de l'état (022001). Notre GRBM vérifie cette propriété ce qui explique que tout chemin où le premier $b-$ a lieu avant le premier $a-$ a été rejeté. Les deux états en rouge foncé sont associés à l'inactivation du SAC.

En outre, il s'avère que le cycle cellulaire canonique introduit une annotation biologique d'un évènement qui appartient à la phase M et qui mérite une attention particulière. Il s'agit du premier $b-$ relié par transitivité à l'inactivation du SAC. C'est uniquement lorsque le mécanisme du SAC est désactivé, c'est à dire que les chromosomes sont tous correctement attachés au fuseau mitotique, que la dégradation de la cycline B (premier $b-$) peut être initiée par APC-cdc20 (ep). Selon le chemin admis considéré pour la phase M, l'état à l'origine de l'inactivation du SAC n'est pas le même (figure 6.9).

Au final, le début de la phase M est fortement contraint par les connaissances formalisées dans nos modèles : seul le préfixe $b+; ep+; a-$ est admis. Il n'en est pas de même pour la fin de la phase M. On peut expliquer la diversité des chemins solutions par un mode de mise à jour asynchrone des variables ou par un ensemble de modèles du cycle cellulaire qui accumulent peu de connaissances autour de la transition $M/G1$. Enfin, tous les chemins solutions du prédicat $estAdmis_\sigma$ sont admis par chacun des 16 modèles de départ. Le triplet de Hoare valide § qui permet de conclure à ce résultat est le suivant :

$$\{P_M\} \quad \forall[(b+; ep+; a-; a-; b-; b-; en+; ep-), \\ (b+; ep+; a-; a-; b-; b-; ep-; en+), \\ (b+; ep+; a-; a-; b-; ep-; b-; en+), \\ (b+; ep+; a-; b-; a-; b-; en+; ep-), \\ (b+; ep+; a-; b-; a-; b-; ep-; en+), \\ (b+; ep+; a-; b-; a-; ep-; b-; en+), \\ (b+; ep+; a-; b-; b-; a-; ep-; en+), \\ (b+; ep+; a-; b-; b-; a-; en+; ep-)]; \quad \{P_{G1}\}$$

Tous les résultats exposés sont disponibles à l'adresse suivante : <https://gitlab.com/deborahboyenval/temporal-separation-prototype/-/tree/master/ProofOfConceptCheckpoint>.

§. La correction du triplet qui suit est prouvée en 200 ms à l'aide de TotemBioNet sur un processeur Intel Core i7-8650U 1.90GHz × 8 sous Ubuntu 20.04.4 LTS.

CHAPITRE 7

Formalisation d'un checkpoint entre deux phases

| | | |
|------------|--|-----|
| 7.1 | Préambule | 131 |
| 7.2 | Le prédicat générique <i>séparation – temporelle</i>(π_i, π_{i+1}) | 132 |
| 7.2.1 | Les prédicats <i>peutClore</i> et <i>peutInitier</i> | 132 |
| 7.2.2 | Le prédicat <i>estRequis</i> | 133 |
| 7.3 | Irréversibilité des transitions de phases | 135 |
| 7.4 | Application à notre modèle et nos phases du cycle cellulaire | 137 |
| 7.4.1 | Implémentation du prédicat <i>séparation – temporelle</i> en Prolog | 137 |
| 7.4.2 | Notre graphe de régulation et ses phases munies de checkpoint | 138 |
| 7.4.3 | La preuve de concept | 140 |

7.1 Préambule

Le chapitre précédent définit une méthodologie formelle qui exhibe les permutations d'une séquence canonique du cycle cellulaire admises par un ensemble de modèles. Ces derniers reproduisent les propriétés classiques du cycle cellulaire, mais *qu'en est-il des checkpoints?* La séquence canonique d'une phase, représentée schématiquement en bleu dans la figure 7.1, exhibe un état checkpoint représenté quant à lui en vert. Il vérifie que tous les évènements associés à une phase ont eu lieu donc est un passage obligatoire pour passer à la phase suivante. Une certaine incertitude dans l'ordre des évènements au sein d'une phase est prise en compte et identifiée. L'objectif est ensuite d'identifier si un graphe de régulation donné, en l'occurrence celui exposé au chapitre 5, vérifie un checkpoint aux endroits cruciaux du cycle cellulaire. Ces checkpoints sont le passage du point de restriction et les transitions G1/S, S/G2, G2/M et métaphase/anaphase. La contrainte qu'exerce un checkpoint à ces endroits doit d'abord prendre en compte *tous* les évènements qui peuvent clore une phase ou portion de phase (notés E plus loin) et *tous* ceux qui peuvent initier la phase ou portion de phase suivante (notés E' plus loin). Un checkpoint vérifie *au minimum* que tout évènement E est requis avant tout évènement E' . La définition 31 formalise cette contrainte qui se base sur le prédicat *estAdmis*, défini dans le chapitre précédent, et introduit le prédicat *estRequis*.

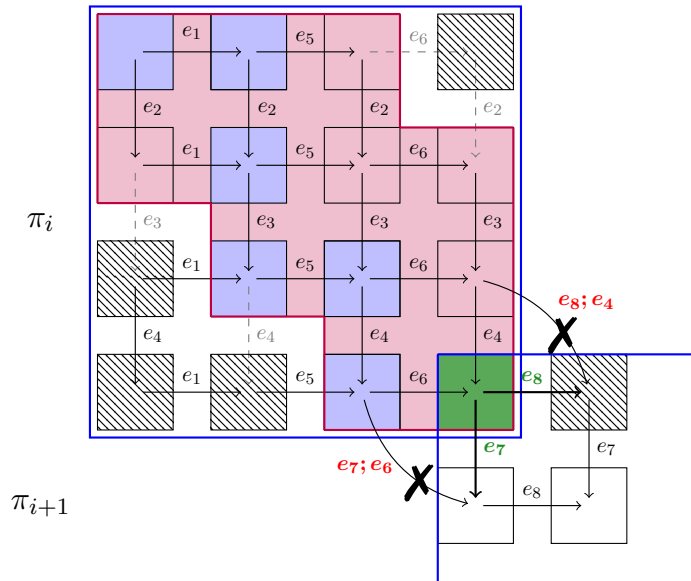


Figure 7.1 – **Schéma de la contrainte minimale pour un checkpoint intrinsèque.** On admet ici que les variables v et v' importent pour la phase π_i , et w et w' pour π_{i+1} . Les évènements e_1 , e_5 et e_6 correspondent à l'incrémentement de la variable v ($v+$). Les évènements e_2 , e_3 et e_4 correspondent à la décrémentation de la variable v' ($v'-$). Enfin l'évènement e_7 correspond à $w-$ et e_8 à $w'+$. La partie rose représente la partie admissible biologiquement de l'hyperpavé de π_i . Un checkpoint entre π_i et π_{i+1} vérifie que les permutations $(e_7; e_6)$, $(e_8; e_4)$, $(e_7; e_4)$ et $(e_8; e_6)$ ne sont pas admises par un ensemble de modèles. Le cas échéant, tous les évènements de $\{e_1, e_2, e_3, e_4, e_5, e_6\}$ sont requis avant ceux appartenant à $\{e_7, e_8\}$.

Le prédicat *estRequis* contraint la valeur de certains paramètres $K_{v,\omega}$ et permet sélectionner les modèles qui vérifient une première formalisation des checkpoints focalisée sur la séparation temporelle des phases. Si l'ensemble des événements de π_i n'est pas requis avant ceux de π_{i+1} alors π_i et π_{i+1} ne sont pas des phases séparées temporellement et ne permettent pas de représenter un cycle cellulaire sous contrôle.

La vérification de l'irréversibilité des transitions de phases suit la même démarche. Une formule définie en CTL équitable (définition 33) permet de sélectionner, parmi les modèles vérifiant que tout E est requis avant tout E' , ceux qui vérifient également l'irréversibilité des transitions π_i/π_{i+1} . L'insatisfiabilité du prédicat *séparation – temporelle*(π_i/π_{i+1}) (définition 31) permet de recentrer la question de modélisation sur la définition canonique des phases π_i et π_{i+1} , ou bien sur la définition du graphe de régulation ou encore sur le domaine des paramètres $K_{v\omega}$.

7.2 Le prédicat générique *séparation – temporelle*(π_i, π_{i+1})

7.2.1 Les prédicats *peutClore* et *peutInitier*

La définition au préalable des chemins d'une phase admis biologiquement à l'aide du prédicat *estAdmis*(p', H) trivialise l'identification des événements qui peuvent initier ou clore une phase définie par H . Il suffit d'identifier l'évènement en tête et en queue de tous les chemins p' qui sont solutions du prédicat *estAdmis*.

Définition 30: Prédicats *peutInitier* et *peutClore*

Étant donné un modèle $\mathcal{R} = (\mathcal{G}, \sigma)$, une phase canonique π définie par un triplet de Hoare élémentaire $H \equiv \{P\} p \{Q\}$ où P (resp. Q) est sa précondition (resp. postcondition) et p son chemin canonique, tout événement E' *peut initier* π si :

$$\exists p' \in \text{permutations}(p) \mid \text{estAdmis}_\sigma(p', H) \wedge E' = \text{premier}(p')$$

et tout événement E *peut clore* π si et seulement si :

$$\exists p' \in \text{permutations}(p) \mid \text{estAdmis}_\sigma(p', H) \wedge E = \text{dernier}(p')$$

On note *peutInitier* $_\sigma(E', \pi)$ et *peutClore* $_\sigma(E, \pi)$ les prédicats correspondants et les fonctions *premier*(p') et *dernier*(p') sont trivialement les premier et dernier événements de p' .

Le prédicat *estAdmis* $_\sigma(p', H)$ a été introduit en définition 29 (chapitre 6). D'après notre introduction, cinq checkpoints régulent le cycle cellulaire. Ainsi, six phases ou portions de phases sont à considérer au lieu des classiques quatre phases du cycle cellulaire.

$$G1_{\text{précoce}} \xrightarrow{\text{PPR}} G1_{\text{tardif}} \xrightarrow{G1/S} S \xrightarrow{S/G2} G2 \xrightarrow{G2/M} M_{\text{précoce}} \xrightarrow{M/A} M_{\text{tardif}} \rightarrow G1_{\text{précoce}}$$

Le passage du point de restriction (noté *PPR*) est un évènement qui sépare la phase G1 en deux

parties. On appelle $G1_{précocce}$ la partie de $G1$ qui précède le passage du point de restriction et $G1_{tardive}$ celle qui lui succède. Une concentration suffisante de facteurs de croissance est requise avant le passage de restriction et permet ainsi de contrôler l'entrée (irréversible) dans un cycle cellulaire. De la même manière, la transition métaphase/anaphase (notée M/A) est un évènement qui sépare la phase M en deux parties : $M_{précocce}$ et M_{tardif} .

Définition 31: Prédicat *séparation – temporelle*

Étant donné un graphe d'interactions \mathcal{G} et un ensemble de paramétrisations Σ de \mathcal{G} , les phases canoniques π_i étant ordonnées suivant la séquence $G1_{précocce}; G1_{tardif}; S; G2; M_{précocce}; M_{tardif}$, les checkpoints du cycle cellulaire sont satisfiables si les *phases sont séparées temporellement* :

$$\exists \sigma \in \Sigma \mid \forall i \in \llbracket 1, 5 \rrbracket,$$

$$\forall E, \forall E', [\text{peutCloure}_\sigma(E, \pi_i) \wedge \text{peutInitier}_\sigma(E', \pi_{i+1}) \implies \text{estRequis}_\sigma(E, E')]$$

On note *séparation – temporelle*(π_i, π_{i+1}) le prédicat correspondant.

La section suivante est dédiée à la définition du prédicat $\text{estRequis}_\sigma(E, E')$ qui assure que l'évènement E' ne peut pas avoir lieu avant l'évènement E . Autrement dit, E et E' sont les évènements clés du cycle cellulaire qui ne doivent pas être permutable.

Remarque 9. *Seuls les évènements qui peuvent initier ou clore une phase important dans la contrainte de séparation temporelle. De plus, tout évènement qui a lieu à partir un état qui n'appartient pas à la partie admissible biologiquement de l'hyperpavé d'une phase π_i est tout simplement ignoré.*

7.2.2 Le prédicat *estRequis*

Maintenant que l'on a identifié les évènements du cycle cellulaire qui ne doivent pas être permutable au sein d'un modèle \mathcal{R} , nous introduisons les notations nécessaires à la caractérisation de la contrainte estRequis_σ . Elle vérifie l'inexistence de la transition associée sémantiquement à l'évènement E' (*i.e.* celle qui peut initier une phase) à partir de l'état source de la transition associée à E (*i.e.* celle qui peut clore la phase précédente).

Notation 3: à propos du prédicat *estRequis*

Avec les notation de la définition 30, on note :

- η l'état source de E ,
- η' l'état cible de E , c'est à dire son état checkpoint déduit de la définition canonique d'une phase (*cf.* notation 2).
- v' (*resp.* v) est la variable associée à l'évènement E' (*resp.* E) de la forme $v'-$ ou $v'+$ (*resp.* $v-$ ou $v+$),
- $K_{v',\omega}$ est le paramètre applicable de v' en η (*cf.* définition 6),
- et $K_{v',\omega'}$ est celui applicable dans l'état η' .

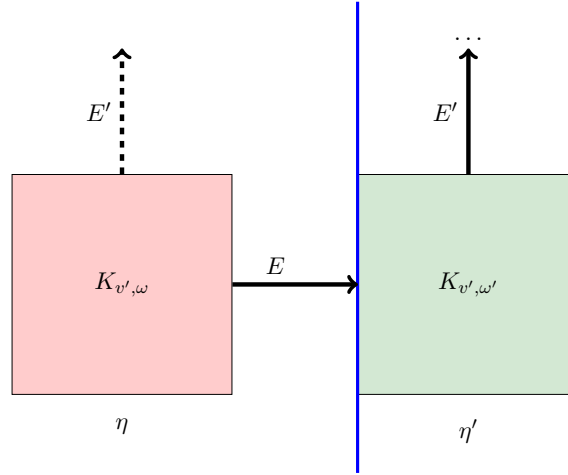


Figure 7.2 – **La notation 3 illustrée.** L'état η appartient exclusivement à une phase π_i et η' est l'état à l'origine de la transition vers la phase suivante. La contrainte exprimée par le prédicat *estRequis* se focalise uniquement sur les transitions qui peuvent initier la phase π_{i+1} : elle ne doit pas être permise dans l'état η . Ici nous illustrons l'approche avec une seule instanciation de E et E' . La bordure bleue représente la transition π_i/π_{i+1} et η' est l'état checkpoint de la phase π_i qui témoigne que tous les événements requis au cours de π_i ont eu lieu.

Définition 32: Prédicat *estRequis*

Étant donné un événement E qui peut clore une phase π_i et un événement E' qui peut initier la phase suivante π_{i+1} , E est requis avant E' dans un modèle $\mathcal{R} = (\mathcal{G}, \sigma)$ si et seulement si :

$$[\sigma(K_{v',\omega}) - \eta(v')] \times [\sigma(K_{v',\omega'}) - \eta'(v')] \leq 0$$

où $K_{v',\omega}$, $K_{v',\omega'}$, η , η' , et v' sont introduits en notation 3 et on rappelle que η , η' et σ sont des substitutions, cf. définitions 5 et 7 du chapitre 2.

On note $estRequis_\sigma(E', E)$ le prédicat correspondant.

Puisque, par définition, l'évènement qui peut initier la phase π_{i+1} est une solution du prédicat *peutInitier* alors la transition décrite par E' est nécessairement admise par au moins un modèle $\mathcal{R} = (\mathcal{G}, \sigma)$. On en déduit que $\sigma(K_{v',\omega'}) - \eta'(v') \neq 0$. L'apprentissage de modèles consiste donc essentiellement à identifier les substitutions σ du paramètre $K_{v',\omega}$ qui n'attirent pas v' dans le même sens que dans l'état η' :

- si $E' \equiv v' +$ alors $\sigma(K_{v',\omega}) \leq \eta'(v')$ puisque $\sigma(K_{v',\omega'}) > \eta'(v')$,
- si $E' \equiv v' -$ alors $\sigma(K_{v',\omega}) \geq \eta'(v')$ puisque $\sigma(K_{v',\omega'}) < \eta'(v')$.

$$\underbrace{[\sigma(K_{v',\omega}) - \eta(v')]}_{E' \text{ est-il admis à partir de } \eta?} \times \underbrace{[\sigma(K_{v',\omega'}) - \eta'(v')]}_{E' \text{ est admis à partir de } \eta'} \leq 0$$

7.3 Irréversibilité des transitions de phases

Un graphe de régulation du cycle cellulaire ne peut être véritablement cohérent biologiquement que si les transitions gouvernées par des checkpoints sont *irréversibles*. Autrement formulé, il ne doit exister aucun chemin qui relie une phase à une quelconque autre phase qui ne passe pas, dans l'ordre, obligatoirement par chacun des états checkpoints définis canoniquement.

Définition 33: Prédicat *irréversible*

Étant donné un graphe d'interactions \mathcal{G} et un ensemble de paramétrisations Σ de \mathcal{G} , les phases canoniques π_i étant ordonnées suivant la séquence $G1_{précocoe}; G1_{tardif}; S; G2; M_{précocoe}; M_{tardif}$, l'irréversibilité des transitions entre π_i et π_{i+1} est satisfiable si :

$$\exists \sigma \in \Sigma \mid \forall i \in \llbracket 1, 5 \rrbracket,$$

$$P_{\pi_i} \Rightarrow \mathbf{A} \left(\psi_{H_{\pi_i}} \vee \neg \left(\bigvee_{\pi' \in \Pi \setminus \pi_i} \psi_{H_{\pi'}} \right) \cup Q_{\pi_i} \right)$$

où :

- $\psi_{H_{\pi}}$ est la formule caractéristique du domaine d'une phase π (définition 28 du chapitre 6),
- P_{π_i} (resp. Q_{π_i}) caractérise l'état initial (resp. final) de π_i ,
- Π est l'ensemble des π_i pour tout $i \in \llbracket 1, 6 \rrbracket$.
- et la modalité $\mathbf{A}[\dots \cup \dots]$ est celle de CTL

On note $irréversible(\pi_i, \pi_{i+1})$ le prédicat correspondant à cette définition.

Cette définition est illustrée en figure 7.3 à la page suivante.

Notation 4: Formule caractéristique l'irréversibilité de la transition π_i/π_{i+1}

On note $\varphi_{irréversibilité}$ la formule en CTL : $(P_{\pi_i} \Rightarrow \mathbf{A}(\psi_{H_{\pi_i}} \vee \neg(\bigvee_{\pi' \in \Pi \setminus \pi_i} \psi_{H_{\pi'}}) \cup Q_{\pi_i}))$.

La formule $P_{\pi_i} \Rightarrow \mathbf{A}(\psi_{H_{\pi_i}} \cup Q_{\pi_i})$ vérifierait qu'une phase π_i est stable tant que l'état checkpoint caractérisé par Q_{π_i} n'est pas atteint. Cette conception d'un checkpoint n'est pas sans rappeler le motif conditionnellement stable de [86] (cf. page 52) qui rend une phase transitoirement stable tant que la condition du motif stable est satisfaite.

La formule $P_{\pi_i} \Rightarrow \mathbf{A}(\neg(\bigvee_{\pi' \in \Pi \setminus \pi_i} \psi_{H_{\pi'}}) \cup Q_{\pi_i})$ vérifie, en cas de sortie du domaine d'une phase π_i , qu'aucune autre phase n'est visitée tant que l'état checkpoint de π_i n'est pas atteint. Tout état qui ne satisfait pas la formule caractéristique d'une phase peut en revanche être visité. Nous prenons en compte, au même titre que l'étude de [87] (cf. page 47), les partitions *inconnues* dans la caractérisation d'un checkpoint. L'inclusion de ces partitions permet de ne pas définir un checkpoint par une contrainte inutilement forte.

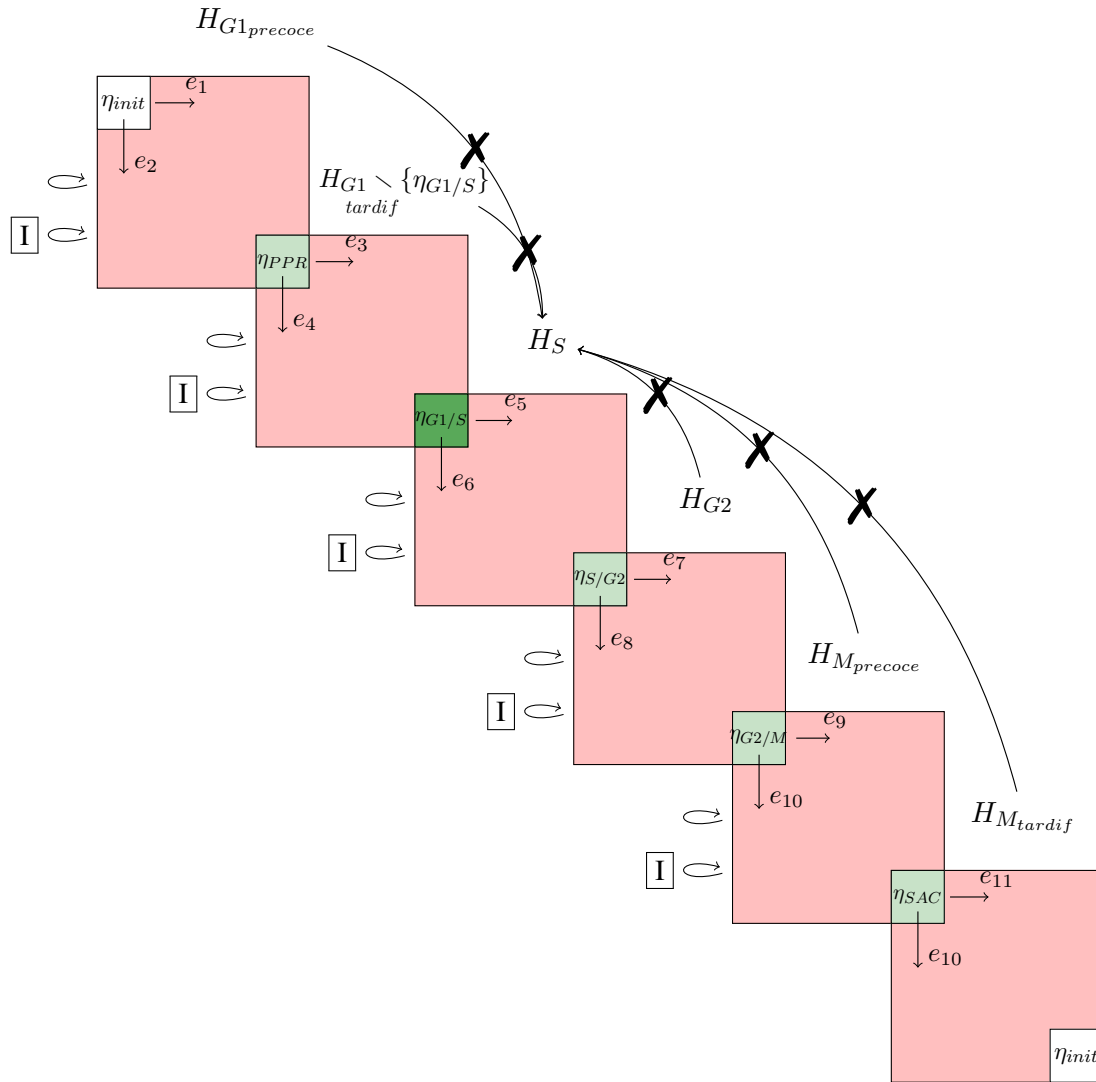


Figure 7.3 – La définition 33 illustrée avec le checkpoint G1/S. L'état initial du cycle cellulaire est représenté par η_{init} (en blanc). Chaque phase ou portion de phase, munie ou non d'un checkpoint, est représentée par un hyperpavé H_π pour tout $\pi \in \{G1_{précoce}, G1_{tardif}, S, G2, M_{précoce}, M_{tardif}\}$. Deux hyperpavés s'intersectent au niveau d'un état checkpoint (en vert) sauf $H_{M_{tardif}}$ et $H_{G1_{précoce}}$ qui s'intersectent au niveau de η_{init} . La portion tardive de la phase M n'est pas munie d'un checkpoint puisque le checkpoint M/G1 n'existe pas (cf. l'introduction). Les évènements qui peuvent initier une phase, à partir d'un état checkpoint, représentent les transitions de phases. Par exemple, e_3 ou e_4 représentent le passage du point de restriction, e_5 ou e_6 représentent la transition G1/S, et, e_9 ou e_{10} représentent la transition métaphase/anaphase. La transition G1/S ne peut s'effectuer que *via* l'état checkpoint $\eta_{G1/S}$ (vert foncé) donc la partie de l'hyperpavé de $G1_{tardif}$ qui exclut cet état doit être stable (boucles). La boucle étiquetée **I** signifie que l'espace des états inconnus, c'est à dire ceux qui ne satisfont aucune formule caractéristique d'une phase, peut être visité avant de revenir dans la phase en cours (représentée ici par la partie rose d'un hyperpavé).

Remarque 10. *La partie admissible biologiquement de l'hyperpavé d'une phase exhibe certains chemins admis par les connaissances biologiques. Certains états de l'hyperpavé peuvent ne pas être atteignables depuis l'état initial de la phase (illustrés par les carrés grisés en figure 7.1) mais cela n'implique pas que ces états ne permettent pas une sortie de phase. Pour cette raison, c'est l'hyperpavé d'une phase (caractérisé par ψ_{H_π}) qui est considéré dans la définition 33 et pas uniquement sa partie admissible biologiquement, comme dans la définition 31.*

In fine, les checkpoints entre phases adjacentes assurent leur *séparation temporelle* lorsque l'ensemble des événements requis pour ces phases ont lieu distinctement. C'est ce qu'assure le prédicat *séparation – temporelle*(π_i, π_{i+1}) de la définition 31, lorsqu'il est satisfiable. S'assurer que tout événement qui peut clore une phase a bien lieu avant tout ceux qui peut initier la phase suivante n'est pas une contrainte suffisante pour prouver qu'une phase est munie d'un checkpoint. Pour cela, il est nécessaire de vérifier additionnellement qu'il n'existe aucune transition, entre une phase π_i et sa phase voisine π_{i+1} , qui a lieu à partir d'un autre état que l'état checkpoint de π_i . C'est ce que vérifie le prédicat *irréversible*(π_i, π_{i+1}) de la définition 33, lorsqu'il est satisfiable. Ce prédicat assure l'*unidirectionnalité* du cycle cellulaire. La *séparation temporelle* des phases et l'*unidirectionnalité* du cycle cellulaire sont les deux caractéristiques essentielles à la caractérisation des checkpoints du cycle cellulaire, d'après l'étude de modélisation de [40]. Ainsi, un graphe de régulation \mathcal{G} admet des phases munies de checkpoints lorsque la condition suivante est vraie :

$$\exists \sigma \in \Sigma \mid \forall i \in \llbracket 1, 5 \rrbracket, \text{ séparation – temporelle}(\pi_i, \pi_{i+1}) \wedge \text{ irréversible}(\pi_i, \pi_{i+1}) \quad (7.1)$$

où :

- Σ est l'ensemble des paramétrisations associées à un GRBM \mathcal{G} ,
- π_i représente les phases (ou portions de phase) canoniques définies par un triplet de Hoare élémentaire (définition 27 du chapitre 6) et sont ordonnées suivant la séquence $G1_{\text{précocoe}}; G1_{\text{tardif}}; S; G2; M_{\text{précocoe}}; M_{\text{tardif}}$.

Et on appelle *checkpoint*(π_i, π_{i+1}) le prédicat correspondant.

7.4 Application à notre modèle et nos phases du cycle cellulaire

7.4.1 Implémentation du prédicat *séparation – temporelle* en Prolog

Le prédicat *séparation – temporelle* est une clause de Horn que l'on peut déclarer directement en Prolog.

$$\text{ séparation – temporelle}(H_{\pi_i}, H_{\pi_{i+1}}, \mathcal{G}, \Sigma, E, E') : -$$

$$\text{ peutClore}(H_{\pi_i}, \mathcal{G}, \Sigma, E), \text{ peutInitier}(H_{\pi_{i+1}}, \mathcal{G}, \Sigma, E'), \text{ estRequis}(H_{\pi_i}, H_{\pi_{i+1}}, \mathcal{G}, \Sigma, E, E').$$

et au préalable :

$$\begin{aligned} & \text{ dernier}(\lfloor _ \rfloor E, E). \\ & \text{ premier}(E \lfloor _ \rfloor, E). \\ \text{ peutClore}(H, \mathcal{G}, \Sigma, E) & : - \text{ estAdmis}(H, \mathcal{G}, \Sigma, p'), \text{ dernier}(p', E). \\ \text{ peutInitier}(H, \mathcal{G}, \Sigma, E) & : - \text{ estAdmis}(H, \mathcal{G}, \Sigma, p'), \text{ premier}(p', E). \end{aligned}$$

Le prédicat *séparation – temporelle* prend en entrée :

- H_{π_i} la définition *canonique* d'une phase π_i où l'on rappelle que l'état initial, le chemin syntaxique élémentaire et l'état final résultant sont définis *a priori*,
- $H_{\pi_{i+1}}$, la définition *canonique* de la phase qui suit directement π_i ,
- \mathcal{G} , un GRBM,
- et Σ , l'ensemble des paramétrisations de \mathcal{G} qui satisfont des contraintes biologiques.

Le lecteur peut se référer à la section 6.4.3.1 page 123 pour l'implémentation `Prolog` du prédicat `estAdmis`.

Le prédicat `estRequis` construit la contrainte éponyme introduite en définition 32 page 134. L'implémentation est au stade du premier prototype. Pour l'instant, ce prédicat délègue la tâche à un programme shell qui :

1. identifie les termes $K_{v,\omega}$ de la contrainte et instancie les termes $\eta(v')$ et $\eta'(v')$ à l'aide de H_{π_i} , $H_{\pi_{i+1}}$ et \mathcal{G} ,
2. teste la valeur de vérité de la contrainte pour chaque occurrence σ de Σ .

D'après la formule 7.1, la séparation temporelle de π_i et π_{i+1} est vérifiée s'il existe au moins une paramétrisation $\sigma \in \Sigma$ qui admet que E est requis avant E' . L'instruction `findall` identifie tous les couples (E, E') solutions :

$$\text{findall}([E, E'], \text{séparation - temporelle}(H_{\pi_i}, H_{\pi_{i+1}}, \mathcal{G}, \Sigma, E, E'))$$

Prouver la satisfiabilité du prédicat revient à évaluer tous les chemins admis biologiquement pour π_i et π_{i+1} , desquels sont extraits E et E' . On note n (resp. m) le nombre d'évènements de π_i (resp. π_{i+1}) et la complexité en temps de calcul pour identifier les solutions du prédicat est :

$$\Theta\left(\frac{n!}{k_1! \dots k_n!}\right) + \Theta\left(\frac{m!}{k_1! \dots k_m!}\right)$$

où les k_i représentent le nombre d'occurrence d'un évènement. De nouveau, nous n'imposerons pas les subtilités d'implémentation. Le prototype est disponible en suivant le lien <https://gitlab.com/deborahboyenval/temporal-separation-prototype>.

7.4.2 Notre graphe de régulation et ses phases munies de checkpoint

La définition 31 suffit à nous questionner sur la pertinence d'un graphe de régulation : *est-il capable de représenter les phases du cycle cellulaire qui sont munies d'un checkpoint, i.e. $G1_{précoce}, G1_{tardif}, S, G2, M_{précoce}, M_{tardif}$?* Les variables et seuils de notre GRBM (figure 5.18 page 102) ont permis de formaliser les phases G1, S, G2 et M à partir d'une définition canonique. Pour les raisons évoquées au chapitre 3 page 54, seuls les checkpoints interphases seront considérés pour la preuve de concept, c'est à dire G1/S, S/G2 et G2/M. On donnera toutefois dans cette section des indications pour la modélisation du checkpoint intraphase M.

Le chapitre précédent conclut sur le flou des connaissances autour de la phase M, plus précisément sur l'emplacement de son état checkpoint. C'est à partir de cet état que la transition métaphase/anaphase peut avoir lieu, après l'inactivation du SAC. Cette transition sépare $M_{précoce}$ de M_{tardif} et est caractérisée dans notre étude par l'évènement $b-$ ($b = 2 \rightarrow b = 1$) appartenant

au chemin de H_M . Contrairement aux autres transitions de phases, la transition métaphase/anaphase (abrégée M/A) n'isole pas un ensemble unique d'évènements requis avant elle :

$$\{P_M\} \quad b+; ep+; a-; a-; \underbrace{b-; b-}_{M/A}; en+; ep- \quad \{P_{G1}\}$$

est le triplet canonique de la phase M. Le multi-ensemble des évènements de $M_{précocoe}$ suggéré par ce triplet est $\{b+ : 1, ep+ : 1, a- : 2\}$. Aucune information de la littérature ne permet de considérer ces quatre évènements requis avant la transition métaphase/anaphase (représentée par $b-$), ce que corroborent les solutions du prédicat $estAdmis_\sigma$ pour la phase M. La figure 7.4 montre que deux séquences sont admises avant la transition métaphase/anaphase : Aucun modèle ne peut être

$$p'_M \in \left\{ \begin{array}{l} b+; ep+; a-; a-; \underbrace{b-; b-}_{M/A}; ep-; en+ \\ b+; ep+; a-; a-; \underbrace{b-; b-}_{M/A}; en+; ep- \\ b+; ep+; a-; a-; \underbrace{b-; ep-}_{M/A}; b-; en+ \\ b+; ep+; a-; \underbrace{b-}_{M/A}; a-; b-; en+; ep- \\ b+; ep+; a-; \underbrace{b-}_{M/A}; a-; b-; ep-; en+ \\ b+; ep+; a-; \underbrace{b-}_{M/A}; a-; ep-; b-; en+ \\ b+; ep+; a-; \underbrace{b-; b-}_{M/A}; a-; ep-; en+ \\ b+; ep+; a-; \underbrace{b-}_{M/A}; b-; a-; en+; ep- \end{array} \right.$$

Figure 7.4 – Les différentes positions de la transition métaphase/anaphase exhibées par les chemins admis en phase M. Le chemin en rouge est le chemin canonique de M tel que défini en table 6.1.

réfuté sur la base du multi-ensemble d'évènements qu'il admet pour $M_{précocoe}$. Le degré de liberté supplémentaire admis en phase M, en ce qui concerne l'ordre de ses évènements, implique qu'il n'est pas possible d'isoler $M_{précocoe}$ et M_{tardif} alors même que la transition métaphase/anaphase est non ambiguë grâce à notre GRBM.

7.4.3 La preuve de concept

| π_i/π_{i+1} | E | E' | $ \sigma \models$ séparation – temporelle (π_i, π_{i+1}) | Temps de calcul * |
|-------------------|----------|-----------|--|--------------------------|
| G1/S | {sk+} | {a+} | 16/16 | 2.1 sec |
| S/G2 | {sk-} | {a+} | 8/16 | 3.3 sec |
| G2/M | {b+} | {b+} | 16/16 | 1h12min |

Table 7.1 – Les solutions du prédicat *séparation – temporelle* pour 3 checkpoints du cycle cellulaire.

Remarque 11. *L'évènement qui peut clore G2 est le même que celui qui peut initier M, mais uniquement d'un point de vue syntaxique. Le premier (resp. second) b+ en G2 (resp. en M) est associé à une monophosphorylation (resp. hyperphosphorylation) du complexe cyclineB/cdk1. L'hyperphosphorylation de ce complexe est associée à l'activation de la mitose. $E \equiv b+$ en G2 est nécessairement requis avant $E' \equiv b+$ en M puisque la sémantique unitaire de René Thomas implique que b passe d'abord par son état $\eta(b) = 1$ (en G2 donc) avant d'atteindre son état $\eta(b) = 2$ (en M donc).*

In stricto sensu, notre graphe de régulation ne satisfait pas le prédicat *séparation-temporelle* de la définition 31. Il satisfait en revanche tous les checkpoints qu'il peut exhiber, cf. la première colonne de la table 7.1. Une des clauses du prédicat vise à identifier le nombre de modèles qui satisfont le checkpoint testé parmi les 16 modèles qui satisfont les propriétés classiques du cycle cellulaire (chapitre 5).

Le checkpoint S/G2 est satisfait par tout modèle (\mathcal{G}, σ) tel que $\sigma(K_{a,\{no_CI,restrict_sk\}}) \leq 1$ puisque la contrainte $estRequis_\sigma(E, E')$ est caractérisée en l'occurrence par :

$$[(\sigma(K_{a,\{no_CI,restrict_sk\}}) - 1) \times (\sigma(K_{a,\{no_CI,end_Rep\}}) - 1)] \leq 0 \quad (7.2)$$

et puisque l'on déduit la contrainte $\sigma(K_{a,\{no_CI,end_Rep\}}) > 1$ du calcul de la plus faible précondition du cycle cellulaire canonique. On réfère le lecteur à la page 74 pour plus de détails. $K_{a,\{no_CI,restrict_sk\}}$ est un paramètre non identifié. Les connaissances accumulées au chapitre 5 permettent de réduire son domaine à l'intervalle $\llbracket 0, 1 \rrbracket$, d'où les 8 modèles sélectionnés sur les 16. Ces connaissances sont en totale cohérence avec le checkpoint S/G2, plus spécifiquement les phase S et G2 sont séparées temporellement. En suivant la même approche, les contraintes $estRequis_\sigma(E, E')$ pour les deux autres checkpoints de la table 7.1 sont respectivement :

$$[(\sigma(K_{a,\{no_CI\}}) - 0) \times (\sigma(K_{a,\{sk2,no_CI\}}) - 0)] \leq 0 \quad (7.3)$$

et

$$[(\sigma(K_{b,\{a2,no_CI\}}) - 1) \times (\sigma(K_{b,\{a2,no_Wce1,no_CI\}}) - 1)] \leq 0 \quad (7.4)$$

La plus faible précondition associée au cycle cellulaire canonique implique $\sigma(K_{a,\{sk2,no_CI\}}) = 1$ et $\sigma(K_{b,\{a2,no_Wce1,no_CI\}}) = 2$ et les 16 modèles vérifient bel et bien que $\sigma(K_{a,\{no_CI\}}) < 1$ et $\sigma(K_{b,\{a2,no_CI\}}) < 2$, donc les checkpoints G1/S et G2/M sont bien satisfiables.

Remarque 12. *D'après la remarque 11, il serait tentant de considérer comme vaine la vérification de la contrainte $estRequis_\sigma$ pour le checkpoint G2/M. En effet, la sémantique unitaire de René Thomas contraint par définition le premier b+ en G2 avant le second en M, même lorsque*

$K_{v',\omega} = K_{v',\omega'}$. En revanche, la contrainte est $Requis_{\sigma}(b+, b+)$ met en évidence que c'est précisément l'ajout de la ressource no_Wee1 à l'ensemble des ressources de la variable b , dans l'état qui précède l'état checkpoint $\eta_{G2/M}$, qui permet à b de passer d'un état monophosphorylé en $G2$ à un état hyperphosphorylé en M , et donc permet la transition $G2/M$.

L'irréversibilité des phases est testée à l'aide de TotemBioNet et de la formule $\varphi_{irréversibilité}$, cf. notation 4. TotemBioNet a permis de vérifier la satisfiabilité de cette formule à l'aide du model-checker Nu-SMV toujours à partir des 16 modèles restant. Les résultats sont reportés dans la table 7.2.

| Checkpoint | $\varphi_{irréversibilité}$ | $ \sigma \models \varphi_{irréversibilité}$ |
|------------|---|--|
| $G1/S$ | $P_{G1} \implies \mathbb{A}[(\psi_{HG1} \vee \neg(\psi_{HS} \vee \psi_{HG2} \vee \psi_{HM})) \cup P_S]$ | 16/16 |
| $S/G2$ | $P_S \implies \mathbb{A}[(\psi_{HS} \vee \neg(\psi_{HG1} \vee \psi_{HG2} \vee \psi_{HM})) \cup P_{G2}]$ | 16/16 |
| $G2/M$ | $P_{G2} \implies \mathbb{A}[(\psi_{HG2} \vee \neg(\psi_{HG1} \vee \psi_{HS} \vee \psi_{HM})) \cup P_M]$ | 16/16 |

Table 7.2 – Vérification de la satisfiabilité de la formule $\varphi_{irréversibilité}$ pour tous les checkpoints que notre graphe de régulation nous permet d'étudier.

Pour rappel :

$$\begin{aligned}
\psi_{HG1} &\equiv (sk \geq 0 \wedge sk \leq 2) \wedge (en \geq 0 \wedge en \leq 1) \wedge (a = 0 \wedge b = 0 \wedge ep = 0 \wedge gf = 1) \\
\psi_{HS} &\equiv (a \geq 0 \wedge a \leq 1) \wedge (sk \geq 0 \wedge sk \leq 2) \wedge (b = 0 \wedge ep = 0 \wedge en = 0 \wedge gf = 1) \\
\psi_{HG2} &\equiv (a \geq 1 \wedge a \leq 2) \wedge (b \geq 0 \wedge b \leq 1) \wedge (sk = 0 \wedge ep = 0 \wedge en = 0 \wedge gf = 1) \\
\psi_{HM} &\equiv (a \leq 2 \wedge a \geq 0) \wedge (b \leq 2 \wedge b \geq 0) \wedge (en \leq 1 \wedge b \geq 0) \wedge (ep \leq 2 \wedge b \geq 0) \\
&\quad \wedge (sk = 0 \wedge gf = 1)
\end{aligned}$$

La vérification de la satisfiabilité de la conjonction des formules $\varphi_{irréversibilité}$ pour ces trois checkpoints est très efficace avec TotemBioNet (82 millisecondes). Ainsi, nous venons de prouver que les phases $G1$, S et $G2$ associées à notre graphe de régulation sont munies d'un checkpoint au sens où elles sont séparées temporellement et où la transition vers la phase suivante est irréversible. Notre preuve de concept de la modélisation du phénomène de checkpoint du cycle cellulaire est donc validée au sens où nous avons identifié des exemples biologiques pertinents biologiquement (nos 8 modèles restants le sont pour la question posée) qui satisfont les définitions 31 et 33. Tous les résultats exposés sont disponibles à l'adresse suivante : <https://gitlab.com/deborahboyenval/temporal-separation-prototype/-/tree/master/ProofOfConceptCheckpoint>.

CHAPITRE 8

Conclusion et perspectives

« Un expert est un homme qui a fait toutes les erreurs qu'on pouvait faire sur un sujet précis. »

-
Niels Bohr

8.1 Bilan

Dans le cadre de la modélisation des systèmes biologiques complexes, la notion de checkpoint est un concept que l'on retrouve relativement souvent et tout particulièrement dans l'exemple du cycle cellulaire. Cette notion porte plusieurs propriétés clés : on pense en premier lieu à la nécessité de terminer un ensemble de tâches avant de débiter l'ensemble de tâches suivant ; mais on doit penser également à l'interdiction de réaliser une marche arrière parmi ces ensembles de tâches identifiés et ordonnés les uns par rapport aux autres. La notion de checkpoint inclut nécessairement un aspect de blocage tant que l'ensemble des tâches en cours n'est pas terminé, *quand bien même ces dernières prendraient du retard*. Il en résulte qu'un ordonnancement correct de ces tâches ne doit pas dépendre de leurs délais de réalisation.

Nous supposons donc qu'il devait être possible de formaliser un checkpoint dans un cadre mathématique où le temps n'est pas quantitatif. C'est pourquoi nous avons opté dans cette thèse pour un formalisme purement discret, celui de René Thomas, où seule la chronologie des événements importe. Notre contribution principale a été de démontrer que la notion de checkpoint pouvait effectivement être formalisée dans un tel cadre. Ce cadre nous a offert la possibilité d'utiliser des outils issus de l'informatique théorique, tels que la vérification formelle de modèles qui prennent la forme d'automates à états finis et la logique de Hoare pour l'analyse statique de traces biologiques.

La tâche aurait pu être aisée si les systèmes biologiques considérés n'étaient pas cycliques, principalement parce que la question se serait résumée à une succession de vérifications de propriétés de non atteignabilité, ce que permettent par exemple les algorithmes de model-checking de formules LTL ou CTL [22, 51, 74] ou encore des algorithmes optimisés de parcours d'automates basés sur des techniques d'interprétation abstraite [59, 66]. Malheureusement, la plupart des systèmes physiologiques sont cycliques ce qui implique qu'à partir de tout état atteignable il est possible d'atteindre tous les autres états atteignables. La notion de checkpoint se devait donc d'être plus subtile que de « simples requêtes » d'atteignabilité.

La thèse montre finalement que deux notions sont fondamentales concernant les checkpoints, à savoir la séparation temporelle des ensembles d'évènements successifs, formalisée par le prédicat

séparation-temporelle au sein du chapitre 7 page 132, et l'unidirectionalité de la dynamique le long de ces ensembles d'évènements, formalisée par le prédicat *irréversibilité* toujours au chapitre 7 page 135.

8.1.1 Contributions méthodologiques

Concrètement, la *séparation temporelle* de deux ensembles d'évènements successifs est assurée si tous les évènements qui peuvent clore le premier ensemble d'évènements sont nécessaires à chacun des évènements qui peuvent initier le second. La logique de Hoare génétiquement modifiée nous a donné l'opportunité de formaliser de manière explicite la notion d'évènement et donc de faciliter la définition de ce prédicat. Les propriétés de séparation temporelle et d'irréversibilité se complètent pour définir la propriété dynamique fondamentale des checkpoints biologiques.

Le *prédicat de séparation temporelle* assure que l'évènement qui peut initier une phase ne peut avoir lieu au sein du domaine caractéristique de la phase précédente. Le prédicat d'irréversibilité des transitions de phases va plus loin en cloisonnant une phase en cours d'exécution vis à vis de toutes les autres phases et force ainsi la succession des phases à avoir lieu uniquement dans le sens défini, par exemple $G1 \rightarrow S \rightarrow G2 \rightarrow M$ dans le cas du cycle cellulaire.

Ce cloisonnement des phases nécessitait de prendre en compte une certaine subtilité, déjà appréhendée par [87]. Notre prédicat d'irréversibilité autorise qu'une trace puisse sortir de l'ensemble des états marqueurs d'une phase (*i.e* un hyperpavé d'une trace canonique) pourvu qu'elle y revienne toujours, sans jamais qu'elle ne visite un état relevant d'une autre phase. Autrement dit, la transition d'une phase vers les suivantes ne se fait qu'à travers l'état final de la phase courante qui marque également l'état initial de la phase qui lui succède directement.

Là encore, le caractère symbolique et formel de la théorie de Thomas a permis de définir l'irréversibilité sous forme de prédicat, alors qu'une première approche plus brutale aurait été d'examiner chacune des transitions qui sortent d'une phase canonique, au moyen de méthodes sémantiques.

8.1.2 Contributions pour la modélisation du cycle cellulaire

Notre preuve de concept a reposé sur l'exemple paradigmatique d'un comportement cyclique avec checkpoints : le cycle cellulaire. Dans ce cas précis, les checkpoints sont garants de l'intégrité de l'ADN lors de la réplication et de la mitose. La thèse prouve que notre concept de checkpoint est directement applicable au cycle cellulaire, sachant qu'une phase peut être décrite par une succession d'évènements qui ont lieu à partir d'un état initial.

Nous avons tiré parti du riche état de l'art de la modélisation mathématique du cycle cellulaire pour modéliser une mécanistique abstraite des checkpoints au moyen d'un graphe de régulation biologique. Le sens biologique précis des variables utiles pour traiter de la question de la succession des phases et des checkpoints a été entièrement réévalué. Essentiellement, les entités biologiques qu'abstraient les variables ont été mises en correspondance avec leur rôle effectif dans le cycle cellulaire. Le graphe d'interaction abstrait de Tyson et Novák [40], mis à jour par [69] et [82], a été complété par des multiplexes qui reflètent explicitement des contraintes dynamiques qu'imposent ces entités biologiques « collectivement » sur le système. Ces contraintes garantissent l'absence de certaines traces et sont cruciales pour les checkpoints. Le succès de la mise en oeuvre de la définition des checkpoints dans le cas du cycle cellulaire n'aurait pas été possible sans notre graphe de régulation enrichi en multiplexe mis à jour sur la base des connaissances biologiques

récentes.

Nous avons ensuite déduit un ensemble fini de graphes de transitions asynchrones à l'aide de la famille \mathcal{K} des paramètres dynamiques. Cette méthode offre l'avantage de s'affranchir d'un choix arbitraire de modèle. En effet, les données expérimentales permettent rarement d'inférer une unique logique d'activation des variables d'un graphe d'interaction, choix susceptible d'introduire des biais dans les prédictions d'un modèle [97]. Des connaissances globales sur la dynamique du système ont également été exploitées pour réduire l'explosion combinatoire.

La définition d'une phase ne pouvait être validée biologiquement que si une séquence canonique des événements était reflétée par au moins un graphe de transitions. Cette validation a été réalisée avec succès par application de la logique de Hoare génétiquement modifiée sur un ensemble de modèles multivalués du cycle cellulaire des mammifères. À partir de ce dernier, le domaine d'une phase est défini par les états que cette phase peut couvrir en réalisant, non pas sa séquence canonique, mais toutes les permutations de la séquence canonique qu'un ensemble de paramétrisations d'un graphe de régulation autorise.

8.2 Perspectives

8.2.1 Inférence de modèles à l'aide de nos prédicats

Notre spécification d'un checkpoint est définie par la conjonction de nos prédicats présentés au chapitre 7 et la mise à jour d'un modèle intervient lorsqu'au moins l'un de ces prédicats est faux. La question de la séparation temporelle est traitée symboliquement, au moyen de la résolution d'une contrainte sur la valeur des paramètres dynamiques d'un graphe de régulation, et informe le modélisateur sur la façon de contraindre localement l'action de certains régulateurs sur certaines variables. La question de l'unidirectionnalité du cycle cellulaire est quant à elle résolue par une énumération des valeurs correctes des paramètres dynamiques à l'aide de `TotemBioNet`, ce qui peut rendre la mise à jour d'un modèle fastidieuse lorsque le prédicat d'irréversibilité n'est pas satisfait. Une formule caractéristique des paramétrisations correctes serait plus facile à manipuler qu'une longue énumération, à l'image de la plus faible précondition de la logique de Hoare génétiquement modifiée.

Une méthode qui consisterait à exhiber toutes les transitions qui sortent de l'hyperpavé d'une phase et à *toutes* les interdire en modifiant le modèle de manière *ad hoc* serait excessivement brute de force. La méthode proposée avec notre prédicat *irréversibilité* est bien plus judicieuse au sens où elle autorise à sortir de la phase courante pourvu que l'on ne rencontre aucune autre phase. Dans certains cas pratiques de débogage de modèle, il peut suffire cependant d'identifier les transitions sortantes qui sont la cause de la non satisfaction du prédicat et de modifier le modèle pour qu'elles ne soit plus admises. On pourrait exploiter la capacité de certains model-checkers à renvoyer des chemins qui constituent un contre-exemple lorsque la propriété testée n'est pas satisfaite. Il serait également possible de développer un algorithme de model-checking qui traite de manière symbolique la question de la satisfiabilité du prédicat irréversibilité. Un tel model-checker renverrait une contrainte sur les paramètres dynamiques qui assurerait que le prédicat est satisfait par tous les états du système de transitions considéré.

8.2.2 À propos de notre modélisation des checkpoints du cycle cellulaire des mammifères

Notre modélisation a mis en évidence cinq points cruciaux au cours du cycle cellulaire. Notre preuve de concept s'est focalisée sur le point de restriction et les checkpoints qui se situent à l'entrée et la sortie de la phase S (G1/S et S/G2), celui situé à l'entrée de la mitose (G2/M), le tout dans un unique contexte environnemental, c'est à dire celui qui admet un attracteur cyclique qui inclut notre définition du cycle cellulaire. L'activité de modélisation nécessite d'être poursuivie afin de caractériser les régulations manquantes qui permettraient de modéliser le checkpoint de sortie de mitose (transition métaphase/anaphase) et donc de compléter notre preuve de concept à tous les checkpoints intrinsèques au cycle cellulaire. Dans une perspective de réduction de modèles ou d'explicabilité, l'algorithme REVEAL [12] ou la méthode d'apprentissage à partir d'un système de transitions (LFIT) [83] pourraient en théorie inférer l'ensemble des règles de régulations suffisantes et minimales pour représenter une phase spécifique munie d'un checkpoint.

8.2.3 Quelques applications en biologie

Le domaine de la biologie synthétique pourrait se saisir de notre démarche de spécification et vérification automatique d'un checkpoint en vue de munir un système oscillatoire d'un ou plusieurs checkpoints, par exemple [77]. Un tour de force expérimental a permis par exemple la conception d'un modèle minimal de la progression du cycle cellulaire chez la levure, où certaines régulations canoniques sont supprimées mais où l'ordonnancement des phases est toujours respecté [45]. Une telle démarche chez des organismes eucaryotes plus complexes mériterait d'être assistée par notre approche.

La capacité des checkpoints à arrêter le cycle cellulaire en réponse à des lésions de l'ADN motive de nombreuses recherches dans le secteur biomédical [103]. Certaines stratégies thérapeutiques anticancéreuses exploitent les propriétés cytostatiques de molécules pharmaceutiques pour bloquer la progression du cycle cellulaire (*e.g.* les inhibiteurs de cdk4 et cdk6) [50]. En empêchant l'activation de certaines cdks, le mécanisme d'action de ces molécules évoque un blocage d'un événement associé à une phase et donc le blocage du cycle cellulaire, si l'on se réfère à notre prédicat *checkpoint*. D'autres thérapies du cancer misent sur des molécules cytotoxiques qui induisent des dommages à l'ADN et activent les checkpoints intrinsèques *via* les checkpoints extrinsèques. Par exemple, des lésions de l'ADN activent p53 qui active à son tour un inhibiteur des complexes *cyc/cdk*, ce qui peut bloquer un événement d'une phase et donc bloquer la progression du cycle cellulaire. Tout modèle « multi-système » qui inclut une description du cycle cellulaire et qui satisfait notre prédicat *checkpoint* est un candidat sérieux pour tester des hypothèses variées.

Nous citerons enfin d'autres pistes de modélisation telles que l'application de nos prédicats à d'autres systèmes biologiques que celui du cycle cellulaire [67, 92, 95], le test de molécules cytostatiques ou cytotoxiques sur la dynamique du cycle cellulaire, l'étude de la synergie entre des mutants qui pourrait conduire à des phénotypes de cancer (*e.g.* une prolifération incontrôlée ou une accumulation de mutations) où encore l'exploration de la coordination entre les phases du cycle cellulaire, du cycle circadien [52, 75] et du métabolisme [31, 91]. Cela pourrait nécessiter l'ajout d'informations chronométriques sous la forme de délais de transition [71, 69] ou la définition de classes de priorité au sein d'un modèle discret [28, 66].

Bibliographie

- [1] Karl R. POPPER. *Conjectures et Réfutations*. Payot, 1963.
- [2] Tony HOARE. “An axiomatic basis for computer programming”. In : *Commun. ACM* 12(10) (1969), p. 576-580.
- [3] S.A. KAUFFMAN. “Metabolic stability and epigenesis in randomly constructed genetic nets”. In : *Journal of Theoretical Biology* 22.3 (1969), p. 437-467. ISSN : 0022-5193. DOI : [https://doi.org/10.1016/0022-5193\(69\)90015-0](https://doi.org/10.1016/0022-5193(69)90015-0). URL : <https://www.sciencedirect.com/science/article/pii/0022519369900150>.
- [4] René THOMAS. “Boolean formalization of genetic control circuits”. In : *Journal of Theoretical Biology* 42.3 (1973), p. 563-585. ISSN : 0022-5193. DOI : [https://doi.org/10.1016/0022-5193\(73\)90247-6](https://doi.org/10.1016/0022-5193(73)90247-6). URL : <http://www.sciencedirect.com/science/article/pii/0022519373902476>.
- [5] Edsger W. DIJKSTRA. “Guarded Commands, Nondeterminacy and Formal Derivation of Programs”. In : *Commun. ACM* 18.8 (août 1975), p. 453-457. ISSN : 0001-0782. DOI : [10.1145/360933.360975](https://doi.org/10.1145/360933.360975). URL : <https://doi.org/10.1145/360933.360975>.
- [6] W.V. QUINE. *Les deux dogmes de l'empirisme*. Gallimard, 1980.
- [7] Leland HARTWELL et Ted WEINERT. “Checkpoints : Controls That Ensure the Order of Cell Cycle Events”. In : *Science (New York, N.Y.)* 246 (déc. 1989), p. 629-34. DOI : [10.1126/science.2683079](https://doi.org/10.1126/science.2683079).
- [8] E.H SNOUSSI. “Qualitative dynamics of a piecewise-linear differential equations : a discrete mapping approach”. In : *Dynamics and stability of Systems* 4 (1989), p. 189-207.
- [9] René THOMAS. “Regulatory networks seen as asynchronous automata : A logical description”. In : *Journal of Theoretical Biology* 153.1 (1991), p. 1-23. ISSN : 0022-5193. DOI : [https://doi.org/10.1016/S0022-5193\(05\)80350-9](https://doi.org/10.1016/S0022-5193(05)80350-9). URL : <http://www.sciencedirect.com/science/article/pii/S0022519305803509>.
- [10] Denis THIEFFRY et René THOMAS. “Dynamical behaviour of biological regulatory networks—II. Immunity control in bacteriophage lambda”. In : *Bulletin of Mathematical Biology* 57.2 (1995), p. 277-297. ISSN : 0092-8240. DOI : [https://doi.org/10.1016/0092-8240\(94\)00037-D](https://doi.org/10.1016/0092-8240(94)00037-D). URL : <https://www.sciencedirect.com/science/article/pii/009282409400037D>.
- [11] Matthew O'CONNELL et al. “O'Connell M, Raleigh J, Verkade H, Nurse P. Chk1 is a wee1 kinase in the G2 DNA damage checkpoint inhibiting cdc2 by Y15 phosphorylation. EMBO J 16 : 545-554”. In : *The EMBO journal* 16 (mar. 1997), p. 545-54. DOI : [10.1093/emboj/16.3.545](https://doi.org/10.1093/emboj/16.3.545).
- [12] Shoudan LIANG, S. FUHRMAN et R. SOMOGYI. “Reveal, a general reverse engineering algorithm for inference of genetic network architectures.” In : *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* (1998), p. 18-29.

- [13] L MENDOZA, D THIEFFRY et E R ALVAREZ-BUYLLA. “Genetic control of flower morphogenesis in *Arabidopsis thaliana* : a logical analysis.” In : *Bioinformatics* 15.7 (juil. 1999), p. 593-606. ISSN : 1367-4803. DOI : [10.1093/bioinformatics/15.7.593](https://doi.org/10.1093/bioinformatics/15.7.593). eprint : <https://academic.oup.com/bioinformatics/article-pdf/15/7/593/9732060/150593.pdf>. URL : <https://doi.org/10.1093/bioinformatics/15.7.593>.
- [14] Paul NURSE. “A Long Twentieth Century of the Cell Cycle and Beyond”. In : *Cell* 100 (fév. 2000), p. 71-8. DOI : [10.1016/S0092-8674\(00\)81684-0](https://doi.org/10.1016/S0092-8674(00)81684-0).
- [15] Mikhail BLAGOSKLONNY et Arthur PARDEE. “The Restriction Point of the Cell Cycle”. In : *Cell cycle (Georgetown, Tex.)* 1 (mar. 2002), p. 103-10. DOI : [10.4161/cc.1.2.108](https://doi.org/10.4161/cc.1.2.108).
- [16] A. CIMATTI et al. “NuSMV 2 : An OpenSource Tool for Symbolic Model Checking”. In : *CAV*. 2002.
- [17] Jerry HSU et al. “E2F-dependent accumulation of hEmi1 regulates S phase entry by inhibiting APC(Cdh1)”. In : *Nature cell biology* 4 (juin 2002), p. 358-66. DOI : [10.1038/ncb785](https://doi.org/10.1038/ncb785).
- [18] Hidde de JONG. “Modeling and Simulation of Genetic Regulatory Systems : A Literature Review”. In : *J. Comput. Biol.* 9 (2002), p. 67-103.
- [19] Markus WELCKER et al. “Multisite Phosphorylation by Cdk2 and GSK3 Controls Cyclin E Degradation”. In : *Molecular cell* 12 (sept. 2003), p. 381-92. DOI : [10.1016/S1097-2765\(03\)00287-9](https://doi.org/10.1016/S1097-2765(03)00287-9).
- [20] Gilles BERNOT et al. “Application of formal methods to biological regulatory networks : Extending Thomas’ asynchronous logical approach with temporal logic”. In : *Journal of Theoretical Biology* 229 (sept. 2004), p. 339-47. DOI : [10.1016/j.jtbi.2004.04.003](https://doi.org/10.1016/j.jtbi.2004.04.003).
- [21] Adrian P. BRACKEN et al. “E2F target genes : unraveling the biology”. In : *Trends in Biochemical Sciences* 29.8 (2004), p. 409-417. ISSN : 0968-0004. DOI : <https://doi.org/10.1016/j.tibs.2004.06.006>. URL : <https://www.sciencedirect.com/science/article/pii/S0968000404001513>.
- [22] Nathalie CHABRIER-RIVIER et al. “Modeling and querying biomolecular interaction networks”. In : *Theoretical Computer Science* 325.1 (2004). Computational Systems Biology, p. 25-44. ISSN : 0304-3975. DOI : <https://doi.org/10.1016/j.tcs.2004.03.063>. URL : <https://www.sciencedirect.com/science/article/pii/S030439750400218X>.
- [23] Béla NOVÁK et John J. TYSON. “A model for restriction point control of the mammalian cell cycle”. In : *Journal of Theoretical Biology* 230.4 (2004). Special Issue in honour of Arthur T. Winfree, p. 563-579. ISSN : 0022-5193. DOI : <https://doi.org/10.1016/j.jtbi.2004.04.039>. URL : <https://www.sciencedirect.com/science/article/pii/S0022519304002449>.
- [24] Robert COSTA. “FoxM1 dances with mitosis”. In : *Nature cell biology* 7 (mar. 2005), p. 108-10. DOI : [10.1038/ncb0205-108](https://doi.org/10.1038/ncb0205-108).

- [25] Yuichi MACHIDA, Joyce HAMLIN et Anindya DUTTA. “Right Place, Right Time, and Only Once : Replication Initiation in Metazoans”. In : *Cell* 123 (nov. 2005), p. 13-24. DOI : [10.1016/j.cell.2005.09.019](https://doi.org/10.1016/j.cell.2005.09.019).
- [26] Yuichi MACHIDA, Jamie TEER et Anindya DUTTA. “Acute Reduction of an Origin Recognition Complex (ORC) Subunit in Human Cells Reveals a Requirement of ORC for Cdk2 Activation”. In : *The Journal of biological chemistry* 280 (août 2005), p. 27624-30. DOI : [10.1074/jbc.M502615200](https://doi.org/10.1074/jbc.M502615200).
- [27] Klaus BECKER et al. “Self-renewal of human embryonic stem cells is supported by a shortened G1 cell cycle phase”. In : *Journal of cellular physiology* 209 (déc. 2006), p. 883-93. DOI : [10.1002/jcp.20776](https://doi.org/10.1002/jcp.20776).
- [28] Adrien FAURÉ et al. “Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle”. In : *Bioinformatics* 22 14 (2006), e124-31.
- [29] Daniele GUARDAVACCARO et Michele PAGANO. “Stabilizers and Destabilizers Controlling Cell Cycle Oscillators”. In : *Molecular cell* 22 (mai 2006), p. 1-4. DOI : [10.1016/j.molcel.2006.03.017](https://doi.org/10.1016/j.molcel.2006.03.017).
- [30] Jonathon PINES. “Mitosis : A matter of getting rid of the right protein at the right time”. In : *Trends in cell biology* 16 (fév. 2006), p. 55-63. DOI : [10.1016/j.tcb.2005.11.006](https://doi.org/10.1016/j.tcb.2005.11.006).
- [31] Zheng CHEN et Steven MCKNIGHT. “A Conserved DNA Damage Response Pathway Responsible for Coupling the Cell Division Cycle to the Circadian and Metabolic Cycles”. In : *Cell cycle (Georgetown, Tex.)* 6 (oct. 2007), p. 2906-12. DOI : [10.4161/cc.6.23.5041](https://doi.org/10.4161/cc.6.23.5041).
- [32] Jane L. LUBISCHER. “The Cell Cycle, Principles of Control. David O. Morgan.” In : *Integrative and Comparative Biology* 47.5 (juin 2007), p. 794-795. ISSN : 1540-7063. DOI : [10.1093/icb/icm066](https://doi.org/10.1093/icb/icm066). eprint : <https://academic.oup.com/icb/article-pdf/47/5/794/1665242/icm066.pdf>. URL : <https://doi.org/10.1093/icb/icm066>.
- [33] A. RICHARD et J.-P. COMET. “Necessary conditions for multistationarity in discrete dynamical systems”. In : *Discrete Applied Mathematics* 155.18 (2007), p. 2403-2413.
- [34] Matt SULLIVAN et David MORGAN. “Finishing mitosis, one step at a time”. In : *Nature reviews. Molecular cell biology* 8 (déc. 2007), p. 894-903. DOI : [10.1038/nrm2276](https://doi.org/10.1038/nrm2276).
- [35] G. BERNOT, J.-P. COMET et Z. KHALIS. “Gene regulatory networks with multiplexes”. In : *European Simulation and Modelling Conference Proceedings*. ISBN 978-90-77381-44-1. France, oct. 2008, p. 423-432.
- [36] Arnaud BESSON, Steve DOWDY et James ROBERTS. “CDK Inhibitors : Cell Cycle regulators and Beyond”. In : *Developmental cell* 14 (mar. 2008), p. 159-69. DOI : [10.1016/j.devcel.2008.01.013](https://doi.org/10.1016/j.devcel.2008.01.013).
- [37] Laurence CALZONE et al. “A comprehensive modular map of molecular interactions in RB/E2F pathway”. In : *Molecular systems biology* 4 (fév. 2008), p. 173. DOI : [10.1038/msb.2008.7](https://doi.org/10.1038/msb.2008.7).
- [38] Helfrid HOCHEGGER, Shunichi TAKEDA et Tim HUNT. “Cyclin-dependent kinases and cell-cycle transitions : Does one fit all ?” In : *Nature reviews. Molecular cell biology* 9 (oct. 2008), p. 910-6. DOI : [10.1038/nrm2510](https://doi.org/10.1038/nrm2510).

- [39] Adrien RICHARD. *Fair paths in CTL*. Technical Report. 2008. URL : <https://gitlab.com/totembionet/totembionet>.
- [40] John J TYSON et Bela NOVAK. “Temporal organization of the cell cycle.” In : *CB Current biology* 18 (2008), R759-R768. DOI : [10.1016/j.cub.2008.07.001](https://doi.org/10.1016/j.cub.2008.07.001).
- [41] Leroy W. WHEELER, Nathan H. LENTS et Joseph J. BALDASSARE. “Cyclin A-CDK activity during G1 phase impairs MCM chromatin loading and inhibits DNA synthesis in mammalian cells”. In : *Cell Cycle* 7.14 (2008). PMID : 18635963, p. 2179-2188. DOI : [10.4161/cc.7.14.6270](https://doi.org/10.4161/cc.7.14.6270).
- [42] C. GÉRARD et A. GOLDBETER. “Temporal self-organization of the cyclin/Cdk network driving the mammalian cell cycle”. In : *Proceedings of the National Academy of Sciences* 106 (2009), p. 21643-21648.
- [43] Arne LINDQVIST, Veronica RODRIGUEZ-BRAVO et René MEDEMA. “The decision to enter mitosis : Feedback and redundancy in the mitotic entry network”. In : *The Journal of cell biology* 185 (mai 2009), p. 193-202. DOI : [10.1083/jcb.200812045](https://doi.org/10.1083/jcb.200812045).
- [44] Marcos MALUMBRES et Mariano BARBACID. “Malumbres M, Barbacid MCell cycle, CDKs and cancer : a changing paradigm. Nat Rev Cancer 9 : 153-166”. In : *Nature reviews. Cancer* 9 (avr. 2009), p. 153-66. DOI : [10.1038/nrc2602](https://doi.org/10.1038/nrc2602).
- [45] Damien COUDREUSE et Paul NURSE. “Driving the cell cycle with a minimal CDK control network”. In : *Nature* 468 (déc. 2010), p. 1074-9. DOI : [10.1038/nature09543](https://doi.org/10.1038/nature09543).
- [46] Barbara DI FIORE et Jonathon PINES. “How cyclin A destruction escapes the spindle assembly checkpoint”. In : *The Journal of cell biology* 190 (août 2010), p. 501-9. DOI : [10.1083/jcb.201001083](https://doi.org/10.1083/jcb.201001083).
- [47] Olivier GAVET et Jonathon PINES. “Progressive Activation of CyclinB1-Cdk1 Coordinates Entry to Mitosis”. In : *Developmental Cell* 18.4 (2010), p. 533-543. ISSN : 1534-5807. DOI : <https://doi.org/10.1016/j.devcel.2010.02.013>. URL : <https://www.sciencedirect.com/science/article/pii/S1534580710001115>.
- [48] Eusebio MANCHADO, Manuel EGUREN et Marcos MALUMBRES. “The anaphase-promoting complex/cyclosome (APC/C) : cell-cycle-dependent and -independent functions”. In : *Biochemical Society Transactions* 38.1 (jan. 2010), p. 65-71. ISSN : 0300-5127. DOI : [10.1042/BST0380065](https://doi.org/10.1042/BST0380065). eprint : <https://portlandpress.com/biochemsoctrans/article-pdf/38/1/65/547226/bst0380065.pdf>. URL : <https://doi.org/10.1042/BST0380065>.
- [49] Adrien RICHARD. “Negative circuits and sustained oscillations in asynchronous automata networks”. In : *Advances in Applied Mathematics* 44.4 (2010), p. 378-392. ISSN : 0196-8858. DOI : <https://doi.org/10.1016/j.aam.2009.11.011>. URL : <https://www.sciencedirect.com/science/article/pii/S0196885809001304>.
- [50] Asako SAKAUE-SAWANO et al. “Drug-induced cell cycle modulation leading to cell-cycle arrest, nuclear mis-segregation, or endoreplication”. In : *BMC Cell Biology* 12 (2010), p. 2-2.

- [51] Gustavo ARELLANO et al. "'Antelope' : A hybrid-logic model checker for branching-time Boolean GRN analysis". In : *BMC bioinformatics* 12 (déc. 2011), p. 490. DOI : [10.1186/1471-2105-12-490](https://doi.org/10.1186/1471-2105-12-490).
- [52] Elisabetta MARIA et al. "Design, optimization and predictions of a coupled model of the cell cycle, circadian clock, DNA repair system, irinotecan metabolism and exposure control under temporal logic constraints". In : *Theoretical Computer Science* 412 (mai 2011), p. 2108-2127. DOI : [10.1016/j.tcs.2010.10.036](https://doi.org/10.1016/j.tcs.2010.10.036).
- [53] Andrew MURRAY. "A brief history of error". In : *Nature cell biology* 13 (oct. 2011), p. 1178-82. DOI : [10.1038/ncb2348](https://doi.org/10.1038/ncb2348).
- [54] Asako SAKAUE-SAWANO, Kenji OHTAWA et Atsushi MIYAWAKI. "Drug-induced cell cycle modulation leading to cell-cycle arrest, nuclear mis-segregation, or endoreplication". In : *BMC cell biology* 12 (jan. 2011), p. 2. DOI : [10.1186/1471-2121-12-2](https://doi.org/10.1186/1471-2121-12-2).
- [55] Marieke AARTS et al. "Forced Mitotic Entry of S-Phase Cells as a Therapeutic Strategy Induced by Inhibition of WEE1". In : *Cancer discovery* 2 (mai 2012), p. 524-39. DOI : [10.1158/2159-8290.CD-11-0320](https://doi.org/10.1158/2159-8290.CD-11-0320).
- [56] Melvin DEPAMPHILIS et al. "'The Octet' : Eight Protein Kinases that Control Mammalian DNA Replication". In : *Frontiers in physiology* 3 (sept. 2012), p. 368. DOI : [10.3389/fphys.2012.00368](https://doi.org/10.3389/fphys.2012.00368).
- [57] Brian GABRIELLI, Kelly BROOKS et Sandra PAVEY. "Defective Cell Cycle Checkpoints as Targets for Anti-Cancer Therapies". In : *Frontiers in pharmacology* 3 (fév. 2012), p. 9. DOI : [10.3389/fphar.2012.00009](https://doi.org/10.3389/fphar.2012.00009).
- [58] Cosetta BERTOLI, Jan SKOTHEIM et Robertus BRUIN. "Control of cell cycle transcription during G1 and S phases". In : *Nature reviews. Molecular cell biology* 14 (août 2013), p. 518-28. DOI : [10.1038/nrm3629](https://doi.org/10.1038/nrm3629).
- [59] Maxime FOLSCHETTE et al. "Under-approximation of Reachability in Multivalued Asynchronous Networks". In : *Electronic Notes in Theoretical Computer Science* 299 (déc. 2013), p. 33-51. DOI : [10.1016/j.entcs.2013.11.004](https://doi.org/10.1016/j.entcs.2013.11.004).
- [60] Carito GUZIOLOWSKI et al. "Exhaustively characterizing feasible logic models of a signaling network using Answer Set Programming". In : *Bioinformatics (Oxford, England)* 30 (juil. 2013). DOI : [10.1093/bioinformatics/btt393](https://doi.org/10.1093/bioinformatics/btt393).
- [61] Tae LEE, Guang YAO et Lingchong YOU. "Cell Cycle Transition, Principles of Restriction Point". In : *Encyclopedia of Systems Biology*. Sous la dir. de Werner DUBITZKY et al. New York, NY : Springer New York, 2013, p. 326-329. ISBN : 978-1-4419-9863-7. DOI : [10.1007/978-1-4419-9863-7_36](https://doi.org/10.1007/978-1-4419-9863-7_36). URL : https://doi.org/10.1007/978-1-4419-9863-7_36.
- [62] Sabrina SPENCER et al. "The Proliferation-Quiescence Decision Is Controlled by a Bifurcation in CDK2 Activity at Mitotic Exit". In : *Cell* 155 (sept. 2013). DOI : [10.1016/j.cell.2013.08.062](https://doi.org/10.1016/j.cell.2013.08.062).
- [63] José Carlos MOMBACH, Cristhian BUGS et Claudine CHAOUIYA. "Modelling the onset of senescence at the G1/S cell cycle checkpoint". In : *BMC Genomics* 15 (oct. 2014). DOI : [10.1186/1471-2164-15-S7-S7](https://doi.org/10.1186/1471-2164-15-S7-S7).

- [64] Gilles BERNOT et al. “A Genetically Modified Hoare Logic”. In : *Theoretical Computer Science* 765 (juin 2015). DOI : [10.1016/j.tcs.2018.02.003](https://doi.org/10.1016/j.tcs.2018.02.003).
- [65] Celine FEILLET et al. “Coupling between the Circadian Clock and Cell Cycle Oscillators : Implication for Healthy Cells and Malignant Growth”. In : *Frontiers in Neurology* 6 (mai 2015). DOI : [10.3389/fneur.2015.00096](https://doi.org/10.3389/fneur.2015.00096).
- [66] Maxime FOLSCHETTE et al. “Sufficient Conditions for Reachability in Automata Networks with Priorities”. In : *Theoretical Computer Science* (2015). DOI : [10.1016/j.tcs.2015.08.040](https://doi.org/10.1016/j.tcs.2015.08.040). URL : <https://hal.archives-ouvertes.fr/hal-01202671>.
- [67] Fritz MELCHERS. “Checkpoints that control B cell development”. In : *The Journal of clinical investigation* 125 (mai 2015), p. 1-8. DOI : [10.1172/JCI78083](https://doi.org/10.1172/JCI78083).
- [68] Santiago VIDELA et al. “Learning Boolean logic models of signaling networks with ASP”. In : *Theoretical Computer Science* 599 (2015). Advances in Computational Methods in Systems Biology, p. 79-101. ISSN : 0304-3975. DOI : <https://doi.org/10.1016/j.tcs.2014.06.022>. URL : <https://www.sciencedirect.com/science/article/pii/S0304397514004587>.
- [69] J. BEHAEGEL et al. “A hybrid model of cell cycle in mammals”. In : *J. Bioinformatics and Comput. Biol.* 14.1 (2016), 1640001 [17 pp.]
- [70] Steven CAPPELL et al. “Irreversible APCCdh1 Inactivation Underlies the Point of No Return for Cell-Cycle Entry”. In : *Cell* 166 (juin 2016), p. 167-180. DOI : [10.1016/j.cell.2016.05.077](https://doi.org/10.1016/j.cell.2016.05.077).
- [71] Emilien CORNILLON et al. “Hybrid Gene Networks : a new Framework and a Software Environment.” In : *Advances in Systems and Synthetic Biology, Modelling Complex Biological Systems in the Context of Genomics*. 2016. URL : <https://hal.archives-ouvertes.fr/hal-01345019>.
- [72] Dávid DERITEI et al. “Principles of dynamical modularity in biological regulatory networks”. In : *Scientific Reports* 6 (mar. 2016), p. 21957. DOI : [10.1038/srep21957](https://doi.org/10.1038/srep21957).
- [73] M. GEBSER et al. “Theory Solving made easy with Clingo 5 (Extended Version *)”. In : 2016.
- [74] Denis THIEFFRY et al. “Logical model specification aided by model- checking techniques : application to the mammalian cell cycle regulation”. In : *Bioinformatics* 32 (sept. 2016). DOI : [10.1093/bioinformatics/btw457](https://doi.org/10.1093/bioinformatics/btw457).
- [75] Pauline TRAYNARD et al. “Model-based investigation of the circadian clock and cell cycle coupling in mouse embryonic fibroblasts : Prediction of RevErb- α up-regulation during mitosis”. In : *Biosystems* 149 (2016). Selected papers from the Computational Methods in Systems Biology 2015 conference, p. 59-69. ISSN : 0303-2647. DOI : <https://doi.org/10.1016/j.biosystems.2016.07.003>. URL : <https://www.sciencedirect.com/science/article/pii/S0303264716301162>.
- [76] Emilien CORNILLON. “Modèles qualitatifs de réseaux génétiques : réduction de modèles et introduction d’un temps continu”. Theses. Université Côte d’Azur, oct. 2017. URL : <https://tel.archives-ouvertes.fr/tel-01682212>.

- [77] Lauren ANDREWS, Alec NIELSEN et Christopher VOIGT. “Cellular checkpoint control using programmable sequential logic”. In : *Science* 361 (sept. 2018), eaap8987. DOI : [10.1126/science.aap8987](https://doi.org/10.1126/science.aap8987).
- [78] Madalena CHAVES et Laurent TOURNIER. “Analysis Tools for Interconnected Boolean Networks With Biological Applications”. In : *Frontiers in Physiology* 9 (2018). ISSN : 1664-042X. DOI : [10.3389/fphys.2018.00586](https://doi.org/10.3389/fphys.2018.00586). URL : <https://www.frontiersin.org/article/10.3389/fphys.2018.00586>.
- [79] Edmund M. CLARKE et al. “Handbook of Model Checking”. In : *Springer International Publishing*. 2018.
- [80] Claude GÉRARD, Didier GONZE et Albert GOLDBETER. “Revisiting a skeleton model for the mammalian cell cycle : From bistability to Cdk oscillations and cellular heterogeneity”. In : *Journal of Theoretical Biology* 461 (oct. 2018). DOI : [10.1016/j.jtbi.2018.10.042](https://doi.org/10.1016/j.jtbi.2018.10.042).
- [81] Hannes KLARNER et al. *Basins of Attraction, Commitment Sets and Phenotypes of Boolean Networks*. Juil. 2018.
- [82] Bela NOVAK, Frank HELDT et John TYSON. “Genome Stability during Cell Proliferation : A Systems Analysis of the Molecular Mechanisms Controlling Progression through the Eukaryotic Cell Cycle”. In : *Current Opinion in Systems Biology* 9 (fév. 2018). DOI : [10.1016/j.coisb.2018.02.004](https://doi.org/10.1016/j.coisb.2018.02.004).
- [83] Tony RIBEIRO et al. “Inductive Learning from State Transitions over Continuous Domains”. In : *27th International Conference on Inductive Logic Programming*. T. LNCS, volume 10759. Inductive Logic Programming. Part of the Lecture Notes in Computer Science book series (LNCS, volume 10759) Also part of the Lecture Notes in Artificial Intelligence book sub series (LNAI, volume 10759). Springer, Cham, mar. 2018, p. 124-139. DOI : [10.1007/978-3-319-78090-0_9](https://doi.org/10.1007/978-3-319-78090-0_9). URL : <https://hal.archives-ouvertes.fr/hal-01655644>.
- [84] Joshua SALDIVAR et al. “An intrinsic S/G 2 checkpoint enforced by ATR”. In : *Science* 361 (août 2018), p. 806-810. DOI : [10.1126/science.aap9346](https://doi.org/10.1126/science.aap9346).
- [85] G. BERNOT et al. “A Genetically Modified Hoare Logic”. In : *Theoretical Computer Science* 765 (2019). <https://doi.org/10.1016/j.tcs.2018.02.003>, p. 145-157.
- [86] Dávid DERITEI et al. “A feedback loop of conditionally stable circuits drives the cell cycle from checkpoint to checkpoint”. In : *Scientific Reports* 9 (nov. 2019). DOI : [10.1038/s41598-019-52725-1](https://doi.org/10.1038/s41598-019-52725-1).
- [87] Ousmane DIOP, Laurent TOURNIEL et Vincent FROMION. “Summarizing complex asynchronous Boolean attractors, application to the analysis of a mammalian cell cycle model”. In : *2019 18th European Control Conference (ECC)*. 2019, p. 1677-1682. DOI : [10.23919/ECC.2019.8795712](https://doi.org/10.23919/ECC.2019.8795712).
- [88] Babatunde EKUNDAYO et Franziska BLEICHERT. “Origins of DNA replication”. In : *PLOS Genetics* 15 (sept. 2019), e1008320. DOI : [10.1371/journal.pgen.1008320](https://doi.org/10.1371/journal.pgen.1008320).

- [89] Maxime FOLSCHETTE. *The Hoare-fol Tool*. Technical Report. Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRISAL - Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France, déc. 2019. URL : <https://hal.archives-ouvertes.fr/hal-02409801>.
- [90] Bennie LEMMENS et Arne LINDQVIST. “DNA replication and mitotic entry : A brake model for cell cycle progression”. In : *The Journal of Cell Biology* 218 (déc. 2019), jcb.201909032. DOI : [10.1083/jcb.201909032](https://doi.org/10.1083/jcb.201909032).
- [91] Cecile MOULIN, Laurent TOURNIER et Sabine PÉRÈS. “Using a hybrid approach to model central carbon metabolism across the cell cycle”. In : *Hybrid Systems Biology*. T. 11705. Lecture Notes in Computer Science Chapter 9. Češka M., Paoletti N. (eds), 2019. DOI : [10.1007/978-3-030-28042-0_9](https://doi.org/10.1007/978-3-030-28042-0_9). URL : <https://hal.inrae.fr/hal-02791386>.
- [92] Suyasha ROY, Zaigham RIZVI et Amit AWASTHI. “Metabolic Checkpoints in Differentiation of Helper T Cells in Tissue Inflammation”. In : *Frontiers in Immunology* 9 (jan. 2019), p. 3036. DOI : [10.3389/fimmu.2018.03036](https://doi.org/10.3389/fimmu.2018.03036).
- [93] Wayne STALLAERT et al. “Bistable switches as integrators and actuators during cell cycle progression”. In : *FEBS Letters* 593.20 (2019), p. 2805-2816. DOI : <https://doi.org/10.1002/1873-3468.13628>. eprint : <https://febs.onlinelibrary.wiley.com/doi/pdf/10.1002/1873-3468.13628>. URL : <https://febs.onlinelibrary.wiley.com/doi/abs/10.1002/1873-3468.13628>.
- [94] Gabriele SULLI, Michael LAM et Satchidananda PANDA. “Interplay between Circadian Clock and Cancer : New Frontiers for Cancer Treatment”. In : *Trends in Cancer* 5 (août 2019). DOI : [10.1016/j.trecan.2019.07.002](https://doi.org/10.1016/j.trecan.2019.07.002).
- [95] Zhihao SUN et Jeffrey BRODSKY. “Protein quality control in the secretory pathway”. In : *The Journal of Cell Biology* 218 (oct. 2019), jcb.201906047. DOI : [10.1083/jcb.201906047](https://doi.org/10.1083/jcb.201906047).
- [96] Déborah BOYENVAL et al. “What is a cell cycle checkpoint ? The ToTemBioNet answer.” In : *18th International Conference on Computational Methods in Systems Biology (CMSB 2020)*. 2020.
- [97] Stéphanie CHEVALIER et al. “Synthesis and Simulation of Ensembles of Boolean Networks for Cell Fate Decision”. In : *18th International Conference on Computational Methods in Systems Biology (CMSB)*. Online, Germany, 2020. URL : <https://hal.archives-ouvertes.fr/hal-02898849>.
- [98] Camilla ELAEK, Valdemaras PETROSIUS et Claus SORENSEN. “WEE1 kinase limits CDK activities to safeguard DNA replication and mitotic entry”. In : *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 819-820 (fév. 2020), p. 111694. DOI : [10.1016/j.mrfmmm.2020.111694](https://doi.org/10.1016/j.mrfmmm.2020.111694).
- [99] Océane MARESCAL et Iain CHEESEMAN. “Cellular Mechanisms and Regulation of Quiescence”. In : *Developmental Cell* 55 (nov. 2020), p. 259-271. DOI : [10.1016/j.devcel.2020.09.029](https://doi.org/10.1016/j.devcel.2020.09.029).

- [100] Loïc PAULEVÉ et al. “Reconciling Qualitative, Abstract, and Scalable Modeling of Biological Networks”. In : *Nature Communications* 11 (2020). DOI : [10.1038/s41467-020-18112-5](https://doi.org/10.1038/s41467-020-18112-5). URL : <https://hal.archives-ouvertes.fr/hal-02518582>.
- [101] Seth M. RUBIN, Julien SAGE et Jan M. SKOTHEIM. “Integrating Old and New Paradigms of G1/S Control”. In : *Molecular Cell* 80.2 (2020), p. 183-192. ISSN : 1097-2765. DOI : <https://doi.org/10.1016/j.molcel.2020.08.020>. URL : <https://www.sciencedirect.com/science/article/pii/S1097276520306055>.
- [102] Xianlin ZOU et al. “A Systems Biology Approach Identifies Hidden Regulatory Connections Between the Circadian and Cell-Cycle Checkpoints”. In : *Frontiers in Physiology* 11 (avr. 2020). DOI : [10.3389/fphys.2020.00327](https://doi.org/10.3389/fphys.2020.00327).
- [103] Helen MATTHEWS, Cosetta BERTOLI et Robertus BRUIN. “Cell cycle control in cancer”. In : *Nature Reviews Molecular Cell Biology* 23 (sept. 2021). DOI : [10.1038/s41580-021-00404-3](https://doi.org/10.1038/s41580-021-00404-3).
- [104] SWIPROLOG. 2021. URL : <https://www.swi-prolog.org/pldoc/man?predicate=permutation/2>.
- [105] AMIGO2 CHECKPOINT. 2022. URL : <http://amigo.geneontology.org/amigo/search/bioentity?q=checkpoint>.
- [106] AMIGO2 REPLICATION. 2022. URL : <http://amigo.geneontology.org/amigo/search/annotation?q=replication>.
- [107] AMIGO2 SAC. 2022. URL : <http://amigo.geneontology.org/amigo/search/annotation?q=Spindle%5C%20Assembly%5C%20Checkpoint>.
- [108] REACTOME CHECKPOINT. 2022. URL : <https://reactome.org/PathwayBrowser/#/R-HSA-69620&DTAB=MT>.

Annexes

ANNEXES A

Paramétrisation de notre GRBM du cycle cellulaire

| $\phi_{K_{v,\omega}}$ | ϕ_{LHGM} | Conjonction des contraintes |
|--|---|---|
| Paramètres de sk interprétés dans le domaine $\llbracket 0, 2 \rrbracket$ | | |
| $K_{sk,\emptyset} = 0$ | - | $K_{sk,\emptyset} = 0$ |
| $K_{sk,\{restrick_gf\}} = 0$ | - | $K_{sk,\{restrick_gf\}} = 0$ |
| $K_{sk,\{rp\}} = 0$ | - | $K_{sk,\{rp\}} = 0$ |
| $K_{sk,\{no_en\}} = 0$ | - | $K_{sk,\{no_en\}} = 0$ |
| $K_{sk,\{no_a\}} = 0$ | - | $K_{sk,\{no_a\}} = 0$ |
| $K_{sk,\{restrick_gf,rp\}} = 0$ | - | $K_{sk,\{restrick_gf,rp\}} = 0$ |
| $K_{sk,\{restrick_gf,no_en\}} = 0$ | - | $K_{sk,\{restrick_gf,no_en\}} = 0$ |
| $K_{sk,\{restrick_gf,no_a\}} < 2$ | $K_{sk,\{restrick_gf,no_a\}} > 0$ | $K_{sk,\{restrick_gf,no_a\}} = 1$ |
| $K_{sk,\{rp,no_en\}} = 0$ | $K_{sk,\{rp,no_en\}} < 1$ | $K_{sk,\{rp,no_en\}} = 0$ |
| $K_{sk,\{rp,no_a\}} < 2$ | - | $K_{sk,\{rp,no_a\}} < 2$ |
| $K_{sk,\{no_en,no_a\}} = 0$ | - | $K_{sk,\{no_en,no_a\}} = 0$ |
| $K_{sk,\{restrick_gf,rp,no_en\}} = 0$ | $K_{sk,\{restrick_gf,rp,no_en\}} < 1$ | $K_{sk,\{restrick_gf,rp,no_en\}} = 0$ |
| $K_{sk,\{restrick_gf,rp,no_a\}} < 2$ | - | $K_{sk,\{restrick_gf,rp,no_a\}} = 1$ (CS) |
| $K_{sk,\{restrick_gf,no_en,no_a\}} < 2$ | - | $K_{sk,\{restrick_gf,no_en,no_a\}} = 1$ (CS) |
| - | $K_{sk,\{rp,no_en,no_a\}} > 1$ | $K_{sk,\{rp,no_en,no_a\}} = 2$ |
| - | $K_{sk,\{restrick_gf,rp,no_en,no_a\}} > 1$ | $K_{sk,\{restrick_gf,rp,no_en,no_a\}} = 2$ |
| Paramètres de a interprétés dans le domaine $\llbracket 0, 2 \rrbracket$ | | |
| $K_{a,\emptyset} = 0$ | - | $K_{a,\emptyset} = 0$ |
| $K_{a,\{sk2\}} = 0$ | - | $K_{a,\{sk2\}} = 0$ |
| $K_{a,\{no_CI\}} = 0$ | - | $K_{a,\{no_CI\}} = 0$ |
| $K_{a,\{restrick_sk\}} = 0$ | - | $K_{a,\{restrick_sk\}} = 0$ |
| $K_{a,\{end_rep\}} = 0$ | $K_{a,\{end_rep\}} < 1$ | $K_{a,\{end_rep\}} = 0$ |
| $K_{a,\{sk2,no_CI\}} < 2$ | $K_{a,\{sk2,no_CI\}} > 0$ | $K_{a,\{sk2,no_CI\}} = 1$ |
| $K_{a,\{sk2,restrick_sk\}} = 0$ | - | $K_{a,\{sk2,restrick_sk\}} = 0$ |
| $K_{a,\{no_CI,restrick_sk\}} < 2$ | - | $K_{a,\{no_CI,restrick_sk\}} < 2$ |
| - | $K_{a,\{end_rep,no_CI\}} > 1$ | $K_{a,\{end_rep,no_CI\}} = 2$ |

$$K_{a,\{sk2,no_CI,restrict_sk\}} < 2 \quad - \quad K_{a,\{sk2,no_CI,restrict_sk\}} = 1 \text{ (CS)}$$

Paramètres de b interprétés dans le domaine $\llbracket 0, 2 \rrbracket$

| | | |
|---------------------------------|------------------------------------|------------------------------------|
| $K_{b,\emptyset} = 0$ | - | $K_{b,\emptyset} = 0$ |
| $K_{b,\{a2\}} < 2$ | - | $K_{b,\{a2\}} < 2$ |
| $K_{b,\{no_Wee1\}} = 0$ | $K_{b,\{no_Wee1\}} < 1$ | $K_{b,\{no_Wee1\}} = 0$ |
| $K_{b,\{no_CI\}} = 0$ | - | $K_{b,\{no_CI\}} = 0$ |
| $K_{b,\{a2,SAC\}} = 1$ | - | $K_{b,\{a2,SAC\}} = 1$ |
| $K_{b,\{a2,no_Wee1\}} > 0$ | $K_{b,\{a2,no_Wee1\}} < 2$ | $K_{b,\{a2,no_Wee1\}} = 1$ |
| - | $K_{b,\{a2,no_CI\}} > 0$ | $K_{b,\{a2,no_CI\}} = 1$ |
| $K_{b,\{no_Wee1,no_CI\}} = 0$ | - | $K_{b,\{no_Wee1,no_CI\}} = 0$ |
| - | $K_{b,\{a2,no_Wee1,no_CI\}} > 1$ | $K_{b,\{a2,no_Wee1,no_CI\}} = 2$ |
| $K_{b,\{a2,no_Wee1,SAC\}} = 2$ | - | $K_{b,\{a2,no_Wee1,SAC\}} = 2$ |

Paramètres de en interprétés dans le domaine $\llbracket 0, 1 \rrbracket$

| | | |
|-------------------------------------|-------------------------------------|-------------------------------------|
| $K_{en,\emptyset} = 0$ | - | $K_{en,\emptyset} = 0$ |
| $K_{en,\{no_sk\}} = 0$ | - | $K_{en,\{no_sk\}} = 0$ |
| $K_{en,\{no_a\}} = 0$ | - | $K_{en,\{no_a\}} = 0$ |
| $K_{en,\{no_b\}} = 0$ | - | $K_{en,\{no_b\}} = 0$ |
| $K_{en,\{no_sk,no_a\}} = 0$ | - | $K_{en,\{no_sk,no_a\}} = 0$ |
| $K_{en,\{no_sk,no_b\}} = 0$ | - | $K_{en,\{no_sk,no_b\}} = 0$ |
| $K_{en,\{no_a,no_b\}} = 0$ | $K_{en,\{no_a,no_b\}} < 1$ | $K_{en,\{no_a,no_b\}} = 0$ |
| $K_{en,\{no_sk,no_a,no_b\}} = 1$ | $K_{en,\{no_sk,no_a,no_b\}} > 0$ | $K_{en,\{no_sk,no_a,no_b\}} = 1$ |

Paramètres de ep interprétés dans le domaine $\llbracket 0, 1 \rrbracket$

| | | |
|-------------------------|----------------------------|-------------------------------|
| $K_{ep,\emptyset} = 0$ | $K_{ep,\emptyset} < 1$ | $K_{ep,\emptyset} = 0$ |
| - | - | $K_{ep,\{b2\}}$ non contraint |
| $K_{ep,\{no_en\}} = 0$ | - | $K_{ep,\{no_en\}} = 0$ |
| - | $K_{ep,\{b2,no_en\}} > 0$ | $K_{ep,\{b2,no_en\}} = 1$ |

Table A.1 – Liste des paramètres de notre GRBM du cycle cellulaire. La conjonction de tous les paramètres de la dernière colonne caractérise 16 modèles cohérents biologiquement. On note ψ_{bio} cette contrainte sous forme conjonctive.

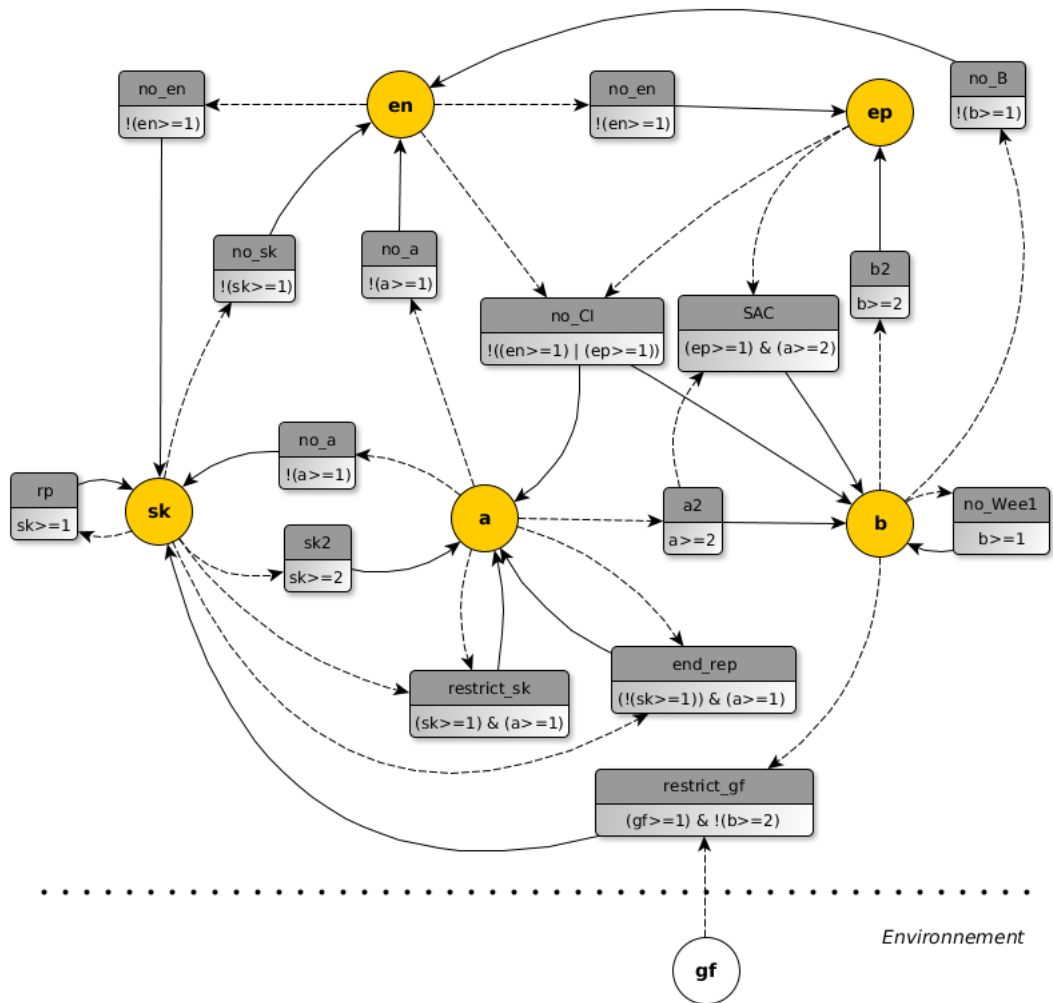


Figure A.1 – Le GRBM du cycle cellulaire utilisé pour la preuve de concept des chapitres 6 et 7

Résumé

Le cycle cellulaire est une séquence d'évènements *via* lesquels une cellule se reproduit tout en préservant l'intégrité de son génome. Cette propriété de préservation est assurée par un ensemble de mécanismes de régulation appelés *checkpoints*. Leur capacité à arrêter le cycle cellulaire en réponse à des lésions de l'ADN motive de nombreuses recherches dans le secteur biomédical. Cette thèse défend l'idée que ce phénomène d'arrêt, observé dans un contexte de perturbation, est un corollaire de propriétés plus fondamentales telles que la séparation temporelle des phases et l'irréversibilité des transitions de phases. Une phase décrit dans ce contexte un aspect particulier d'une cellule en prolifération, par exemple la phase de réplication du génome (phase S) ou celle de division de la cellule (phase M). Les propriétés fondamentales des checkpoints n'ont jamais été formalisées essentiellement parce que la modélisation des phases n'avait elle-même jamais été abordée jusque-là. La thèse présente un modèle discret de la régulation du cycle cellulaire et propose une méthodologie de formalisation d'une phase. Elle est initialement décrite par un multiensemble fini d'évènements discrets qui ont lieu à partir d'un état initial. En choisissant le cadre de modélisation de René Thomas, on s'autorise à appliquer la logique de Hoare génétiquement modifiée pour calculer l'ensemble des séquences d'évènements, exhibées par le multiensemble initial, qui sont compatibles avec un certain ensemble de paramétrisations d'un graphe de régulation. L'ensemble des séquences d'évènements qui caractérise une phase isole temporellement cette phase vis-à-vis des autres s'il existe une paramétrisation telle que tout évènement qui peut clore une phase est requis avant tout évènement qui peut initier la phase suivante. Ce problème a été spécifié à l'aide d'un prédicat Prolog. La propriété d'irréversibilité des transitions de phases est quant à elle formalisée à l'aide d'une logique temporelle et le model-checking identifie les paramétrisations d'un modèle du cycle cellulaire qui satisfont cette propriété. Nous montrons que cette méthodologie de formalisation et de vérification des propriétés fondamentales des checkpoints permet de cibler les contradictions d'un graphe de régulation en s'affranchissant de certains choix arbitraires que l'on retrouve dans la plupart des études de modélisation du cycle cellulaire.

Abstract

The cell cycle is a series of events by which a cell reproduces itself while maintaining the integrity of its genome. This property is ensured by a set of regulatory mechanisms called *checkpoints*. Their ability to stop the cell cycle in response to DNA damage has motivated numerous studies in the field of biomedicine. This thesis argues that the phenomenon of cell cycle arrest, which is observed in the context of perturbation, is a consequence of more fundamental properties such as the temporal separation of phases and the irreversibility of phase transitions. In this context, a phase describes a particular aspect of a proliferating cell, for example the phase of genome replication (S-phase) or that of cell division (M-phase). The fundamental properties of checkpoints have never been formalized, mainly due to the fact that phase modelling itself has never been tackled before. The thesis presents a discrete model of the regulation of the cell cycle and proposes a methodology for formalizing a phase. The latter is firstly characterized by a finite multi-set of discrete events which occur from an initial state. By opting for René Thomas' modelling framework, the genetically modified Hoare logic can be applied to compute the set of sequences of events, exhibited by the initial multiset, that are consistent with a certain set of parameterizations of a regulation graph. The set of event sequences that characterizes a phase temporally insulates it from the others if there is a parameterization such that any event that can complete a phase is required before any event that can trigger the next phase. This problem has been specified using a Prolog predicate. The irreversibility property of phase transitions is formalized using a temporal logic and model-checking identifies the parameterizations of a cell cycle model that satisfy this property. We show that this method for formalizing and verifying the fundamental properties of checkpoints allows us to target the inconsistencies of a regulatory graph while getting rid of certain arbitrary choices found in most cell cycle modeling studies.