



Analyse mathématique de la dynamique de réseaux de régulation biologique

Ousmane Diop

► To cite this version:

Ousmane Diop. Analyse mathématique de la dynamique de réseaux de régulation biologique. Automatique / Robotique. Université Paris-Saclay, 2020. Français. $\langle$ NNT : 2020UPASG013 $\rangle$. $\langle$ tel-03166968 $\rangle$

HAL Id: tel-03166968

<https://theses.hal.science/tel-03166968v1>

Submitted on 11 Mar 2021

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.



HAL Authorization

Analyse mathématique de la
dynamique de réseaux de
régulation biologique
Mathematical analysis of the
dynamics of biological
regulatory networks

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 580, Sciences et technologies de
l'information et de la communication (STIC)
Spécialité de doctorat : Automatique
Unité de recherche : Université Paris-Saclay, INRAE, MaIAGE, 78350,
Jouy-en-Josas, France
Réfèrent : Faculté des sciences d'Orsay

**Thèse présentée et soutenue à Paris-Saclay, le 8 décembre
2020, par**

Ousmane Diop

Composition du jury :

Madalena Chaves Directrice de Recherche, Inria, Université Côte d'Azur	Présidente
Claudine Chaouiya Maître de Conférences HDR, Université d'Aix-Marseille	Rapporteuse & Examinatrice
Étienne Farcot Professeur Assistant, University of Nottingham	Rapporteur & Examineur
Franck Delaunay Professeur, Université Côte d'Azur	Examineur
Dorothée Normand-Cyrot Directrice de Recherche, CNRS, Université Paris-Saclay	Examinatrice
Thomas Schiex Directeur de Recherche, INRAE, Université de Toulouse	Examineur
Vincent Fromion Directeur de Recherche, INRAE, Université Paris-Saclay	Directeur
Laurent Tournier Chargé de Recherche, INRAE, Université Paris-Saclay	Coencadrant

A mon père qui m'a quitté six mois après le début de ma thèse. Voici quelques strophes pour me souvenir de sa sagesse à chaque fois que je relis ce manuscrit :

Mon père, mon repère

Je t'ai tendu mes deux mains avant que je parte,
Hélas ! Je ne suis même pas revenu avant que tu ne partes.
Si seulement tu m'avais dit que je n'allais plus te revoir,
J'aurais rendu une éternité notre au-revoir.

Tu es parti me laissant un vide pour toujours ;
Me voilà en fin noyé dans cet océan de souvenirs.
Souvenirs, souvenirs, témoins de la tendresse paternelle,
Mon amour pour toi ne peut être qu'éternel.

Je t'écris ce poème en l'encre du beau sang que tu m'as transmis
Tout en espérant que de cette souffrance, je me sois remis.
Souffrance qui me hante depuis que je t'ai perdu, père
Tu as toujours été pour beaucoup, un véritable repère.

Tu as choisi de nous quitter en ce jour du onze mars,
Après avoir mené toute une vie de patriarche.
Sagesse était le maître mot te définissant ;
Car il a été le témoignage des plus reconnaissants.

Je suis fier de t'avoir comme papa, je le dis !
Mais puisses-tu l'être pour ton fils, même au paradis ?
De toute mes forces, je m'agripperai à ce que tu m'as dit.

Ousmane Diop, ton enfant.

Remerciements

A l'issue de ces trois années de doctorat, je suis plus que convaincu que le travail d'une thèse ne peut être un exercice complètement solitaire. Heureusement que j'ai eu la chance d'être entouré par des personnes dont chacune, à sa manière, a contribué à la réussite de ce travail. Il est temps de leur dire merci !

Tout d'abord, j'adresse mes remerciements les plus chaleureuses à l'endroit de Vincent Fromion et de Laurent Tournier pour leur accompagnement sans faille durant tout le long de cette thèse. De par leurs qualités scientifiques et humaines, je ne peux imaginer la chance que j'ai d'avoir travaillé avec eux. Vincent, mon directeur de thèse, qui, sa passion pour la recherche, ses qualités de chercheur ainsi que sa culture scientifique font de lui un abreuvoir pour l'assoiffé de science que je suis. Laurent, mon encadrant, je ne sais même pas quoi dire ... ! Il m'a fait confiance. Il m'a assisté et il a été très patient avec moi surtout lorsqu'il fallait qu'il m'initie à l'approche des systèmes dynamiques discrets. Merci à vous deux, j'ai beaucoup appris avec vous.

Merci également à Claudine Chaouiya et Étienne Farcot d'avoir accepté de rapporter ma thèse. La pertinence de leurs suggestions et commentaires a montré toute l'attention qu'ils ont accordé à mon manuscrit malgré leur emploi du temps très chargé.

Merci à Madalena Chaves pour tout son soutien depuis le début à travers le projet ANR Icycle dans lequel ma thèse est inscrite. En plus de tout son accompagnement, elle m'a fait l'honneur de présider mon jury de soutenance. Merci également à Franck Delaunay, Dorothée Normand-Cyrot et Thomas Schiex d'avoir accepté d'examiner ce travail et de faire partie de mon jury de soutenance.

Merci à tous les membres du laboratoire MaiAGE de l'INRAE et plus particulièrement ceux de l'équipe BioSys pour leur sens de l'écoute et pour leur serviabilité. Malheureusement, je ne peux citer tout le monde mais sachez qu'en écrivant ces lignes, vos noms bourdonnent dans ma tête.

Merci à toutes ces personnes qui, de façon consciente ou inconsciente, ont participé à la réussite de cette thèse.

En fin un grand merci à tous les membres de ma famille pour leur soutien depuis ma venue au monde.

Table des matières

Introduction générale	11
1 Réseaux de régulation cellulaires et oscillateurs biologiques	13
1.1 Description du cycle de division cellulaire et de l'horloge circadienne . . .	14
1.1.1 Deux oscillateurs cellulaires centraux	14
1.1.2 Régulation du cycle de division des cellules mammifères	15
1.1.3 Régulation de l'horloge circadienne des cellules mammifères	19
1.2 Modélisation quantitative des oscillateurs biologiques par des équations différentielles ordinaires	21
1.3 Modélisation qualitative des oscillateurs biologiques par des réseaux Booléens	24
1.3.1 Les réseaux Booléens asynchrones	25
1.3.2 Ajout de probabilités de transition dans le modèle Booléen	26
1.3.3 Attracteurs asynchrones complexes et oscillations cellulaires . . .	28
2 Analyse d'attracteurs Booléens complexes par un graphe résumé : application à un modèle du cycle cellulaire	31
2.1 Définition du graphe résumé d'un attracteur	32
2.1.1 Graphe résumé d'un attracteur sur une partition	32
2.1.2 Graphe résumé d'un attracteur sur une semi-partition	34
2.1.3 Propriété centrale du graphe résumé d'un attracteur	36
2.2 Utilisation du graphe résumé pour l'analyse d'un modèle du cycle cellulaire	37
2.2.1 Construction d'une première partition des états de l'attracteur .	37
2.2.2 Raffinement de la partition	39
2.2.3 Analyse de l'attracteur grâce au graphe résumé	40
2.2.4 Validation de mutants à l'aide du graphe résumé	42
2.3 Graphe résumé probabilisé d'un attracteur	45
2.3.1 Probabilités de transition sur un graphe résumé	45
2.3.2 Estimation du temps passé dans les phases du cycle cellulaire . .	46
2.3.3 Prise en compte des vitesses relatives dans la chaîne de Markov .	48
3 Construction et analyse d'un modèle du cycle circadien	53
3.1 Construction d'un modèle Booléen de l'horloge circadienne	54
3.1.1 Modélisations continues de la dynamique de l'horloge	54
3.1.2 Élaboration du modèle Booléen	55
3.1.3 Analyse préliminaire de la dynamique du réseau	60
3.2 Analyse du modèle par un graphe résumé	61

3.2.1	Identification de phases qualitatives du cycle circadien	61
3.2.2	Construction et analyse du graphe résumé	63
3.2.3	Validation des mutations à l'aide du graphe résumé	64
3.3	Analyse temporelle des oscillations circadiennes	65
3.3.1	Estimation du temps passé dans les phases du cycle	66
3.3.2	Extension : étude du déphasage du cycle	67
4	Utilisation du graphe résumé pour ajuster la dynamique d'un attracteur complexe	71
4.1	Analyse avancée de l'attracteur du cycle cellulaire	72
4.1.1	Vers un modèle sans court-circuits	72
4.1.2	Suppression séparée des court-circuits dans l'attracteur	75
4.2	Vers un modèle simplifié du cycle cellulaire	76
4.2.1	Projection des états et construction d'un graphe de transition réduit	77
4.2.2	Analyse du modèle réduit	78
4.3	Utilisation du graphe résumé comme aide à la modélisation, application au réseau du cycle circadien	80
4.3.1	Elimination de certains états non classés	80
4.3.2	Analyse des court-circuits du graphe résumé	82
	Conclusion et perspectives	85
A	Articles parus pendant la thèse	87

Table des figures

1.1	Schéma simplifié d'une cellule eucaryote.	13
1.2	Les quatres phases $G1$, S , $G2$ et M du cycle cellulaire et la phase de quiescence $G0$	16
1.3	Les étapes de la mitose : Prophase, Métaphase, Anaphase, Télophase.	17
1.4	Mécanismes de contrôle du cycle de cellule mammifère par le système cycline/Cdk.	18
1.5	Schéma simplifié du mécanisme moléculaire régissant le réseau de régulation de l'horloge circadienne chez les mammifères.	20
1.6	Illustration des oscillations temporelles du modèle de Gérard et Golbeter [32].	23
1.7	Simulation temporelle du modèle d'Almeida <i>et al</i> 2020 [4].	24
2.1	Attracteur de 12 états du réseau (2.1).	32
2.2	Graphe résumé de l'attracteur du réseau (2.1).	33
2.3	Graphe résumé de l'attracteur du réseau (2.1) sur la partition $\mathcal{P}_2$	34
2.4	Graphe résumé du réseau (2.1) sur la semi-partition $\mathcal{S}$	36
2.5	Graphe résumé $A(\mathcal{P}(\mathcal{S}))$	36
2.6	Graphe résumé sur la semi-partition obtenue avec la table 2.1	39
2.7	Graphe résumé $A(\mathcal{P})$ de l'attracteur complexe du modèle dans [26]	41
2.8	Graphe résumé $A(\mathcal{S})$ avec $\mathcal{S} = \{G1, G1/S, SG2, G2/M, M, M/G1\}$	42
2.9	Graphes résumés des mutants $Rb=0$, $P27=0$, $CycE=1$ et $E2F=1$	44
2.10	Proportions du temps passé dans chaque phase du cycle cellulaire.	47
2.11	Évolution de la proportion du temps passé dans chacune des trois phases ω_{G1} , ω_{SG2} et ω_M et les proportions cumulées ($\omega_{G1/S} + \omega_{G2/M} + \omega_{M/G1}$) des phases de transitions en fonction des poids ρ_1 ou ρ_3 . Chaque ω_i correspond à la i -ième composante de la distribution stationnaire ω obtenue en variant de façon indépendante ρ_1 ou ρ_3	49
2.12	Proportions du temps passé dans chaque phase du cycle, calculées avec les poids $\rho = (50, 1, 100, 1)$	50
3.1	Graphe d'interaction du modèle Booléen de l'horloge.	59
3.2	Attracteur synchrone de 5 états du modèle de l'horloge.	60
3.3	Attracteur asynchrone complexe du modèle de l'horloge.	60
3.4	Description des principales étapes qualitatives de l'horloge circadienne.	62
3.5	Graphe résumé de l'attracteur sur la partition $\mathcal{P} = \{V_1, \dots, V_5, U\}$, les V_i étant définis par (3.33).	63
3.6	Graphe résumé du modèle Booléen de l'horloge ainsi que des trois mutants $PER = 0$, $CRY = 0$ et $REV = 0$	66

3.7	Distributions stationnaires ω des chaînes de Markov du modèle wild type et du mutant $REV = 0$	67
3.8	Analyse de sensibilité du modèle Booléen de l'horloge.	69
4.1	Graphe des dépendances entre les variables $E2F, CycE, CycA$ et $CycB$. .	74
4.2	Graphe résumé de l'attracteur du modèle du tableau 4.1.	74
4.3	Projection de l'attracteur initial sur les quatre variables $E2F, CycE, CycA$ et $CycB$: 16 états et 44 transitions asynchrones.	78
4.4	Modèle réduit.	78
4.5	Graphe résumé du modèle réduit sur la partition $\tilde{P}$	79
4.6	Graphes résumés du modèle de l'horloge, avant et après suppression de court-circuits.	81

Liste des tableaux

1.1	Règles d'activation des variables du modèle de Fauré <i>et al.</i> , [26]	27
2.1	Une première semi-partition de l'attracteur du modèle de [26].	39
2.2	Raffinement de la semi-partition : table de vérité des six phases en fonction des quatre variables E2F, CycE, CycA et CycB.	40
2.3	Quatre mutations oscillantes du modèle de [26].	43
2.4	Tables de vérité pour reclasser les états des mutants CycE=1 et E2F=1.	44
2.5	Quatre classes de priorité de transition proposées dans [26].	49
3.1	Activateurs et inhibiteurs des trois CCEs (clock controlled elements)	55
3.2	Inhibiteur et activateur des CCEs avec prise en compte des variables multivaluées.	57
3.3	Règles Booléennes des 11 variables du modèle de l'horloge.	59
3.4	Effet de quelques mutations sur la dynamique du modèle Booléen de l'horloge.	64
4.1	Version du modèle de [26] inférée à partir du nouvel attracteur.	73
4.2	Les 16 états projetés de l'attracteur de [26] sur les quatre variables <i>E2F</i> , <i>CycE</i> , <i>CycA</i> et <i>CycB</i> .	77

Introduction

La vie des organismes est rythmée par des phénomènes divers et variés agissant à des échelles de temps très différentes, de l'échelle de la seconde comme les battements cardiaques à celle du jour comme le cycle veille-sommeil, voire au-delà. L'étude de ces phénomènes est au centre de la discipline récente nommée chronobiologie. En particulier, de nombreux phénomènes biologiques montrent un caractère cyclique, observable à l'échelle d'une population comme à l'échelle d'un individu. Au sein même d'un organisme, on observe des oscillations internes qui vont jusqu'au niveau de la cellule qui est la plus petite unité de vie. Ces phénomènes cycliques sont centraux pour appréhender la vie des cellules et leur compréhension constitue un enjeu majeur de la biologie cellulaire. Ces dernières décennies, les progrès des sciences expérimentales ont permis de mieux en mieux cerner ce qui se passe à l'intérieur des cellules, et notamment d'observer la variation périodique de différents constituants intracellulaires. A l'origine de ces variations, on trouve des réseaux de régulation structurés. Ces réseaux assurent un ordre précis dans les oscillations et garantissent leur robustesse. Certains de ces réseaux sont très étudiés et leurs bases moléculaires ont déjà été largement élucidées, grâce notamment à de nombreux travaux expérimentaux. Néanmoins, la compréhension fine de la dynamique de ces réseaux complexes reste délicate. Dans ce domaine, la modélisation mathématique occupe une place centrale.

De manière générale, un modèle mathématique est la représentation abstraite d'un phénomène. Sa construction se base sur des hypothèses rationnelles, simplifiant le phénomène biologique afin de rendre son analyse plus intuitive, tout en tenant compte, le plus possible, des connaissances disponibles. Dans le cas de la biologie cellulaire, la modélisation des réseaux de régulation fait largement appel à la théorie des systèmes dynamiques. Une fois construit, la première utilisation d'un modèle réside dans sa simulation, qui permet de visualiser comment il se comporte dans le temps et ainsi le comparer au phénomène qu'il modélise. Grâce à sa nature abstraite, un modèle peut être ensuite simulé facilement dans des contextes différents, comme une mutation par exemple, offrant ainsi la possibilité de valider le modèle à partir des informations biologiques disponibles. On peut même simuler des conditions qui n'ont pas encore été testées expérimentalement. Ces "prédictions" permettent alors de suggérer de futures expériences à tester, afin d'avancer dans la compréhension du mécanisme biologique modélisé. Une fois suffisamment validé, l'analyse approfondie du modèle permet d'éclairer la dynamique de ce mécanisme. Par exemple, on peut analyser la réponse du modèle à la variation de certains paramètres clés, indiquant ainsi la sensibilité ou au contraire la robustesse de certains comportements.

Dans cette thèse, nous nous intéressons à l'analyse de la dynamique de deux cycles biologiques centraux, le cycle de division cellulaire et l'horloge circadienne, par une classe de systèmes dynamiques discrets. Les réseaux de régulation derrière ces deux cycles sont

bien connus à l'échelle de la cellule. Nous proposons d'analyser leurs dynamiques d'un point de vue qualitatif, en nous focalisant sur la succession de phases discrètes dans un ordre déterminé. Pour cela, nous utilisons des réseaux Booléens asynchrones, adaptés à de telles analyses qualitatives. Dans ces réseaux, les oscillations sont capturées par des *attracteurs*, qui sont des graphes fortement connexes pouvant contenir des centaines voire des milliers d'états. Afin de comparer ces attracteurs avec les oscillations biologiques, nous proposons une méthode originale pour les représenter de façon simplifiée, en nous basant sur des informations biologiques. Cette méthode nous permet ainsi de repérer dans les attracteurs l'enchaînement de phases qualitatives rythmant les deux cycles précédemment cités pour ensuite valider les modèles Booléens. A travers l'étude de ces deux cycles, la méthode proposée dans cette thèse se présente comme un outil tout à fait pertinent pour analyser des systèmes biologiques présentant des oscillations complexes.

Plan du manuscrit. Dans le chapitre 1, après une brève description des réseaux derrière les deux cycles au cœur de cette thèse, nous faisons un état des lieux succinct de différents modèles de ces réseaux basés sur des équations différentielles ordinaires. Ensuite, nous introduisons les réseaux Booléens asynchrones qui sont les modèles dynamiques que nous utiliserons principalement dans la thèse.

Dans le chapitre 2, nous donnons une définition du concept de graphe résumé, au cœur de notre approche. Nous l'utilisons pour analyser le modèle Booléen du cycle cellulaire de [26], et nous montrons notamment que ce modèle capture bien la succession attendue des phases du cycle. En combinant le graphe résumé avec une chaîne de Markov, nous montrons également comment estimer, au sein d'un modèle Booléen, le temps passé dans les phases du cycle.

Nous proposons dans le chapitre 3 un nouveau modèle Booléen de l'horloge circadienne des cellules mammifères basé sur le modèle différentiel de [4]. En appliquant notre méthode, nous montrons la succession de différentes phases qualitatives dans notre modèle. Nous poussons ensuite l'étude du modèle pour analyser un déphasage du cycle.

Enfin, dans le chapitre 4, nous proposons une analyse approfondie des deux modèles précédents. Nous cherchons notamment à supprimer certaines transitions de phase intempestives, en suggérant des modifications locales des règles des modèles. Nous montrons ainsi comment notre méthode permet d'ajuster localement la dynamique asymptotique de réseaux Booléens asynchrones, fournissant un outil puissant dans la modélisation de réseaux de régulation biologique.

Chapitre 1

Réseaux de régulation cellulaires et oscillateurs biologiques

La cellule est le plus petit ensemble cohérent de structures et de fonctions vitales. Elle peut être définie comme la plus petite “unité de vie” capable de manifester les propriétés essentielles d’un être vivant : se nourrir, croître et se reproduire [16]. Toutes les formes vivantes sont constituées de cellules, et chaque cellule contient en elle-même tous les attributs du vivant. On distingue deux grands types de cellules : les cellules procaryotes sont dépourvues de noyau et les cellules eucaryotes qui contiennent un noyau à l’intérieur duquel est conservé le matériel génétique. Dans le cadre de cette thèse, nous nous intéressons principalement à des cellules mammifères, qui sont des cellules eucaryotes. La figure 1.1 décrit de manière très simplifiée une cellule eucaryote. Elle est entourée par une membrane biologique appelée membrane plasmique, qui délimite l’intérieur de l’extérieur de la cellule. La membrane plasmique sépare la cellule de son environnement et lui permet notamment d’échanger avec celui-ci. L’intérieur de la cellule (le cytoplasme) est compartimenté en plusieurs sous-systèmes, appelés organites, eux-même délimités par des membranes. Le noyau est l’un de ces organites, il représente un compartiment central de la cellule eucaryote car il contient le matériel génétique, sous la forme de molécules linéaires d’ADN (acide désoxyribonucléique) qui, associées à d’autres molécules, forment les chromosomes.

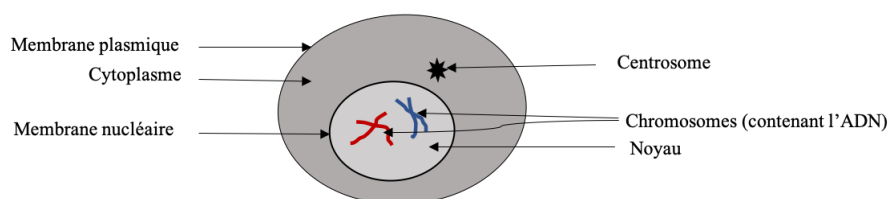


FIGURE 1.1 – Schéma simplifié d’une cellule eucaryote. Le centrosome est un organite impliqué dans le processus mitotique.

La division d’une cellule en deux cellules filles, contenant chacune exactement le même matériel génétique que la cellule mère, est au cœur de la vie d’un organisme. Ce processus de division, appelé *mitose*, est cyclique. La cellule commence par croître, ce faisant elle est amenée à produire notamment de nouveaux organites. Durant cette croissance,

le matériel génétique est dupliqué en deux copies identiques. A la fin du processus, deux nouveaux noyaux sont produits, chacun contenant une copie, puis la membrane elle-même se sépare pour produire les deux cellules filles, chacune identique à la cellule mère, terminant ainsi un cycle. Ce processus cyclique, au cœur de la vie de l'organisme, illustre le côté fondamental des oscillations en biologie, et notamment en biologie cellulaire. Deux cycles centraux rythment ainsi la vie des cellules eucaryotes. Outre le cycle de division (simplement appelé cycle cellulaire en général), le cycle circadien permet à la plupart des organismes de rythmer leurs activités selon l'alternance jour/nuit. Ces deux "systèmes" oscillatoires sont tous les deux commandés, au sein des cellules, par des *réseaux de régulation* qui ont pour rôle de générer ces cycles et d'assurer leur bon déroulement.

Dans cette thèse, nous nous intéressons précisément aux deux réseaux de régulations qui génèrent le cycle de division et le cycle circadien. Nous proposons d'analyser la dynamique de ces deux systèmes par la modélisation mathématique, et notamment par l'utilisation de modèles discrets, particulièrement bien adaptés à ce type de systèmes. Ils permettent en particulier de proposer une description qualitative de la dynamique des réseaux de régulation. Dans le cas de systèmes présentant des oscillations, cette description qualitative est plutôt judicieuse car proche de la description biologique. En effet, le plus souvent ce genre de cycles biologiques est décrit en terme de succession d'événements discrets (ou de "phases"), faisant de la modélisation discrète un outil de description très pertinent.

Dans ce chapitre, nous proposons tout d'abord une brève description des deux oscillateurs au cœur de cette thèse, en faisant notamment dans la partie 1.1 un état des lieux des connaissances sur les réseaux de régulation sous-jacents. Dans la partie 1.2, nous présentons différents modèles de ces réseaux basés sur des équations différentielles ordinaires. Dans la partie 1.3, nous introduisons les réseaux Booléens asynchrones, qui sont plus qualitatifs et qui sont les modèles que nous utiliserons principalement dans la thèse.

1.1 Description du cycle de division cellulaire et de l'horloge circadienne

1.1.1 Deux oscillateurs cellulaires centraux

En biologie, de nombreux processus essentiels présentent des oscillations. On peut citer les battements cardiaques, la respiration, le cycle de division cellulaire ou encore le cycle veille-sommeil. L'analyse des "rythmes" biologiques est au cœur de la chronobiologie, à l'interconnexion de plusieurs disciplines scientifiques. Parmi les nombreux cycles biologiques existant, ceux qui nous intéressent ici sont les cycles ayant lieu à l'échelle de la cellule, c'est à dire ceux qui sont marqués par une oscillation de constituants intracellulaires. Ces oscillations se répètent, de manière remarquablement robuste, sous une même forme et à des intervalles de temps réguliers [35]. Elles rythment la vie de la cellule, en conditionnant la survenue d'un ensemble d'événements, se succédant de façon ordonnée. La période des oscillations constitue une grandeur caractéristique du rythme et, selon cette valeur, les rythmes peuvent être répartis en trois catégories : ultradien, infradien et circadien. Les oscillations ultradiennes s'étendent sur une période inférieure

à 20 heures, comme par exemple le rythme cardiaque (de l'ordre de la seconde). Les oscillations infradiennes ont une période supérieure à 28 heures, comme le cycle ovarien (environ 28 jours). Quant aux oscillations circadiennes (*circa diem*, *i.e* environ un jour), elles se produisent sur une période autour de 24 heures et incluent les deux systèmes qui nous intéressent ici, l'horloge circadienne et le cycle de division cellulaire. Ces deux systèmes sont au centre de la vie de la plupart des cellules eucaryotes et ils interviennent, directement ou indirectement, dans de nombreux mécanismes intracellulaires.

Grâce aux nombreux travaux expérimentaux réalisés sur plusieurs organismes modèles, les mécanismes moléculaires qui sous-tendent le fonctionnement de ces deux oscillateurs sont aujourd'hui relativement bien connus. Les oscillations du cycle cellulaire, en particulier, sont décrites depuis de nombreuses années en terme de succession de phases [21], différentes tant au niveau de leurs durées que des processus intracellulaires en jeu (croissance cellulaire, duplication de l'ADN, *etc*). Quant à l'horloge circadienne, des travaux ont montré depuis longtemps des oscillations intracellulaires qui se synchronisent à l'alternance jour-nuit, par deux phases grossières se déroulant respectivement durant la journée et durant la nuit [48]. Dans ces deux systèmes biologiques, les oscillations intracellulaires se décrivent à chaque fois sous la forme de phases qualitatives, dont la succession respecte un ordre fixe et prédéterminé, et chacune étant régulée par un mécanisme sous-jacent. L'existence de ces phases qualitatives est fondamentale : une perturbation dans l'ordre de succession de ces événements est souvent à l'origine de pathologies sévères telles que le développement de certains cancers par exemple.

En raison de leur caractère central, ces deux systèmes biologiques ont été au centre de nombreuses recherches en biologie cellulaire. D'ailleurs, les travaux réalisés sur ces deux systèmes ont fait l'objet de deux prix Nobel de physiologie et de médecine, l'un en 2001 sur le cycle cellulaire (Leland Hartwell, Timothy Hunt et Paul Nurse) et l'autre sur l'horloge en 2017 (Jeffrey C. Hall, Michael Rosbash et Michael W. Young). Ces travaux, entre autres, ont permis de générer de nombreuses données, qualitatives comme quantitatives, et ainsi de mettre au jour les réseaux de régulation moléculaire qui sont à l'origine des deux cycles. Nous proposons dans la suite un bref résumé des principaux mécanismes intracellulaires en jeu, pour le cycle cellulaire puis pour l'horloge circadienne.

1.1.2 Régulation du cycle de division des cellules mammifères

Déroulement dynamique du cycle de division

Comme évoqué plus haut, le cycle de division cellulaire est le processus par lequel une cellule mère donne naissance à deux cellules filles dotées du même matériel génétique. S'il peut y avoir des nuances selon le type de cellule, ce cycle consiste principalement en deux grandes étapes : l'interphase et la mitose. L'interphase est composée de trois phases qui sont la phase *G1*, la phase *S* et la phase *G2*. La mitose *M*, quant à elle, représente la division proprement dite de la cellule (voir Fig. 1.2).

Lorsque les cellules ne se divisent pas, elles sont dites dans une phase de quiescence, généralement notée *G0*. Lorsque les conditions extérieures l'exigent, indiqué par la présence en quantité suffisante d'un *facteur de croissance*, le cycle de division se déclenche par l'entrée dans la phase *G1*. Durant la phase *G1*, la cellule augmente en taille et effectue toutes les fonctions pour lesquelles elle est génétiquement programmée. La phase *S*, appelée phase de synthèse, consiste en la duplication du matériel génétique, et donc la

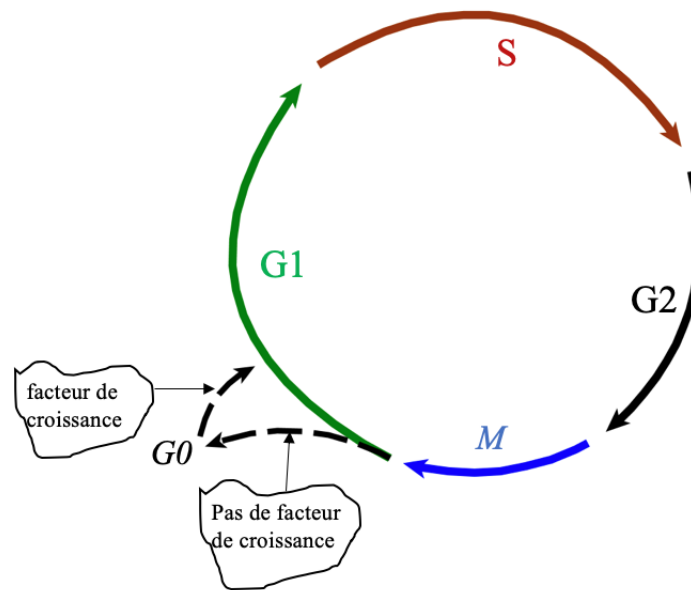


FIGURE 1.2 – Les quatre phases $G1$, S , $G2$ et M du cycle cellulaire et la phase de quiescence $G0$. Les cellules restent en $G0$ en l’absence de facteur de croissance et commencent leur cycle de division en présence d’une quantité suffisante de facteur de croissance.

production de nouvelles molécules d’ADN. La phase S est une étape clé du cycle puisqu’elle permet de préparer le matériel génétique que les cellules filles vont se partager à la fin du cycle. Après la phase S , la cellule entre à nouveau dans une phase de croissance qui est la phase $G2$. Au même titre que la phase $G1$ prépare la phase S , la phase $G2$ permet à la cellule de préparer la phase de mitose M . La mitose est la phase de division cellulaire proprement dite. Elle est plutôt rapide (relativement aux précédentes) et se découpe également en différentes étapes temporelles, décrites dans la figure 1.3. La première étape est appelée prophase : la membrane nucléaire commence à disparaître et les chromosomes fils issus de la duplication pendant la phase S se condensent. Après la prophase, la cellule entre en métaphase : les deux centrosomes fils migrent vers les pôles opposés et maintiennent à travers les fibres du fuseau mitotique (ou microtubules) les chromosomes fils le long de la plaque équatoriale. Dans le cas où la réplication s’est correctement effectuée et que tous les chromosomes se sont alignés, la cellule entre en anaphase avec une séparation des chromatides fils et leur migration vers les pôles. La fin de la mitose est marquée par la télophase, divisant la membrane cellulaire pour permettre la différenciation des deux cellules filles qui retournent dans la phase de quiescence $G0$.

Mécanismes de régulation et de surveillance

Le déroulement du cycle cellulaire et son organisation en phases est rendu possible grâce à un réseau de régulation contrôlé par trois familles de protéines : les cyclines, les Cdks (*Cyclin-dependent kinases* ou Cyclines Kinases dépendantes) et les CKI (*Cyclin kinase inhibitor* ou Cyclines inhibitrices de Kinases). Ces protéines agissent en complexes,

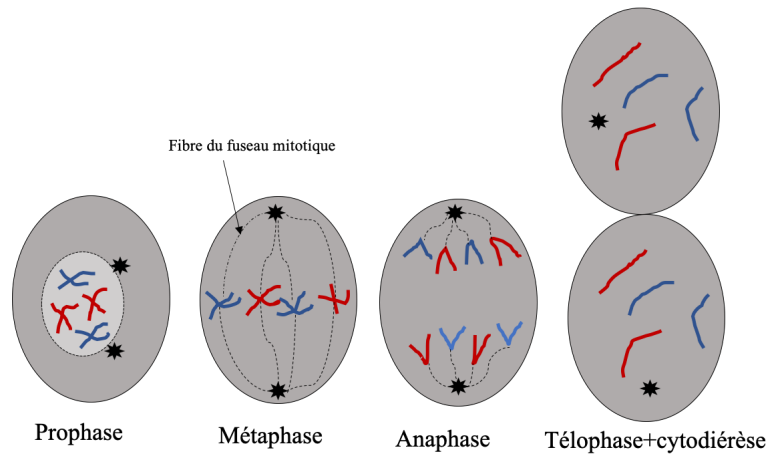


FIGURE 1.3 – Les étapes de la mitose : Prophase, Métaphase, Anaphase, Télophase.

portant le nom de “système Cycline/Cdk”. L’élucidation du rôle de ces familles de protéines est à l’origine du prix Nobel de 2001 cité plus haut. Il a été montré notamment que le principe de fonctionnement de ces systèmes est le même chez la levure, les plantes et l’Homme. Les Cdk sont des enzymes qui ont besoin, pour être actifs, de se lier à des activateurs spécifiques qui sont les cyclines [28]. L’activité du complexe Cycline/Cdk dépend directement du niveau de production des cyclines. Dans le cas du cycle de cellules mammifères, quatre types de cyclines nommées D, E, A, et B et quatre types de Cdk nommées Cdk4, Cdk6, Cdk2, et Cdk1 sont les plus connues. La cycline D s’associe à Cdk4 ou à Cdk6, les cyclines E et A s’associent à Cdk2. La Cycline A s’associe également à Cdk1, au même titre que la cycline B. Le long des phases du cycle, les cyclines sont produites successivement dans un ordre nécessaire à l’activation des différents complexes : d’abord D, puis E, puis A et enfin B (voir Figure 1.4). En présence d’une quantité suffisante de facteur de croissance, une cellule en phase G0 commence son cycle de division [11]. La phase G1 constitue la porte d’entrée du cycle cellulaire : au cours de cette phase la cellule décide soit de continuer la progression dans le cycle, soit d’arrêter définitivement le cycle. Un certain nombre d’interactions entre les complexes cyclines/Cdk et leur inhibiteurs CKI sont à l’origine de cette prise de décision. Dès le début de la phase G1, la cycline D s’accumule sous la forme de complexes cycline D/Cdk4 puis cycline D/Cdk6 permettant la progression du cycle en inhibant par phosphorylation la famille des “protéines poches” (pRb, protéine du rétinoblastome p107 et p130) [79, 90]. Ces protéines sont les principaux répresseurs des facteurs de transcription de la famille E2F. Sous l’action maintenue des complexes cycline D/Cdk4,6, le facteur de transcription E2F favorise l’accumulation de la cycline E qui forme un complexe avec la Cdk2 [21]. La concentration du complexe Cycline E/Cdk2 est élevée à la fin de la phase G1 [72] et contribue à déclencher la phase S en favorisant, avec l’aide de E2F, la synthèse de la cycline A qui se lie avec la Cdk2 [58].

Afin d’assurer que les futures cellules filles ne présentent pas d’anomalies, la cellule a tout intérêt à vérifier de façon sûre un certain nombre de critères avant de passer à la phase S. Ce mécanisme de surveillance inclut la vérification que le ratio du volume du cytoplasme sur le volume du noyau a atteint un certain niveau. De plus, l’ADN de la cellule mère ne

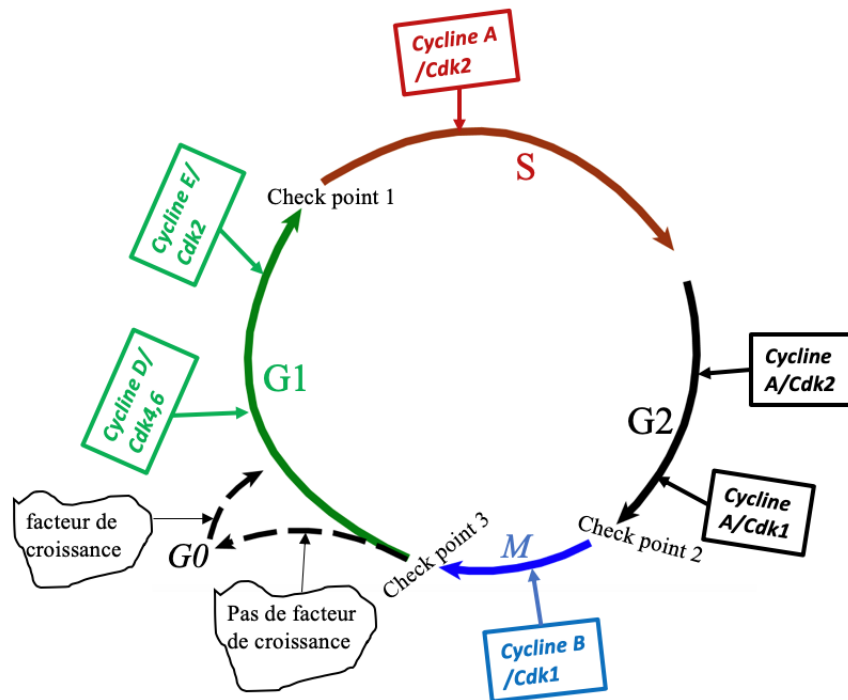


FIGURE 1.4 – Mécanismes de contrôle du cycle de cellule mammifère par le système cycline/Cdk.

doit présenter aucune défaillance dont les cellules filles pourraient hériter. Du point de vue de la régulation, grâce à l'action inhibitrice de la cycline D, les CKI restent dans un état hyperphosphorylé entraînant une accumulation des cyclines E et A essentielles à l'entrée en *S*. Durant cette transition *G1/S*, la décision consistant à autoriser la duplication de l'ADN est prise de façon irréversible [61, 10]. De façon similaire, la transition *G2/M* constitue un deuxième point de contrôle du cycle. Avant de passer à la phase de division *M*, la cellule vérifie que l'ADN est correctement et complètement dupliqué. Dans le cas où des défaillances se présentent dans l'ADN, un mécanisme de correction faisant intervenir des protéines comme p53 est déclenché. Durant tout le long des phases *S* et *G2*, le complexe Cycline B/Cdk1 s'accumule, mais dans un état inactif car la Cdk1 est phosphorylée par la kinase Wee1. A la fin de la phase *G2*, la phosphatase Cdc25 permet d'activer la Cdk1. Cette activation permet l'activation du complexe CycB/Cdk1[42]. En parallèle, la cycline A se lie également à Cdk1 pour former le complexe CycA/Cdk1. Ces deux complexes permettent alors le déclenchement de la phase *M*. Dans le chapitre 2, nous nous baserons essentiellement sur le modèle proposé dans [26], qui propose une vue assez complète des régulations du cycle chez les cellules mammifères.

1.1.3 Régulation de l’horloge circadienne des cellules mammifères

La notion de cycle circadien au niveau cellulaire

L’horloge circadienne est observée pour la première fois par Jean-Jacques Dortous de Mairan en 1729 [35]. Mairan remarque en effet que les feuilles d’une plante (la sensitive) s’ouvrent et se ferment en suivant le rythme du jour et de la nuit. Des années plus tard, Augustin de Candolle montre que ces oscillations circadiennes persistent même en obscurité constante et sont donc indépendantes de l’alternance jour-nuit, confirmant ainsi leur caractère endogène [35]. Depuis, l’horloge circadienne est observée chez la plupart des organismes vivants [97, 83, 3, 47] et elle leur permet de s’adapter au caractère périodique de l’environnement terrestre et d’anticiper l’alternance du cycle lumière-obscurité.

Chez la plupart des organismes vivants, il est observé une horloge centrale, qui représente la pièce maîtresse du système circadien, et des horloges périphériques, de l’échelle des organes jusqu’à celle de la cellule. L’horloge centrale a pour rôle d’assurer le caractère endogène des oscillations et de synchroniser les horloges périphériques [59]. Chez les mammifères, l’horloge centrale est localisée dans le cerveau, plus précisément dans les noyaux suprachiasmatiques (NSC) de l’hypothalamus [63]. Il est à noter que la période des oscillations circadiennes n’est pas exactement égale à 24 h mais elle est souvent variable et est un peu plus longue chez la plupart des individus [18]. Cette variabilité de la période de l’horloge s’explique par l’influence de différents facteurs appelés *zeitgebers* (“donneurs de temps”, en allemand) que sont la lumière, la température, la nourriture, l’activité physique, *etc* [43]. Des perturbations de l’horloge peuvent intervenir à différents niveaux. En particulier, des perturbations d’ordre moléculaire peuvent entraîner un dysfonctionnement continu dans la synchronisation de l’horloge avec le cycle veille-sommeil. Ces perturbations conduisent à des syndromes tels que le syndrome d’avance de phase (FASPS, Familial Advanced Sleep Phase Syndrome). Les sujets atteints de ce syndrome ont un cycle veille-sommeil décalé, l’ordre chronologique des événements caractéristiques de l’horloge étant perturbé.

Mécanismes de régulation de l’horloge circadienne

L’élucidation des mécanismes moléculaires qui sous-tendent l’horloge circadienne des cellules mammifères a largement profité des travaux sur la drosophile des trois prix Nobel précédemment cités, Jeffrey C. Hall, Michael Rosbash et Michael W. Young [6, 40, 96]. Le réseau de régulation de l’horloge des cellules mammifères implique trois boucles de rétroaction principales faisant intervenir différents activateurs et inhibiteurs aux niveaux transcriptionnel et traductionnel [76] (voir Fig.1.5). Au cœur du réseau se trouvent les gènes *Bmal1* (Brain-muscle Arnt like 1) et *Clock* (Circadian Locomotor Output Cycles Kaput). Ces deux gènes codent pour les protéines Bmal1 et Clock qui jouent le rôle d’activateur principal en formant un complexe Clock:Bmal1. Dans la première boucle de rétroaction, ce complexe se lie à des régions du génome appelées *E-boxes* pour activer la transcription des familles de gènes *Per* (*Per1*, *Per2*, *Per3*) et *Cry* (*Cry1*, *Cry2*) [84]. Les protéines correspondantes Per et Cry s’accumulent avant de former un complexe Per:Cry, qui entre à son tour dans le noyau pour réprimer l’activité de Clock:Bmal1. En plus de cette première boucle négative, le complexe Clock:Bmal1 est impliqué dans

une seconde boucle négative en activant le gène *Rev-Erb α* qui code pour la protéine Rev-Erb α , inhibitrice de la transcription de Bmal1 et Clock [71]. Enfin, le complexe Clock:Bmal1 active sa propre transcription par une troisième boucle de rétroaction, cette fois positive, en activant la transcription du gène *Rora* dont la protéine correspondante active à son tour la transcription de Bmal1.

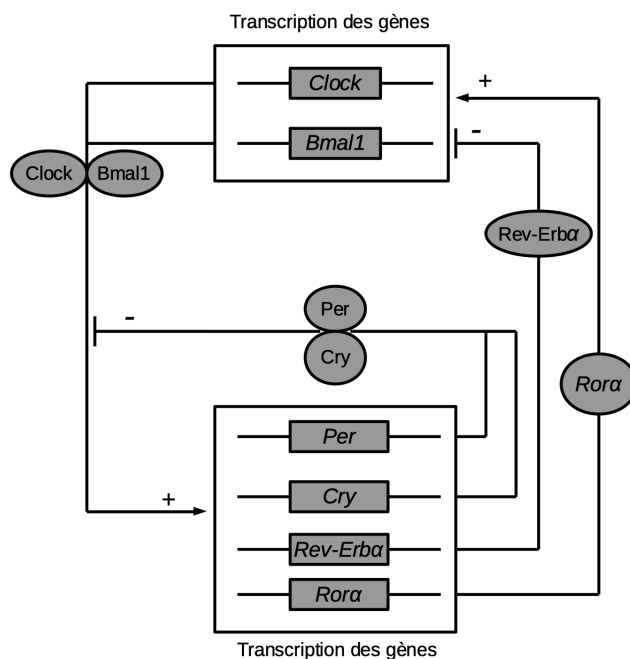


FIGURE 1.5 – Schéma simplifié du mécanisme moléculaire régissant le réseau de régulation de l’horloge circadienne chez les mammifères. Trois boucles de rétroactions (deux négatives et une positive) centrées autour du complexe Clock:Bmal1. Les gènes sont sous forme de rectangle et les protéines sont encadrées. Schéma adapté de [35].

La combinaison de ces trois boucles de rétroaction conduit à une oscillation des constituants cellulaires, qui suit l’alternance jour-nuit. Durant le cycle, la transcription des gènes et la production des protéines correspondantes sont organisées de façon chronologique, offrant la possibilité de repérer des zones temporelles durant lesquelles les étapes clés du processus biologique se produisent. En particulier, Clock:Bmal1 est globalement actif durant la journée tandis que le complexe Per:Cry est actif durant la nuit. Cette opposition de phase entre les deux complexes est une caractéristique de l’horloge circadienne des cellules mammifères : une forte activité de l’un correspond à une faible activité de l’autre. Dans le chapitre 3, nous reviendrons sur ce réseau en proposant un modèle dynamique discret et en définissant un ensemble de phases qualitatives.

1.2 Modélisation quantitative des oscillateurs biologiques par des équations différentielles ordinaires

Comme nous l'avons vu dans la partie précédente, le comportement dynamique des cellules est directement commandé par la présence de réseaux de régulation, impliquant divers activateurs ou inhibiteurs, reliés entre eux par des graphes d'interaction plus ou moins complexes. Ces descriptions sont essentiellement statiques, et différentes approches peuvent être utilisées pour modéliser la dynamique de ces réseaux du point de vue mathématique. L'approche la plus utilisée dans la littérature reste celle des équations différentielles ordinaires (EDO), que nous présentons brièvement à présent. Dans cette approche, nous considérons que les variables qui entrent en jeu dans les mécanismes de régulation sont au nombre de n . On dit que n est la dimension du système. On pose pour tout temps t , $x_i(t) \geq 0$ la concentration de l'entité x_i pour $i = 1, \dots, n$. Ces entités peuvent être des protéines, des ARN messagers, des facteurs de transcription *etc.* L'évolution temporelle des concentrations x_i est alors donnée par un système d'équations différentielles de la forme (1.1)

$$\frac{dx_i(t)}{dt} = f_i(x(t)) - \lambda_i x_i(t), 1 \leq i \leq n, \quad (1.1)$$

où $-\lambda_i x_i$ est un terme de dégradation linéaire et $f_i(x)$ regroupe tous les termes de production et de consommation de la variable x_i . Le champ de vecteur $F := (f_1, \dots, f_n)$ peut être constitué de termes de synthèse, dégradation, liaison, dissociation, phosphorylation, déphosphorylation, *etc.* Ces termes recouvrent l'ensemble des régulations biologiques connues, et peuvent être de nature variée. Les réactions de catalyse enzymatique par exemple, telles que la phosphorylation ou la déphosphorylation, sont classiquement représentées par des termes du type Michaelis-Menten (voir [62]). S'agissant des réactions de liaison ou de dissociation, elles sont souvent une simple traduction du principe de loi d'action de masse. Parfois, il est aussi commode d'utiliser des termes moins mécanistiques, comme les fonctions de Hill par exemple (ou d'autres sigmoïdes), pour représenter des régulations plus complexes ou moins connues. On peut se référer à l'article de revue [80] qui décrit de façon très pédagogique les étapes de construction de ce type de modèle dans un cas pratique du réseau de régulation du cycle cellulaire.

On peut citer en guise d'exemple le modèle de Novàk et Tyson [66], qui regroupe une grande partie des régulations du cycle cellulaire citées plus haut dans un système de $n = 18$ équations différentielles. On peut citer également le modèle de Leloup et Golbeter [51] qui propose une représentation de la régulation de l'horloge circadienne de cellules mammifères en $n = 16$ équations différentielles. Ces deux exemples illustrent bien le type de systèmes d'EDO (1.1) utilisés pour modéliser les réseaux de régulations cellulaires. Ce sont généralement des systèmes de grande dimension (dizaines de variables), non linéaires, et fortement structurés autour d'un graphe d'interaction représentant les différentes activations et inhibitions entre les variables. De manière générale, la plupart des termes utilisés pour construire ce type d'EDO sont suffisamment réguliers pour assurer l'existence et l'unicité d'une solution $x(t) = (x_1(t), \dots, x_n(t))$, dépendant continûment d'une condition initiale $x^0 = x(t = 0)$. Ces solutions sont généralement calculées numériquement à l'aide d'algorithmes classiques, pour être ensuite comparées avec le phénomène biologique modélisé.

Parfois, l'équation (1.1) prend en compte un signal extérieur $u(t)$ comme la concen-

tration d'une substance par exemple. Dans ce cas, (1.1) s'écrit sous la forme :

$$\frac{dx_i(t)}{dt} = f_i(x(t), u(t)) - \lambda_i x_i(t), 1 \leq i \leq n. \quad (1.2)$$

Ceci permet notamment de reproduire le fait qu'un processus soit déclenché par un stimulus extérieur. Ainsi, dans un modèle étudiant le cycle cellulaire comme [66], $u(t)$ peut représenter la concentration d'un facteur de croissance dont le rôle est de déclencher le cycle. Pour un modèle de l'horloge circadienne comme [51] $u(t)$ peut représenter un signal lumineux, afin de moduler la période ou la phase des oscillations.

Dans la suite, nous proposons une liste de modèles basés sur des EDO du cycle cellulaire et de l'horloge circadienne, pris dans la littérature. Notre but n'est pas d'être exhaustif, mais plutôt de montrer plusieurs types d'utilisation pratique de ces modèles quantitatifs pour représenter la dynamique des oscillateurs cellulaires, leurs intérêts et leurs limitations.

Dans [32], Gérard et Golbeter proposent un modèle de $n = 39$ équations différentielles étudiant le réseau de régulation du cycle de cellules mammifères. La structure du réseau est organisée en quatre modules, centrés autour des complexes cycline/cdk cités plus haut (voir Figure 1.4). Dans leur système, Gérard et Golbeter prennent en compte l'effet des facteurs de croissance dans le déclenchement du cycle avec l'introduction d'une variable d'entrée GF . Les oscillations du modèle, obtenues par simulation numérique, montrent bien que le système cycline/Cdk aboutit à son auto-organisation temporelle, conduisant à l'activation répétitive et séquentielle des modules et la progression ordonnée le long des phases du cycle. De plus, ils ont montré que les oscillations ne surviennent que lorsque la concentration de GF dépasse un seuil (voir Figure 1.6). En dessous de ce seuil, les simulations convergent vers un point d'équilibre stable, identifiable à la phase de quiescence G0. Ce résultat reproduit la présence d'un point de restriction dans le modèle, avec déclenchement irréversible du cycle uniquement lorsque les facteurs de croissance sont suffisants, conformément aux résultats expérimentaux de Zetterberg et Larson [98]. Les simulations numériques d'un tel modèle permettent donc d'avoir une bonne idée du comportement dynamique du modèle et ainsi de le comparer à des données expérimentales quantitatives. Néanmoins, la manipulation mathématique d'un tel système non linéaire de dimension 39 reste délicate. Dans [33], les auteurs réduisent leur modèle en un "squelette" de 5 équations qui capture encore les oscillations et reproduit le point de restriction. Ce modèle réduit est plus maniable, mais sa construction n'est pas automatique et nécessite plusieurs choix de modélisation.

D'autres modèles EDO de la littérature se sont attachés à reproduire des comportements caractéristiques du cycle cellulaire [66, 95, 78], notamment les oscillations et les points de restriction. Dans [66], les auteurs proposent même une estimation (numérique) de la période des oscillations et situent le point de restriction environ 3 heures après la mitose (donc en G1), ce qui est quantitativement conforme à l'expérience de Zetterberg et Larson [98]. Ceci montre qu'en plus d'observer l'évolution temporelle des concentrations, ces modèles offrent la possibilité d'évaluer des propriétés quantitatives, comme la période des oscillations, cruciales pour comparer les simulations du modèle avec les caractéristiques biologiques obtenues expérimentalement.

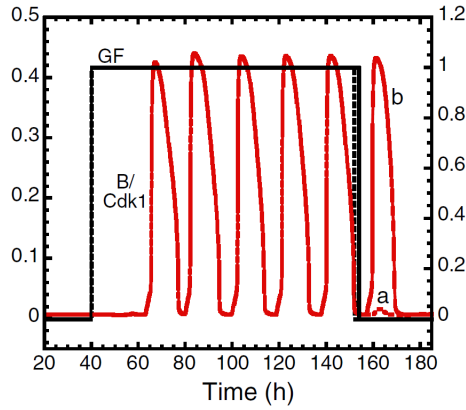


FIGURE 1.6 – Illustration des oscillations temporelles du modèle de Gérard et Golbeter, 2009 : réponse de la variable CycB/Cdk1 à l'introduction de facteur de croissance GF (figure extraite de [32]).

Dans [51], Leloup et Golbeter proposent un modèle différentiel de l'horloge circadienne de 16 équations reproduisant des oscillations dans une obscurité constante. En plus d'être soutenues, les oscillations de ce modèle présentent une opposition de phase entre les complexes Clock:Bmal1 et Per:Cry , caractéristique de l'horloge des cellules mammifères. De plus, les auteurs montrent que les oscillations peuvent être entraînées par la lumière selon des cycles lumière/obscurité, pour ensuite étudier la sensibilité de la phase des oscillations par rapport à certains paramètres. Cette analyse de sensibilité conclut à la génération de déphasages similaires à ceux observés dans les syndromes d'avance ou de retard de phases. Un an après la publication de ce modèle, les auteurs proposent une extension dans [50]. En utilisant différents ensembles de valeurs de paramètres, ils montrent que la survenue et la période des oscillations sont généralement plus sensibles aux paramètres liés à la synthèse ou à la dégradation de l'ARNm Bmal1 que de la protéine Bmal1 .

Dans [4], un système d'EDO de $n = 8$ équations est proposé pour modéliser le réseau de régulation de l'horloge circadienne des cellules mammifères et la simulation d'une trajectoire est reproduite dans la figure 1.7. Cette simulation montre l'oscillation de toutes les variables du modèle sur une période d'environ 20 heures et reproduit également l'opposition de phase entre les complexes Clock:Bmal1 et Per:Cry . Nous reviendrons plus en détail sur ce modèle dans le chapitre 3, dans l'optique d'en proposer une version qualitative.

A travers ces différents exemples, nous avons vu plusieurs avantages des modèles basés sur des équations différentielles. Outre la production de courbes temporelles, qui peuvent être plus ou moins directement mises en relation avec des mesures expérimentales, les simulations numériques de tels modèles permettent également d'estimer des propriétés quantitatives clés, comme la période des oscillations par exemple. Néanmoins, la dimension de tels systèmes étant souvent élevée (de l'ordre de la dizaine de variables), les analyses mathématiques peuvent être rapidement limitées. Dans le cadre de systèmes oscillants, en particulier, la démonstration de l'existence d'un cycle limite stable peut rapidement devenir délicate. Souvent les auteurs sont amenés à réduire le système de façon

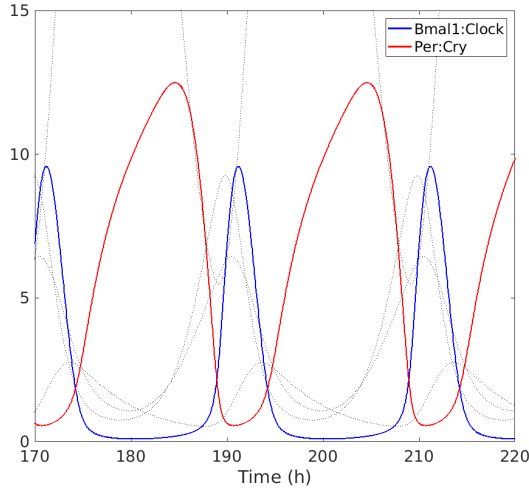


FIGURE 1.7 – Simulation temporelle du modèle d’Almeida *et al* 2020, montrant l’opposition de phase entre les complexes Clock:Bmal1 et Per:Cry. Les valeurs des paramètres sont tirées de [4], et la simulation numérique est faite avec Matlab (MathWorks, Inc).

drastique, comme dans [33] ou [5] pour utiliser des outils théoriques comme le théorème de Poincaré-Bendixon (voir [17]), par exemple.

Une autre difficulté de ces modèles réside dans leur construction. D’une part, comme nous l’avons vu plus haut les termes mathématiques choisis pour représenter les différentes régulations peuvent être relativement variés, et dépendent en général du niveau de connaissances sur l’ensemble des processus entrant en jeu. Ainsi, au fur et à mesure de l’introduction de nouvelles variables dans le modèle, les choix de modélisation deviennent plus nombreux, et nécessitent un ensemble de plus en plus important de connaissances biologiques. D’autre part, les équations différentielles construites dépendent en général de nombreux paramètres décrivant les cinétiques intracellulaires, dont l’estimation nécessite de nombreuses données expérimentales, pas toujours disponibles. Pour ces raisons, on leur préfère parfois des modèles plus qualitatifs, qui nécessitent moins de données quantitatives comme les systèmes affines par morceaux [25] ou les réseaux Booléens. Dans la suite, nous présentons les réseaux Booléens asynchrones, qui sont particulièrement bien adaptés pour analyser la dynamique de réseaux de régulation cellulaire. Pour les phénomènes oscillatoires qui nous intéressent dans cette thèse, leur nature qualitative est importante car elle permet notamment d’aborder de façon directe la description des cycles en terme de successions de phases discrètes.

1.3 Modélisation qualitative des oscillateurs biologiques par des réseaux Booléens

Dans cette partie, nous présentons une approche de modélisation basée sur les réseaux Booléens asynchrones. A la différence des EDOs, les réseaux Booléens se focalisent sur la présence ou l’absence des régulateurs clefs, simplifiant les régulations à des expressions Booléennes. Par exemple, la cycline E est peu présente en début de cycle (représentée

par $CycE = 0$), alors que sa concentration est élevée à la fin de la phase G1 (représentée par $CycE = 1$). C'est cette forte concentration qui est l'élément déclencheur de la phase S. C'est donc l'évènement $CycE = 1$ qui servira de signal, oubliant ainsi la valeur exacte de la concentration.

La modélisation Booléenne a été introduite dans l'étude des réseaux biologiques par Kauffman en 1969 [44] puis affinée par Thomas en 1973 [85]. Le premier proposait l'utilisation de réseaux synchrones tandis que le second utilisait une approche asynchrone (définie ci-après) plus adaptée aux réseaux biologiques. Depuis, de nombreux modèles Booléens sont apparus, représentant toute une variété de systèmes biologiques. La régulation du cycle cellulaire, en particulier, a souvent été modélisée par des réseaux Booléens : par exemple [24, 52, 68] en synchrone, ou [26, 89, 27] en asynchrone. Les modèles Booléens de l'horloge circadienne sont moins nombreux, on peut néanmoins citer [2, 19, 14]. D'autres réseaux de régulations cellulaires, notamment la prise de décision d'entrée en apoptose [12] ont également été modélisés par de tels systèmes. Ci-après, nous proposons une description rapide du concept de réseaux Booléens asynchrones, plus de détails se trouvent par exemple dans [1, 87, 37, 77].

1.3.1 Les réseaux Booléens asynchrones

Dans cette partie, nous considérons que les entités impliquées dans le réseau (protéines, ARN messagers, facteurs de transcriptions, *etc*) sont représentées par des variables Booléennes $x_i(t) \in \Omega = \{0, 1\}$. La variable $x_i = 1$ si sa concentration est supérieure à un certain seuil, ou si elle est sous forme active, par exemple. La variable $x_i = 0$ si sa concentration est inférieure à un certain seuil, ou si elle est sous une forme inactive. Le réseau de régulation entre les n variables se représente alors directement par un graphe orienté, appelé *graphe d'interaction*, tel que les sommets sont les entités et les arcs représentent les régulations (inhibitions ou activations). Les fonctions f_i de (1.1) sont maintenant remplacées par des fonctions Booléennes. Le plus souvent, ces fonctions s'expriment à l'aide des trois opérateurs logiques *et* $\wedge$, *ou* $\vee$ et *non* $\overline{(\cdot)}$. A chaque variable x_i est donc associée une règle d'activation $f_i : \Omega^n \rightarrow \Omega$. Cette règle d'activation f_i est construite en prenant en compte toutes les variables qui ont un effet direct sur x_i , c'est à dire les variables qui activent ou qui inhibent x_i . Ceci mène à une application Booléenne $F : \Omega^n \rightarrow \Omega^n$ qui encode la structure discrète du réseau.

L'état du réseau est donc un vecteur Booléen $x = (x_1, \dots, x_n)$, l'ensemble des états est Ω^n . Pour définir une dynamique, il s'agit à présent de préciser comment le système évolue d'un état $x(t)$ vers un état successeur $x(t + 1)$. En d'autres mots, il faut définir pour chaque état $x \in \Omega^n$ l'ensemble de ses successeurs $Succ(x)$. Pour cela, on choisit une stratégie de mise à jour des variables. Deux stratégies classiques existent : synchrone ou asynchrone.

Dans la stratégie synchrone, toutes les variables sont mises à jour simultanément. L'ensemble des successeurs de x est simplement :

$$Succ(x) = \{F(x)\}.$$

Ainsi, chaque état du système a un et un seul successeur. Cette stratégie suppose que toutes les variables ont la même vitesse. Cette hypothèse est bien sûr loin d'être réaliste en biologie, car les échelles de temps des processus de régulation peuvent être très différentes. On lui préférera donc la stratégie asynchrone, plus réaliste.

La stratégie asynchrone suppose que, dans chaque état $x \in \Omega^n$, une et une seule variable est mise à jour à la fois. Posons, pour $x \in \Omega^n$ et $i = 1, \dots, n$, $\tilde{x}^i := (x_1, \dots, \bar{x}_i, \dots, x_n)$. L'ensemble $Succ(x)$ est alors défini par :

$$Succ(x) = \begin{cases} \{\tilde{x}^i, f_i(x) \neq x_i\}, & \text{si } F(x) \neq x, \\ \{x\}, & \text{si } F(x) = x. \end{cases}$$

Selon cette définition, chaque état peut avoir plusieurs successeurs. Cette particularité rend la mise à jour asynchrone plus complexe. En effet, à partir de l'ensemble des successeurs $Succ$, on peut construire le graphe de transition $G = (V = \Omega^n, E)$ tel que $E = \{(x, y), x \in V, y \in Succ(x)\}$, qui est donc non déterministe. Le graphe de transition G est un graphe orienté de 2^n sommets, qui contient toute la dynamique du modèle.

Pour analyser le graphe de transition asynchrone, nous utilisons deux algorithmes classiques de la théorie des graphes : la décomposition en composantes fortement connexes (SCC, strongly connected components) et le tri topologique (voir [22] pour plus de détails). La décomposition du graphe de transition en SCCs permet d'obtenir un graphe acyclique. En appliquant le tri topologique, nous obtenons alors une représentation hiérarchique de la dynamique, en particulier nous isolons les SCCs terminales, dont le système ne peut s'échapper. Ces SCCs terminales sont les attracteurs du réseau.

Lorsqu'un attracteur est constitué d'un seul état, c'est un *point fixe* (remarquons que les points fixes sont indépendants de la stratégie de mise à jour utilisée). Les attracteurs qui contiennent plus d'un état seront dits dans la suite *complexes*. En synchrone, les attracteurs complexes sont nécessairement des cycles (car le graphe de transition synchrone est déterministe). Par conséquent, les attracteurs synchrones sont relativement faciles à comparer avec des oscillations biologiques. En revanche, en asynchrone, les attracteurs complexes peuvent être beaucoup plus compliqués à analyser, car ils contiennent beaucoup plus de transitions.

En guise d'exemple, nous considérons le modèle Booléen du cycle cellulaire proposé dans [26]. Ce modèle contient $n = 10$ variables centrales dans la régulation du cycle : *CycD*, *Rb*, *E2F*, *CycE*, *CycA*, *P27*, *Cdc20*, *Cdh1*, *UbcH10*, *CycB*. Les règles du réseau sont reproduites dans la Table 1.1. Dans ce modèle, la variable *CycD*, représentant le facteur de croissance est considérée comme une entrée. Lorsque *CycD* = 0, le graphe de transition admet un unique attracteur qui est le point fixe (0, 1, 0, 0, 0, 1, 0, 1, 0, 0) correspondant à la phase de quiescence G_0 . Lorsque *CycD* = 1, c'est à dire en présence de facteur de croissance, le graphe de transition admet également un unique attracteur contenant 112 états et 338 transitions (voir [26] pour plus de détails). La comparaison d'un tel attracteur complexe avec des connaissances biologiques n'est pas évidente. Dans le chapitre 2, nous proposons une méthode pour simplifier un tel attracteur et rendre la comparaison plus aisée.

1.3.2 Ajout de probabilités de transition dans le modèle Booléen

Les réseaux Booléens asynchrones sont un bon outil pour capturer la dynamique qualitative des réseaux de régulation cellulaires. Dans chaque état x , chaque variable susceptible de bouger donne un successeur potentiel. Cependant, aucune hypothèse n'est faite *a priori* concernant l'ordre dans lequel les variables sont mises à jour. Pour ajouter un tel ordre, nous pouvons associer à chaque successeur une probabilité. Ce faisant, on

Variable	Règle
CycD	$CycD;$
Rb	$(\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge \overline{CycD} \wedge \overline{CycB});$
E2F	$(\overline{Rb} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge \overline{Rb} \wedge \overline{CycB});$
CycE	$E2F \wedge \overline{Rb};$
CycA	$(E2F \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge \overline{(Cdh1 \wedge Ubc)}) \vee (CycA \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge \overline{(Cdh1 \wedge Ubc)});$
p27	$(\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge \overline{(CycE \wedge CycA)} \wedge \overline{CycB} \wedge \overline{CycD});$
Cdc20	$CycB;$
Cdh1	$(\overline{CycA} \wedge \overline{CycB}) \vee Cdc20 \vee (p27 \wedge \overline{CycB});$
Ubc	$\overline{Cdh1} \vee (Cdh1 \wedge Ubc \wedge (Cdc20 \vee CycA \vee CycB));$
CycB	$\overline{Cdc20} \wedge \overline{Cdh1};$

TABLE 1.1 – Règles d’activation des variables du modèle de Fauré *et al.*, [26]

construit alors un graphe de transition probabilisé. Ce graphe est en fait le graphe d’une chaîne de Markov de matrice de transition $P \in [0, 1]^{2^n \times 2^n}$ qui peut être définie par

$$P_{x,y} = \begin{cases} \frac{1}{|Succ(x)|}, & \text{si } y \in Succ(x) \\ 0, & \text{sinon.} \end{cases}$$

Ceci revient à supposer que pour chaque état, tous les successeurs sont équiprobables. En d’autres termes, les vitesses de tous les processus de régulation sont équivalentes. Si des informations, qualitatives, sont disponibles sur ces vitesses relatives, on peut facilement les inclure dans la chaîne de Markov, en associant à chaque variable x_i un poids $\rho_{x_i} > 0$. En asynchrone, chaque transition $x \longrightarrow y$ correspond à la mise à jour d’une variable. Ainsi, nous pouvons définir le poids de la transition :

$$\rho(x \longrightarrow y) := \rho_{x_i},$$

où x_i est la variable qui a bougé *i.e* $y = \tilde{x}^i$. La formule de la matrice de transition devient alors :

$$P_{x,y} = \begin{cases} \frac{\rho(x \longrightarrow y)}{\sum_{z \in Succ(x)} \rho(x \longrightarrow z)}, & \text{si } y \in Succ(x), \\ 0, & \text{sinon.} \end{cases} \quad (1.3)$$

Grâce à cette matrice, nous pouvons régler les probabilités pour prendre en compte les vitesses relatives des différentes variables du réseau.

En utilisant la formule (1.3), nous associons le même poids à une transition dans laquelle la variable x_i s’active (passe de 0 à 1) et à une transition dans laquelle x_i s’inactive (passe de 1 à 0). Ceci n’est pas forcément réaliste, car les processus de synthèse et de dégradation d’une entité peuvent être très différents. Ceci peut être très facilement pris en compte en associant à chaque variable, non pas un poids mais deux poids $\rho_{x_i}^+$ pour l’activation de x_i et $\rho_{x_i}^-$ pour sa dégradation. En modifiant la formule (1.3) en conséquence, nous produisons alors une chaîne de Markov plus réaliste.

L’introduction d’une chaîne de Markov dans un réseau Booléen asynchrone est intéressante car elle permet d’ajouter des informations biologiques quantitatives (ou semi-quantitatives) au modèle, notamment sur les vitesses relatives des processus de régulation

du réseau. D'un point de vue mathématique, un graphe de transition asynchrone et une chaîne de Markov (discrète) sont bien sûr des objets très proches, car reliés par la théorie des matrices à coefficients positifs, comme décrit dans [9].

1.3.3 Attrappeurs asynchrones complexes et oscillations cellulaires

Comme nous l'avons vu, les réseaux Booléens asynchrones sont un bon moyen de représenter la dynamique qualitative des réseaux de régulation cellulaires. En particulier, leur nature discrète est bien adaptée pour représenter des "signaux" biologiques, comme par exemple l'augmentation de la concentration d'une cycline avant la transition vers une prochaine phase. D'ailleurs, lorsqu'on lit des descriptions biologiques des oscillations du cycle cellulaire, on se rend compte que ces descriptions sont faites le plus souvent de façon qualitative, indépendamment des valeurs prises des concentrations. Ainsi, l'attrapeur d'un modèle Booléen est un bon moyen de capturer la dynamique qualitative de telles oscillations. Si l'identification de tels attrappeurs est relativement simple d'un point de vue théorique (SCC terminales d'un graphe), d'un point de vue pratique, il faut pouvoir comparer l'attrapeur avec le phénomène oscillatoire. Lorsque le graphe contient des dizaines voire des centaines d'états, cette comparaison devient rapidement délicate.

Dans cette thèse, nous proposons une méthode pour analyser de tels attrappeurs, afin de les comparer avec les phénotypes biologiques qu'ils modélisent. Nous nous intéressons particulièrement aux phénotypes oscillatoires comme le cycle cellulaire et l'horloge circadienne. Tirant profit de la nature discrète des réseaux Booléens, le premier travail consiste à découper l'attrapeur afin de retrouver les phases biologiques du cycle. Une fois ces phases identifiées, nous pouvons alors repérer les différentes transitions de phases dans l'attrapeur. Parfois, l'identification des phases est rendue aisée par la description biologique de l'oscillateur, comme c'est le cas pour le cycle cellulaire. Dans d'autres cas, comme pour l'horloge circadienne, l'établissement des phases peut être moins évident et nécessiter un travail de modélisation en amont.

La méthode proposée consiste en la construction d'un graphe qui résume l'attrapeur. Un tel graphe nous permet de valider ou non un modèle, en reproduisant les transitions de phases telles qu'elles sont connues au niveau biologique. Ce graphe est particulièrement utile pour analyser finement les dynamiques contenues dans l'attrapeur. Par exemple, la modélisation d'une mutation dans le réseau peut entraîner la conservation d'un attrapeur complexe, mais l'analyse des deux graphes résumés peut révéler la disparition de certaines transitions de phases dans le modèle mutant. Dans ce cas, nous pouvons conclure que même si la mutation n'a pas fait disparaître les oscillations complètement, elle a impacté la dynamique globale de l'attrapeur. Au delà de son utilisation directe pour la validation, l'outil *graphe résumé* ouvre la voie vers une analyse plus fine des dynamiques oscillatoires, en permettant notamment au ssmodélisateur d'avoir un accès direct aux transitions de phases.

Dans le chapitre 2, nous donnons une définition précise du concept de graphe résumé et nous l'utilisons pour analyser le modèle du cycle cellulaire de [26]. Nous verrons notamment que ce modèle capture bien la succession attendue des phases du cycle. En introduisant en plus la notion de chaîne de Markov, comme plus haut, nous utilisons également le graphe résumé pour estimer le temps passé dans les phases du cycle. Nous proposons ensuite dans le chapitre 3 un modèle Booléen de l'horloge circadienne, et nous

appliquons notre méthode pour identifier des phases du cycle. Enfin, nous approfondissons notre étude dans le chapitre 4, en analysant de façon plus fine les deux graphes résumés des chapitres précédents. En particulier, nous montrons comment cet outil permet d'ajuster localement la dynamique asymptotique de réseaux Booléens asynchrones fournissant ainsi un outil puissant dans la modélisation de réseaux de régulation biologiques.

Chapitre 2

Analyse d'attracteurs Booléens complexes par un graphe résumé : application à un modèle du cycle cellulaire

Dans ce chapitre, nous introduisons le concept de *graphe résumé* d'un attracteur Booléen asynchrone, et nous montrons son utilité pour analyser des réseaux de régulation cellulaires présentant des oscillations. Du point de vue de la théorie des graphes, il est basé sur le concept bien connu de graphe quotient. Appliqué aux modèles Booléens, il constitue un outil intéressant pour pouvoir comparer un attracteur complexe, contenant par exemple des centaines d'états et de transitions, avec un phénomène biologique, notamment un phénomène oscillant.

Nous appliquons cette démarche pour analyser l'attracteur asynchrone du modèle de [26]. Ce réseau de 10 variables modélise les principales régulations impliquées dans le cycle de division des cellules mammifères, décrit dans le chapitre 1. En présence de facteur de croissance (cycline D dans le modèle), ce réseau admet un unique attracteur de 112 états et de nombreux circuits imbriqués. En construisant un graphe résumé de cet attracteur, nous montrons notamment que la succession des phases du cycle G1-S-G2-M est bien respectée. Au-delà de l'outil de visualisation, permettant de repérer simplement des événements qualitatifs dans le graphe d'un attracteur complexe, le graphe résumé offre également l'opportunité d'analyser plus en détail ce type d'attracteur. Ainsi, nous pouvons réanalyser le comportement de certains mutants avec le graphe résumé, afin de repérer si les oscillations d'un mutant sont similaires au modèle initial (d'un point de vue qualitatif) ou si elles sont dégradées. En combinant le graphe résumé avec l'outil des chaînes de Markov évoqué plus haut, nous pouvons également avoir accès à des propriétés plus quantitatives, comme la proportion du temps passé dans chaque phase par exemple.

Après avoir donné les définitions et propriétés principales (partie 2.1), nous décrivons dans la partie 2.2 comment construire le graphe résumé d'un attracteur complexe, en prenant en exemple le modèle de [26]. Nous analysons par la suite ce graphe, en prenant soin de repérer ses caractéristiques intéressantes du point de vue de la biologie. Enfin, dans la partie 2.3 nous décrivons comment construire une chaîne de Markov sur un graphe résumé. Nous utilisons cet outil pour étudier l'impact des vitesses relatives des différents

processus sur les proportions de temps passé dans les phases du cycle.

2.1 Définition du graphe résumé d'un attracteur

On considère $A = (V, E)$ le graphe d'un attracteur asynchrone avec $V \subseteq \{0, 1\}^n$. Soit $k \in \mathbb{N} \setminus \{0\}$ et $\mathcal{P} = \{V_1, V_2, \dots, V_k\}$ une k -partition de V , c'est à dire une collection de k sous ensembles de V vérifiant les conditions (C1), (C2) et (C3) suivantes :

- (C1) $\forall i = 1, \dots, k : V_i \neq \emptyset$,
- (C2) $\forall i, j = 1, \dots, k, i \neq j : V_i \cap V_j = \emptyset$,
- (C3) $\cup_{i=1}^k V_i = V$.

Une collection de sous ensembles de V vérifiant seulement (C1) et (C2) est dite (k) -semi-partition de V .

Dans la suite, nous allons définir, dans un premier temps, le graphe résumé de A lorsque $\mathcal{P}$ est une partition de V . Puis, dans un second temps, nous définissons le graphe résumé de A dans le cas où $\mathcal{P}$ est une semi-partition pour généraliser la première définition. Afin d'illustrer ces définitions, nous introduisons l'exemple suivant, qui est un réseau Booléen de dimension 4 donné par les règles :

$$\begin{cases} f_1(x_1, x_2, x_3, x_4) &= (\overline{x_3} \wedge \overline{x_4}) \vee \overline{x_2}, \\ f_2(x_1, x_2, x_3, x_4) &= \overline{x_4} \vee x_3 \vee x_1, \\ f_3(x_1, x_2, x_3, x_4) &= x_2 \wedge \overline{x_4}, \\ f_4(x_1, x_2, x_3, x_4) &= x_3. \end{cases} \quad (2.1)$$

Ce réseau admet un unique attracteur de 12 états, représenté dans la figure 2.1.

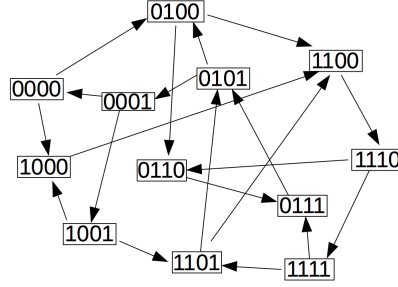


FIGURE 2.1 – Attracteur de 12 états du réseau (2.1).

2.1.1 Graphe résumé d'un attracteur sur une partition

Définition 1 (Graphe résumé) Soit $A = (V, E)$ un attracteur asynchrone et $\mathcal{P} = \{V_1, \dots, V_k\}$ une k -partition de V . On appelle graphe résumé de A associé à $\mathcal{P}$, le graphe $A(\mathcal{P}) = (V(\mathcal{P}), E(\mathcal{P}))$, où $V(\mathcal{P})$ et $E(\mathcal{P})$ sont respectivement l'ensemble des sommets et des arcs, définis par :

- $V(\mathcal{P}) = \{V_1, \dots, V_k\}$,
- $E(\mathcal{P}) = \{(V_i, V_j) \in (V(\mathcal{P}))^2 : i \neq j, \exists x \in V_i, y \in V_j : (x, y) \in E\}$.

A chaque couple $(i, j) \in \{1, \dots, k\}$, est associé un ensemble de transitions dans l'attracteur : $T_{ij} := \{(x, y) \in E : x \in V_i, y \in V_j\}$. Ainsi, il y aura une flèche $V_i \rightarrow V_j$ dans le graphe résumé $A(\mathcal{P})$ si et seulement si l'ensemble de transitions T_{ij} est non vide. Par convention les boucles $V_i \rightarrow V_i$ seront ignorées.

Pour illustrer la définition 1 sur le graphe de la figure 2.1, nous considérons la partition $\mathcal{P}_1 = \{V_1, V_2, V_3, V_4\}$ suivante

$$\begin{cases} V_1 = \{0000, 1000, 0100, 1100\}, \\ V_2 = \{1110, 0110, 1111\}, \\ V_3 = \{1101, 0111\}, \\ V_4 = \{0101, 0001, 1001\}. \end{cases}$$

Le graphe résumé $A(\mathcal{P}_1)$ obtenu avec la partition $\mathcal{P}_1$ est donné par la figure 2.2.

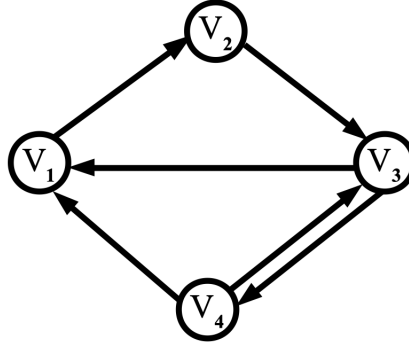


FIGURE 2.2 – Graphe résumé de l'attracteur du réseau (2.1).

Notons que, dans la définition 1, nous utilisons le symbole V_i pour désigner à la fois un sommet du graphe résumé $A(\mathcal{P})$ et également un ensemble d'états de l'attracteur. Cet abus de notation nous permettra de décrire plus facilement les chemins dans le graphe résumé $A(\mathcal{P})$. Le graphe résumé tel que nous l'avons défini sur une partition correspond, en théorie des graphes, à la notion de graphe quotient [30]. D'ailleurs, cette définition a été déjà introduite dans [86] dans le cadre plus général des graphes (non orientés) de base de données de grandes tailles.

L'utilisation de cette notion de graphe résumé dans le cadre des réseaux Booléens asynchrones, nous permet de capturer la dynamique globale des attracteurs complexes. En effet, le graphe de la figure 2.1, est représenté beaucoup plus simplement dans la figure 2.2. A supposer que chaque V_i a un sens biologique et représente une étape clef dans la dynamique du processus que l'attracteur modélise qualitativement, le graphe résumé $A(\mathcal{P}_1)$ permet une visualisation plus directe du déroulement du processus. Ainsi, on peut voir une orientation globale du processus selon le cycle $V_1 \rightarrow V_2 \rightarrow V_3 \rightarrow V_4 \rightarrow V_1$ mais avec deux arcs retour $V_4 \rightarrow V_3$ et $V_3 \rightarrow V_1$.

Bien entendu, le graphe résumé d'un attracteur dépend fortement du choix de la partition. En effet, en considérant la partition $\mathcal{P}_2 = \{V'_1, V'_2, V'_3, V'_4\}$ définie par :

$$\begin{cases} V'_1 = \{0000, 0100\}, \\ V'_2 = \{1110, 1100, 0110, 1000, 0111\}, \\ V'_3 = \{1101, 1111, 0001\}, \\ V'_4 = \{1001, 0101\}, \end{cases}$$

nous obtenons le graphe résumé $A(\mathcal{P}_2)$ de la figure 2.3 ,fortement différent du graphe résumé $A(\mathcal{P}_1)$.

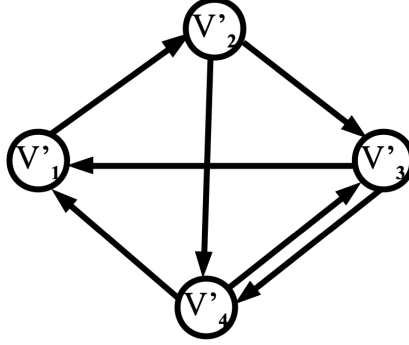


FIGURE 2.3 – Graphe résumé de l’attracteur du réseau (2.1) sur la partition $\mathcal{P}_2$.

Cette sensibilité du graphe résumé par rapport à la partition nous permet de nous rendre compte que le choix de la partition est crucial. Par ailleurs, le nombre de k –partitions d’un ensemble de p éléments correspond au nombre de Stirling de second espèce σ_p^k (voir [20]) et il explose très vite. Par exemple, avec le graphe de l’attracteur de la figure 2.1, le nombre de partitions des 12 états en 4 sous-ensembles est $\sigma_{12}^4 = 611501$. Cela reste calculable et tous les graphes résumés associés peuvent être obtenus. Par contre, vu l’explosion de σ_n^k , l’idée consistant à calculer exhaustivement toutes les partitions possibles devient rapidement irréalisable. Dans le contexte de cette thèse, plutôt que d’utiliser une approche purement mathématique, nous tirons profit du fait que les graphes que nous manipulons sont sensés capturer des phénotypes biologiques pour sélectionner une partition. Par exemple, dans le cas où le graphe de l’attracteur est sensé capturer les oscillations du cycle cellulaire, nous cherchons à représenter chacune des phases du cycle par un ensemble d’états. Ainsi, nous allons chercher, de façon générale, à construire des partitions en classant les états sur la base d’informations relatives à ces phénomènes. Ce faisant, certains états peuvent être difficiles à classer car ils ne correspondent pas à des états biologiques acceptables. Dans ce cas, on se retrouve avec une semi-partition au lieu d’une partition. Dans la section 2.1.2, nous donnons la définition du graphe résumé sur une semi-partition qui constitue en une généralisation de la définition 1.

2.1.2 Graphe résumé d’un attracteur sur une semi-partition

La représentation d’un graphe résumé sur une semi-partition est d’autant plus pertinente lorsqu’on étudie l’attracteur d’un modèle biologique comme dans notre cas. Pour construire une partition, on peut choisir de répartir les états en utilisant la nature de leur

configuration comme critère. Cependant, les connaissances biologiques sur le système modélisé ne suffisent généralement pas pour classer tous les états de l'attracteur. Les états non classés, bien que n'apparaissant pas explicitement dans le graphe résumé, relient certaines étapes clefs de manière indirecte en des trajectoires qu'on aimerait visualiser pour comprendre leur nature.

Avant de construire un graphe résumé sur une semi-partition, nous donnons la définition d'un chemin direct dans le graphe d'un attracteur.

Définition 2 (Chemin direct) Soit $A = (V, E)$ un attracteur asynchrone et $\mathcal{S} = \{V_1, \dots, V_k\}$ une k -semi-partition de V . Il existe un chemin direct de V_i vers V_j , $i \neq j$, s'il existe un chemin dans A d'un $x \in V_i$ vers un $y \in V_j$ qui ne passe que par V_i , V_j et les états non classés dans $\mathcal{S}$.

Dans la suite, nous noterons $V_i \supseteq V_j$ s'il existe un chemin direct de V_i vers V_j . Nous donnons la définition 3 qui nous permet d'introduire un graphe résumé sur une semi-partition. La définition 3 se présente aussi comme une généralisation de la définition 1 puisqu'une partition est aussi une semi-partition.

Définition 3 (Généralisation de la définition 1) Soit $A = (V, E)$ un attracteur asynchrone et $\mathcal{S} = \{V_1, \dots, V_k\}$ une k -semi-partition de V . On appelle graphe résumé de A associé à $\mathcal{S}$, le graphe $A(\mathcal{S}) = (V(\mathcal{S}), E(\mathcal{S}))$, où $V(\mathcal{S})$ et $E(\mathcal{S})$ sont respectivement l'ensemble des sommets et des arcs, définis par :

- $V(\mathcal{S}) = \{V_1, V_2, \dots, V_k\}$,
- $E(\mathcal{S}) = \{(V_i, V_j) \in (V(\mathcal{S}))^2 : i \neq j, V_i \supseteq V_j\}$.

A titre d'illustration, nous considérons la semi-partition

$$\mathcal{S} = \begin{cases} V_1 = \{0000, 1101, 0111\}, \\ V_2 = \{0100, 0101\}, \\ V_3 = \{1100, 1110, 0110\}, \\ V_4 = \{1001, 1111\} \end{cases}$$

des états de l'attracteur donné dans la figure 2.1. Dans $\mathcal{S}$, 10 états sur 12 sont classés et les deux états $\{1000, 0001\}$ sont non classés. La semi-partition $\mathcal{S}$ nous donne le graphe résumé $A(\mathcal{S})$ de la figure 2.4.

Dans $A(\mathcal{S})$, les flèches en trait plein représentent les transitions entre les V_i , comme dans les graphes résumés $A(\mathcal{P}_1)$ et $A(\mathcal{P}_2)$ montrés dans la section 2.1.1. En plus, nous visualisons maintenant d'autres chemins, passant par les états non classés, par des flèches pointillées. Avec $A(\mathcal{S})$, nous obtenons un graphe résumé qui permet de voir une structure globale de l'attracteur même si tous les états n'ont pas été classés dans les sous ensembles. De façon générale, l'intérêt de cette représentation est de visualiser l'attracteur malgré qu'il y ait des états non pris en compte par le critère de classement qu'on s'est donné.

Notons qu'on peut aussi compléter la semi-partition $\mathcal{S}$ en la partition

$$\mathcal{P}(\mathcal{S}) = \{V_1, V_2, \dots, V_k, U\}$$

dans laquelle les états non classés $U := V \setminus \left(\bigcup_{i=1}^k V_i\right)$ apparaissent explicitement. Nous pouvons alors définir le graphe résumé $A(\mathcal{P}(\mathcal{S}))$ en utilisant la définition 1. Ce graphe est

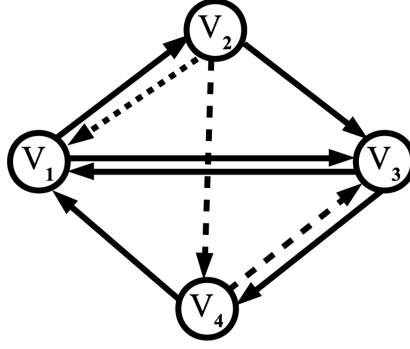


FIGURE 2.4 – Graphe résumé du réseau (2.1) sur la semi-partition $\mathcal{S}$. Les flèches en trait plein sont les chemins directs qui contournent les états non classés. Les flèches en pointillées sont les chemins directs qui passent par des états non classés $\{1000, 0001\}$.

représenté dans la figure 2.5. Dans le graphe $A(\mathcal{S})$, les états non classés sont donc cachés, et uniquement représentés par les “court-circuits” qu’ils génèrent éventuellement entre les V_i . Dans le graphe $A(\mathcal{P}(\mathcal{S}))$, en revanche, les états non-classés apparaissent explicitement.

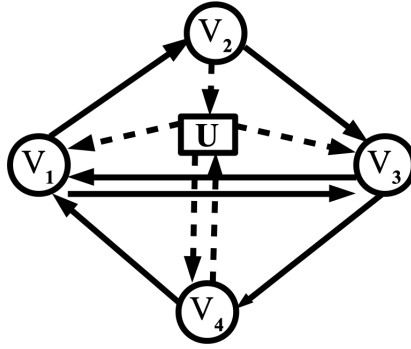


FIGURE 2.5 – Graphe résumé $A(\mathcal{P}(\mathcal{S}))$

2.1.3 Propriété centrale du graphe résumé d’un attracteur

Nous démontrons le théorème 1 ci après, établissant un lien important entre un attracteur et son graphe résumé. Ce théorème sera notamment utilisé dans la section 2.3 lorsque nous associons un graphe résumé à une chaîne de Markov.

Théorème 1 Soit $A = (V, E)$ un attracteur et $\mathcal{P} = \{V_1, V_2, \dots, V_k\}$ une k -partition de V . Le graphe résumé $A(\mathcal{P})$ est fortement connexe.

Preuve. A étant un attracteur, il est donc fortement connexe. Montrons que $A(\mathcal{P})$ l’est aussi.

Soit $V_i, V_j \in \mathcal{P}$ tels que $i \neq j$. Le principe de la preuve est de construire un chemin de V_i vers V_j dans le graphe résumé.

On a, par définition de $\mathcal{P} : V_i \neq \emptyset$ et $V_j \neq \emptyset$. Soit $x \in V_i$ et $y \in V_j$. Puisque A est fortement connexe, il existe un chemin de x vers y dans A . Donc il existe un entier naturel $r > 0$ et $x_0, x_1, \dots, x_r \in V$ tels que $x = x_0 \longrightarrow_A x_1 \longrightarrow \dots \longrightarrow_A x_{r-1} \longrightarrow_A x_r = y$. On peut considérer que ce chemin est élémentaire, c'est à dire aucun des sommets $x_0, \dots, x_r$ ne se répète. Puisque $i \neq j$, il existe $d_1 \leq r$ tel que $x_0, \dots, x_{d_1-1} \in V_i$ et $x_{d_1} \notin V_i$. Si $x_{d_1} \in V_j$, alors selon la définition 1 il existe une arête $V_i \rightarrow V_j$ ce qui termine la preuve. Sinon, soit $V_{i_1} \ni x_{d_1}$. Nous considérons alors le chemin $x_{d_1} \rightarrow \dots \rightarrow y$ et nous posons $d_2 \leq r$ tel que $x_{d_1}, \dots, x_{d_2-1} \in V_{i_1}$ et $x_{d_2} \notin V_{i_1}$. Puisque $y \in V_j$, en réitérant le processus précédent nous construisons progressivement un chemin $V_i \rightarrow V_{i_1} \rightarrow \dots \rightarrow V_j$ dans le graphe résumé, ce qui termine la preuve. ■

Remarque 1 Jusqu'ici, nous avons défini le graphe résumé d'un attracteur. La définition d'un graphe résumé peut bien sûr s'étendre sur n'importe quel type de graphe. Le théorème 1 peut alors être étendu : le graphe résumé d'un graphe fortement connexe est fortement connexe. En revanche, la réciproque est fausse.

2.2 Utilisation du graphe résumé pour l'analyse d'un modèle du cycle cellulaire

Pour illustrer les définitions précédentes sur l'attracteur complexe de [26], il nous faut d'abord chercher une partition pertinente. Cette partition doit nous permettre de comparer l'attracteur avec les propriétés connues du cycle cellulaire. La recherche d'une partition des états de l'attracteur constitue une étape stratégique dans la construction d'un graphe résumé. Comme évoqué plus haut, l'ensemble des partitions est trop grand pour une recherche exhaustive. Plutôt, nous allons utiliser des informations biologiques pour construire cette partition, en identifiant notamment des zones de l'attracteur associables à des étapes clefs de la dynamique du modèle. La nature de ces étapes dépend bien sûr des propriétés biologiques que l'on cherche à vérifier dans l'attracteur. Dans le cas du cycle cellulaire, ces étapes peuvent correspondre à l'interphase et à la mitose, ou plus précisément aux différentes phases $G1, S, G2, M$ du cycle. Dans la suite, nous avons choisi de répartir les 112 états de l'attracteur en six étapes clefs.

2.2.1 Construction d'une première partition des états de l'attracteur

Au premier abord, il serait naturel de choisir $k = 4$, de sorte que chacun des sous-ensembles soit représentatif d'une phase. Mais, compte tenu des variables du modèle, nous préférons prendre $k = 6$ groupes, découpés ainsi :

- $G1$: début de la phase G1.

Nous considérons le début de la phase G1 comme la première étape du cycle cellulaire. Nous la séparons de la fin de la phase car elles ne sont pas régulées de la même manière. En effet, d'après [21, 16], le début de la phase G1 est marqué par une activation de la cycline D. La cycline D libère le facteur de transcription $E2F$ en phosphorylant les protéines de la famille Rb (protéine du rétinoblastome). Dans le modèle [26], la cycline D est supposée constante et elle inhibe également

P27 qui est un inhibiteur des cyclines au même titre que *Rb* [70]. Ainsi, le facteur de transcription *E2F*, une fois actif, favorise la synthèse de la cycline E pour une progression de la phase G1 [79, 29]. Dans cette étape, nous regroupons les états de l'attracteur dans les quels $E2F = 1$ et *CycE* peut prendre les valeurs 0 ou 1.

- *G1/S* : fin de la phase G1 et entrée en phase S.
L'étape *G1/S* correspond à la fin de la phase G1 et au début de la phase S. C'est dans la transition *G1/S* que la cycline E admet une forte activité [72]. En plus de la cycline E, la cycline A est aussi connue pour être active dans cette transition [21, 27]. Ainsi, nous regroupons les états tels que $CycE = CycA = 1$ et $E2F$ vaut 0 ou 1.
- *SG2* : fusion des deux phases S et G2.
L'étape *SG2* est associée à la fusion de la phase S et de la phase G2. La raison de cette fusion est liée au fait que, dans le modèle, ces deux phases sont régulées par la seule variable cycline A. Nous regroupons les états tels que $CycA = 1$ (et $E2F = CycE = 0$).
- *G2/M* : fin de la phase G2 et entrée en mitose.
Cette étape est la fin de la phase G2 et le début de la phase de mitose M. Comme pour *G1/S*, cette étape représente à la fois la fin de la phase G2 et le début de la phase M. La cycline A qui est active depuis la fin de la phase G1, est aussi active dans cette transition *G2/M* [21]. Il est intéressant de souligner que l'activation de la cycline A dans *G1/S* se fait à travers un complexe avec la *Cdk2* alors que dans *G2/M*, cette activation se fait à travers un complexe avec la *Cdk1* [23, 28]. Nous ne faisons pas la différence puisque le modèle ne considère, par hypothèse, que les cyclines et non leurs partenaires *Cdk*. De plus, la cycline B est active à partir de cette transition [65]. Ainsi, nous regroupons les états de l'attracteur tels que $CycA = CycB = 1$.
- *M* : phase de mitose (avant l'anaphase).
Cette étape correspond à la phase de mitose M de l'attracteur. C'est à dire l'étape de la mitose partant de la transition *G2/M* jusqu'à l'anaphase. Ce découpage nous permet de différencier les étapes de la mitose dans lesquelles la Cycline B (ou MPF, Mitosis Promoting Factor) est active et les étapes de sortie de mitose dans lesquelles la cycline B est inhibée. Ainsi, nous regroupons les états de l'attracteur tels que $CycB = 1$ (et $CycA = 0$).
- *M/G1* : fin de la phase M et entrée en G1.
Enfin, *M/G1* correspond à l'étape de sortie de mitose et d'entrée en phase G1. Comme la Cycline D est constante ($CycD = 1$), la sortie de mitose s'enchaîne directement avec l'entrée dans d'une nouvelle phase G1.

Maintenant, il s'agit de classer les 112 états dans les sous-ensembles *G1*, *G1/S*, *SG2*, *G2/M*, *M* et *M/G1*. Le classement d'un état se fait selon sa configuration. Dans le tableau 2.1, nous avons sur chaque ligne le niveau attendu des variables du modèle selon la description précédente. La construction de ce tableau se base sur les informations de la littérature sur la régulation du cycle cellulaire. Nous mettons le symbole (*) pour indiquer qu'une variable peut prendre la valeur 0 ou 1 dans la phase.

En remplaçant les (*) par 0 et 1 dans le tableau 2.1, nous obtenons un classement de 24 états sur les 112 que compte l'attracteur. Ce classement forme donc une semi-partition, avec un nombre d'états classés assez faible. Nous obtenons alors le graphe résumé de la

Variables	$G1$	$G1/S$	$SG2$	$G2/M$	M	$M/G1$
CycD	1	1	1	1	1	1
Rb	0	0	0	0	0	0
E2F	1	*	0	0	0	0
CycE	*	1	0	0	0	0
CycA	0	1	1	1	0	0
P27	0	0	0	0	0	0
Cdc20	0	0	0	0	*	*
Cdh1	1	*	0	0	0	1
UbcH10	*	*	*	*	*	*
CycB	0	0	0	1	1	0
Nombre	4	8	2	2	4	4

TABLE 2.1 – Une première semi-partition $\mathcal{S} = \{G1, G1/S, SG2, G2/M, M, M/G1\}$ de l'attracteur du modèle de Fauré. Le nombre d'états classés dans chaque sous-ensemble est donné dans la dernière ligne.

figure 2.6 dans lequel les flèches en pointillés représentent les chemins passant par les états non classés.

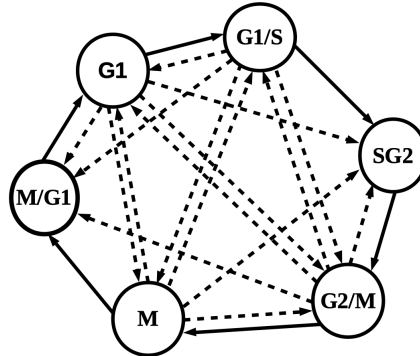


FIGURE 2.6 – Graphe résumé sur la semi-partition obtenue avec la table 2.1

Ce graphe résumé retrouve bien l'ordre attendu des phases biologiques. En revanche, il présente également de nombreux chemins parasites (court-circuits) entre ces phases. Ceci est certainement dû à l'insuffisance du nombre d'états classés. Ainsi, nous avons tenté de raffiner cette semi-partition dans la suite, en ajoutant d'autres états de l'attracteur qui n'avaient pas été pris en compte par ce premier classement.

2.2.2 Raffinement de la partition

Pour raffiner la partition précédente, nous faisons l'hypothèse de ne considérer que quatre variables particulières, E2F, CycE, CycA et CycB pour classer les états. En effet, ces quatre variables suffisent pour différencier les changements de phase (table 2.1). Nous construisons alors la Table 2.2.

$E2F$	$CycE$	$CycA$	$CycB$	<i>ensembles</i>
0	0	0	0	$M/G1$
0	0	0	1	M
0	0	1	0	$SG2$
0	0	1	1	$G2/M$
0	1	0	0	$G1$
0	1	0	1	$\emptyset$
0	1	1	0	$G1/S$
0	1	1	1	$\emptyset$
1	0	0	0	$G1$
1	0	0	1	$\emptyset$
1	0	1	0	$G1/S$
1	0	1	1	$\emptyset$
1	1	0	0	$G1$
1	1	0	1	$\emptyset$
1	1	1	0	$G1/S$
1	1	1	1	$\emptyset$

TABLE 2.2 – Raffinement de la semi-partition : table de vérité des six phases en fonction des quatre variables $E2F$, $CycE$, $CycA$ et $CycB$.

A partir de cette table, nous déduisons pour chaque phase une règle Booléenne qui constitue son critère d'appartenance :

$$\left\{ \begin{array}{lcl} G1 & = & (E2F \vee CycE) \wedge \overline{CycA} \wedge \overline{CycB} \\ G1/S & = & (E2F \vee CycE) \wedge CycA \wedge \overline{CycB} \\ SG2 & = & CycA \wedge \overline{E2F} \wedge \overline{CycE} \wedge \overline{CycB} \\ G2/M & = & CycA \wedge CycB \wedge \overline{CycE} \wedge \overline{E2F} \\ M & = & CycB \wedge \overline{E2F} \wedge \overline{CycE} \wedge \overline{CycA} \\ M/G1 & = & \overline{CycB} \wedge \overline{E2F} \wedge \overline{CycE} \wedge CycA \end{array} \right. \quad (2.2)$$

En utilisant ces règles, nous classons de façon automatique 76 sur les 112 états de l'attracteur. Ce qui est intéressant de noter est que le premier classement de 24 états est conservé et que celui-ci n'a fait qu'ajouter 52 autres états. Cette semi-partition de 76 états peut se compléter par les états non classés, donnant la partition $\mathcal{P} = \{G1, G1/S, SG2, G2/M, M, M/G1, U\}$ où U l'ensemble des 36 états non classés.

2.2.3 Analyse de l'attracteur grâce au graphe résumé

Avec la partition $\mathcal{P}$ définie dans la section 2.2.2, nous obtenons le graphe résumé donné dans la figure 2.7. La première chose que nous observons dans ce graphe est la présence du cycle $G1 \rightarrow G1/S \rightarrow SG2 \rightarrow G2/M \rightarrow M \rightarrow M/G1 \rightarrow G1$. Ce cycle correspond qualitativement au cycle cellulaire dans le sens où il respecte le sens et l'ordre de progression des phases.

Le graphe résumé montre aussi la présence de deux points de vérification (check points) qui sont connus dans le déroulement du cycle cellulaire (voir Murray, 1994[61], Barnum

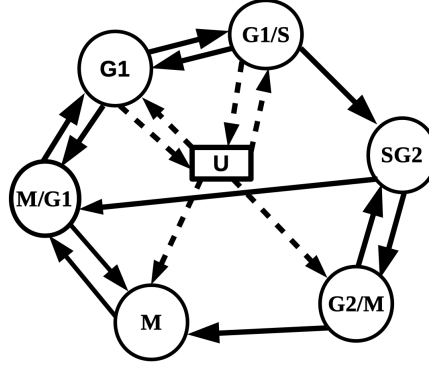


FIGURE 2.7 – Graphe résumé $A(\mathcal{P})$ de l'attracteur complexe du modèle dans [26]

& O'Connel 2014 [7]). Ces points correspondent à des moments du cycle où plusieurs conditions sont vérifiées, ils se situent avant la transition $G1/S$ et avant la transition $G2/M$. Dans le premier check point ($G1/S$), la cellule vérifie deux événements, à savoir si l'environnement est favorable à la duplication de l'ADN et si le rapport nucléocytoplasmique¹ (RNP) satisfait une certaine proportion. Tant que ces conditions ne sont pas favorables, la cellule ne passe pas à la phase de réplication. Dès que ce point est franchi, il se fait de manière irréversible (voir [10],[91]). Ceci se traduit dans le graphe résumé par l'irréversibilité de la transition $G1/S \rightarrow SG2$. Dans le deuxième check point ($G2/M$), la cellule, sortie de la phase de réplication, contient une quantité double de chromosome (cellule diploïde). Cette étape permet à la cellule de s'assurer que le matériel génétique est correctement et complètement répliqué avant d'entamer la mitose (voir Blagosklonny & Pardee, 2002 [10]). Dans ce cas, il s'agit pour la cellule de faire en sorte que la mitose ne soit pas déclenchée tant que la réplication n'est pas totalement achevée ou tant que les éventuelles lésions détectées dans l'ADN répliqué ne sont pas réparées. A nouveau, une fois cette transition effectuée, elle est irréversible. Cela est représenté dans le graphe résumé par l'irréversibilité la transition $G2/M \rightarrow M$. Un troisième point de contrôle intervient durant la mitose lors du passage de la métaphase à l'anaphase et elle correspond à la condition de sortie de la mitose. Ce point de contrôle n'est pas observable dans le graphe résumé.

Au delà de ces bonnes propriétés, le graphe résumé montre aussi des chemins irréalistes, qui court-circuitent le cycle. Afin de mieux visualiser ces chemins, nous représentons le graphe résumé sur la semi-partition obtenue en ne considérant que les 76 états classés (figure 2.8). Les flèches en pointillés et celle de $SG2 \rightarrow M/G1$ sont les court-circuits. Les flèches en pointillés sont des chemins qui traversent les états non classés alors que celles associées à la flèche de $SG2$ à $M/G1$ ne passent pas par les états non classés. Par exemple, le court-circuit $G1 \rightarrow M$ montre que la cellule peut quitter la phase $G1$ et entamer une phase de mitose sans avoir dupliqué son ADN. La flèche de $SG2$ à $M/G1$ montre un cycle qui saute les étapes de la mitose ; donc qui saute l'activation de la cycline B. La présence de ces types de chemins dans l'attracteur n'est pas surprenante d'autant plus qu'elle a été déjà signalée par les auteurs du modèle dans [26]. On voit bien ici l'intérêt du graphe résumé : il permet d'identifier de façon précise ces chemins. Mieux

1. le rapport volume du noyau cellulaire / volume du cytoplasme

encore, on peut remonter aux transitions à l'origine des chemins ainsi qu'aux variables du modèle qui en sont responsables. D'ailleurs, l'identification des variables responsables de ces transitions sera utilisée dans la section 2.3 et dans le chapitre 4.

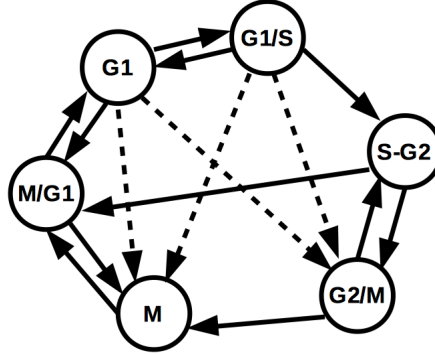


FIGURE 2.8 – Graphe résumé $A(\mathcal{S})$ avec $\mathcal{S} = \{G1, G1/S, SG2, G2/M, M, M/G1\}$

2.2.4 Validation de mutants à l'aide du graphe résumé

Nous proposons à présent d'utiliser l'outil graphe résumé pour étudier l'impact de quelques mutations sur le modèle. Nous nous intéressons d'abord à l'analyse des mutations non oscillantes, c'est à dire des mutations pour lesquelles le modèle admet des points fixes. Ensuite, nous utilisons le graphe résumé de la figure 2.7 pour comparer les oscillations de quelques mutants connus par rapport au modèle initial.

Dans [26], certaines mutations ont été déjà testées par les auteurs et comparées aux phénotypes biologiques correspondants. Certaines de ces mutations admettent un point fixe, indiquant une perte d'oscillations et l'arrêt du cycle, alors que d'autres conservent un attracteur complexe. Pour ces dernières, le graphe résumé est particulièrement utile pour étudier ces oscillations altérées.

Surexpressions conduisant à un arrêt du cycle : RB, P27, UBCH10. Dans la littérature, la surexpression de chacun des trois gènes Rb, P27 et Ubch10 entraîne un arrêt du cycle cellulaire [34, 53, 73]. Cet arrêt du cycle se traduit dans le modèle par une perte des oscillations lorsque $CycD=1$ et l'apparition d'un point fixe à la place de l'attracteur complexe. La partition $\mathcal{P}$ définie plus haut nous permet de situer ce point fixe par rapport aux étapes du cycle cellulaire dans l'attracteur, et ainsi de voir dans quelle étape le cycle s'est arrêté.

Rb=1. Le modèle admet le point fixe (1100000100). Dans ce point fixe, les quatre variables E2F, CycE, CycA et CycB sont toutes à 0. Ce point fixe se situe dans la phase M/G1, c'est à dire que le cycle cellulaire s'arrête à l'entrée de la phase G1. Cette observation est cohérente avec le résultat rapporté dans [34] : la surexpression de Rb bloque l'activité de E2F et empêche le passage au point de restriction [69]. Par conséquent, les cyclines E et A, responsables du passage de G1 vers S ne sont pas présentes.

P27=1. Là encore, on obtient une disparition des oscillations avec un point fixe (1011110100). Dans ce point fixe, les variables E2F, CycE, et CycA prennent toutes la valeur 1. Ceci montre que l'arrêt du cycle s'effectue dans la transition G1/S. Cet arrêt

du cycle dans $G1/S$ est confirmé dans [53] qui montre que P27 se lie à la Cycline E en fin de phase $G1$ pour empêcher le passage en S .

Ubc=1. On obtient une disparition de l'attracteur complexe et l'apparition du point fixe (1011000110). Ce point est dans $G1$ car les variables E2F et CycE prennent la valeur 1 et CycA et CycB sont à 0. L'absence de la variable CycA lorsque $Ubc = 1$ est cohérente avec le résultat évoqué dans [73] qui stipule que UbcH10 a pour rôle d'empêcher, à la fin de la phase $G1$, l'accumulation de la cycline A, indispensable à la phase S .

Mutations oscillantes : Rb=0, P27=0, CycE=1, E2F=1. Nous nous intéressons aux quatre mutations Rb-/-, P27-/-, CycEop et E2Fop qui ont été testées et pour lesquelles le modèle maintient des oscillations². Notre objectif est de construire un graphe résumé, pour chaque modèle mutant, afin de visualiser l'impact de la mutation sur les oscillations. Le tableau 2.3 ci-après résume les principaux résultats.

Mutation	Comportements asymptotiques du modèle
Rb=0	▷ CycD=0 : attracteur de 224 états ▷ CycD=1 : attracteur de 112 états
P27 =0	▷ CycD=0 : point fixe G0 ▷ CycD=1 : attracteur de 112 états
CycE =1	▷ CycD=0 : attracteur de 56 états dans lequel P27=0 et point fixe dans lequel P27=Rb=Cdh1=1 (G0) ; ▷ CycD=1 : attracteur de 56 états
E2F =1	▷ CycD=0 : attracteur de 28 états ; et point fixe dans lequel P27=Rb=Cdh1=1 (G0) ; ▷ CycD=1 : attracteur de 28 états

TABLE 2.3 – Quatre mutations oscillantes du modèle de [26] (voir <http://ginsim.org/taxonomy/term/5?page=1> pour plus de détails).

Rb=0. Ce mutant retrouve le résultat évoqué dans [66, 8], qui est l'obtention d'un cycle cellulaire aussi bien en présence qu'en l'absence du facteur de croissance CycD. On obtient un attracteur de 224 états lorsque CycD=0 et un autre de 112 états lorsque CycD=1. Construire un graphe résumé pour chacun des deux attracteurs nous permet de vérifier que ces oscillations se comportent comme le modèle initial (voir tableau 2.3). Autrement dit, cette mutation n'affecte pas le comportement qualitatif du cycle en présence de facteur de croissance (succession des phases correcte) et génère le même type d'oscillations en l'absence de facteur de croissance.

P27=0. Cette mutation n'a pas d'impact sur les oscillations par rapport à celles du modèle non perturbé : elle conserve le point fixe G0 lorsque CycD=0 et un attracteur de 112 états lorsque CycD=1, dont le graphe résumé correspond exactement à celui du modèle initial.

CycE=1. Cette mutation conserve un attracteur complexe en présence du facteur de croissance CycD, conformément à [67]. Lorsque nous construisons le graphe résumé, les oscillations semblent dégradées (tableau 2.3), certaines phases disparaissant du cycle. Ceci

2. <http://ginsim.org/taxonomy/term/5?page=1>

est dû au fait que la variable CycE fait partie des quatre variables prises en compte dans la construction de la partition. Il nous faut donc adapter la partition par rapport à cette mutation. Pour cela, nous proposons d'utiliser la table de vérité de la table 2.4 (gauche), qui sont des adaptations de la table de vérité 2.2. Nous notons cette nouvelle partition $\mathcal{P}' = \{G1', G1/S', SG2', M/G2', M', M/G1', U'\}$, et nous avons $|G1'| = 8$, $|G1/S'| = 8$, $|SG2'| = 8$, $|G2/M'| = 6$, $|M'| = 6$, $|M/G1'| = 8$ et $|U'| = 12$. Nous remarquons que seuls les ensembles $G1$ et $G1/S$ voient leur nombre d'états réduit, ceci correspondant à l'accélération de la phase $G1$, observée dans [94]. Le nouveau graphe résumé est représenté dans la figure 2.9 (gauche). Ce graphe montre le même cycle cellulaire que l'attracteur du type sauvage à l'exception de l'arc $G1 \rightarrow M/G1$ qui a disparu.

<i>E2F</i>	<i>CycA</i>	<i>CycB</i>	<i>ensembles</i>
0	0	0	$M/G1'$
0	0	1	M'
0	1	0	$SG2'$
0	1	1	$G2/M'$
1	0	0	$G1'$
1	0	1	U'
1	1	0	$G1/S'$
1	1	1	U'

<i>CycE</i>	<i>CycA</i>	<i>CycB</i>	<i>ensembles</i>
0	0	0	$M/G1'$
0	0	1	M'
0	1	0	$SG2'$
0	1	1	$G2/M'$
1	0	0	$G1'$
1	0	1	U'
1	1	0	$G1/S'$
1	1	1	U'

TABLE 2.4 – Gauche : table de vérité pour classer les états du cas CycE=1. Droite : table de vérité pour classer les états du cas E2F=1.

E2F=1. Comme dans le cas précédent, cette mutation affecte une variable de classement. Nous devons donc à nouveau adapter la partition, en utilisant la table 2.9(droite). En revanche, cette fois certaines phases ne sont plus représentées dans l'attracteur : $|G1'| = 16$, $|G1/S'| = 16$, $|U'| = 24$ et $|SG2'| = |G2/M'| = |M'| = |M/G1'| = 0$. Le graphe résumé, représenté dans la figure 2.9 (droite), montre une forte dégradation des oscillations du point de vue qualitatif. Ceci confirme l'observation faite par les auteurs de [26], indiquant que ce mutant ne se comporte pas comme attendu.

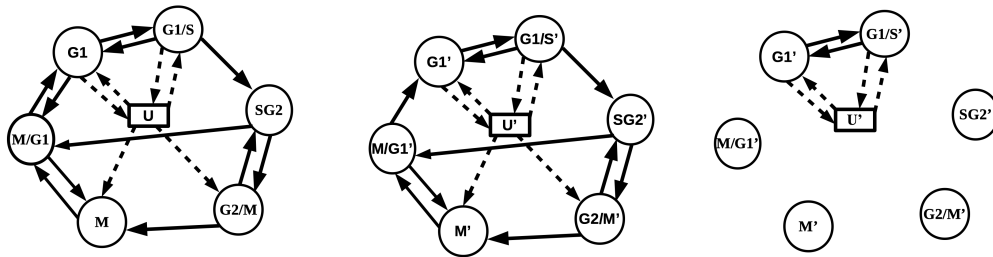


FIGURE 2.9 – Graphes résumés des mutants Rb=0, P27=0 (gauche), CycE=1 (milieu) et E2F=1 (droite).

On voit bien ici l'intérêt du graphe résumé dans l'analyse de mutants. Il permet en effet de repérer si une mutation altère qualitativement un attracteur, en affectant l'ordre

de succession des phases. Ici par exemple, les mutations CycE=1 et E2F=1, bien que similaires dans leurs effets globaux (attracteur complexe conservé dans les deux cas) révèlent en fait des phénotypes distincts car les deux attracteurs ont des structures sous-jacentes différentes.

2.3 Graphe résumé probabilisé d'un attracteur

Jusque-là, nous avons considéré le graphe résumé comme un moyen de visualisation, simple, d'un attracteur complexe. A présent, nous proposons d'associer une chaîne de Markov au graphe résumé, permettant ainsi d'estimer la proportion de temps pris par chaque phase du cycle cellulaire. Comme évoqué dans le chapitre 1 l'idée d'associer une chaîne de Markov à un graphe de transition asynchrone n'est pas nouvelle. Dans la littérature, plusieurs utilisations des chaînes de Markov pour comprendre une dynamique asynchrone ont été étudiées [13, 87]. Dans [56], Mendes *et al.* proposent notamment deux algorithmes permettant de quantifier l'atteignabilité des attracteurs asynchrones complexes.

Dans la suite, nous décrivons tout d'abord comment définir une chaîne de Markov sur un graphe résumé puis nous verrons comment utiliser cette chaîne de Markov pour compléter l'analyse d'un modèle.

2.3.1 Probabilités de transition sur un graphe résumé

Nous reprenons les notations des parties précédentes : $A = (V, E)$ est le graphe d'un attracteur asynchrone, $\mathcal{P} = \{V_1, V_2, \dots, V_k\}$ est une k -partition de V et $A(\mathcal{P}) = (V(\mathcal{P}), E(\mathcal{P}))$ est le graphe résumé de A associé à $\mathcal{P}$. L'espace d'états de la chaîne de Markov est donné par $\mathbb{E} = \{V_1, V_2, \dots, V_k\}$ et sa matrice de transition sera notée $P = (P_{ij})_{1 \leq i, j \leq k}$.

Comme évoqué plus haut, pour définir la matrice de transition P , nous nous appuyons implicitement sur une probabilité uniforme des transitions de l'attracteur, et plus précisément entre les ensembles V_i . Nous avons introduit plus haut, pour chaque couple (i, j) l'ensemble T_{ij} des transitions entre les états de V_i et ceux de V_j . Par extension, nous définissons T_{ii} comme l'ensemble des transitions d'un état de V_i vers un autre état de V_i . La probabilité associée à l'arc $V_i \rightarrow V_j$, notée P_{ij} , est alors donnée par :

$$P_{ij} = \frac{|T_{ij}|}{\sum_{l=1}^k |T_{il}|} \quad (2.3)$$

D'après la formule (2.3), $P = (P_{ij})_{1 \leq i, j \leq k}$ est une matrice stochastique, c'est à dire $P_{ij} \geq 0$ et

$$\forall i = 1 \dots k, \sum_{j=1}^k P_{ij} = 1.$$

P est donc la matrice de transition d'une chaîne de Markov. Par définition, cette chaîne est la chaîne de Markov associée au graphe résumé $A(\mathcal{P})$. L'idée de construire une chaîne de Markov sur un graphe résumé se rapproche de la notion de chaîne agrégée (lumped Markov chain, en anglais), telle que définie dans [45]. L'approche choisie dans la formule

(2.3) est légèrement plus générale, car elle peut s'appliquer à n'importe quel attracteur et pas seulement à une chaîne agrégeable ("lumpable"), définie dans [45].

Le graphe A étant un attracteur, le théorème 1 nous assure que $A(\mathcal{P})$ est fortement connexe, et donc la chaîne P est ergodique [9]. Dans le cadre de cette thèse, nous nous intéressons plus particulièrement au cas où P est régulière. P étant ergodique, il nous suffit de vérifier que les P_{ii} sont strictement positifs [9]. Dans notre cas, cela signifie que la partition $\mathcal{P}$ est construite de telle façon qu'il existe au moins une transition interne à chaque groupe V_i , en d'autres mots, les T_{ii} sont non vides. Cette condition suffisante sera vérifiée pour tous les graphes résumés étudiés dans cette thèse.

Lorsque P est régulière, il existe $\omega = (\omega_1, \omega_2, \dots, \omega_k)$ tel que

$$\begin{cases} 0 \leq \omega_i \leq 1, \\ \omega P = \omega. \end{cases} \quad (2.4)$$

Cette relation signifie que ω est un vecteur propre à gauche associé à la valeur propre $\lambda = 1$. La régularité de P implique que la valeur propre $\lambda = 1$ est simple et dominante (théorème de Perron-Frobenius, [9]). Ainsi, ω est l'unique vecteur propre à gauche associé à la valeur 1. Ce vecteur ω est appelée la distribution stationnaire de la chaîne de Markov et elle représente l'état asymptotique de la chaîne. Plus précisément, la i -ème composante ω_i s'interprète comme la fraction du temps passé dans l'ensemble V_i [39].

Dans la partie 2.3.2, nous utilisons d'abord cette distribution stationnaire pour estimer la durée des phases du cycle cellulaire dans l'attracteur de [26], et ainsi les comparer avec les durées biologiques. Ensuite, dans la partie 2.3.3, nous verrons comment généraliser cette chaîne de Markov en ajoutant des poids aux différentes variables, afin d'ajuster le modèle pour le rapprocher au comportement réel du cycle cellulaire.

2.3.2 Estimation du temps passé dans les phases du cycle cellulaire

Nous considérons le graphe résumé $A(\mathcal{P})$ de la figure 2.7. Nous calculons la matrice P de la chaîne de Markov à partir de la formule (2.3) :

$$\begin{array}{c} \begin{array}{ccccccc} G1 & G1/S & SG2 & G2/M & M & M/G1 & U \end{array} \\ \begin{array}{c} G1 \\ G1/S \\ SG2 \\ G2/M \\ M \\ M/G1 \\ U \end{array} \end{array} \begin{bmatrix} 0.7326 & 0.1056 & 0 & 0 & 0 & 0.1000 & 0.0618 \\ 0.1674 & 0.5382 & 0.2007 & 0 & 0 & 0 & 0.0938 \\ 0 & 0 & 0.5521 & 0.1875 & 0 & 0.2604 & 0 \\ 0 & 0 & 0.2639 & 0.4722 & 0.2639 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5278 & 0.4722 & 0 \\ 0.4583 & 0 & 0 & 0 & 0.0729 & 0.4688 & 0 \\ 0.1190 & 0.0861 & 0 & 0.0944 & 0.1106 & 0 & 0.5898 \end{bmatrix}. \quad (2.5)$$

Nous pouvons tout d’abord noter que le long du cycle cellulaire $G1 \rightarrow G1/S \rightarrow SG2 \rightarrow G2/M \rightarrow M \rightarrow M/G1 \rightarrow G1$, la probabilité de transition d’un sommet vers son successeur est toujours supérieure à celle de ce sommet vers les états non classés U . Ceci est un premier constat indiquant que les états non classés ne semblent pas prédominants dans la dynamique.

Afin de confirmer ce constat, nous calculons la distribution stationnaire

$$\begin{aligned}\omega &= (\omega_{G1}, \quad \omega_{G1/S}, \quad \omega_{SG2}, \quad \omega_{G2/M}, \quad \omega_M, \quad \omega_{M/G1}, \quad \omega_U) \\ &= (0.4227, \quad 0.1133, \quad 0.0762, \quad 0.0431, \quad 0.0732, \quad 0.1819, \quad 0.0896).\end{aligned}$$

Cette distribution est représentée sous la forme d’un camembert dans la figure 2.10 (droite). En effet, même si U est le groupe contenant la majorité d’états (Fig. 2.10, milieu), le temps passé dans les états non classés ω_U reste faible (Fig. 2.10, droite). De manière générale, la comparaison de ces deux figures montre une certaine différence entre les proportions du nombre d’états dans les groupes d’une part et les probabilités asymptotiques d’autre part, confirmant l’intérêt de la chaîne de Markov pour l’estimation du temps des phases.

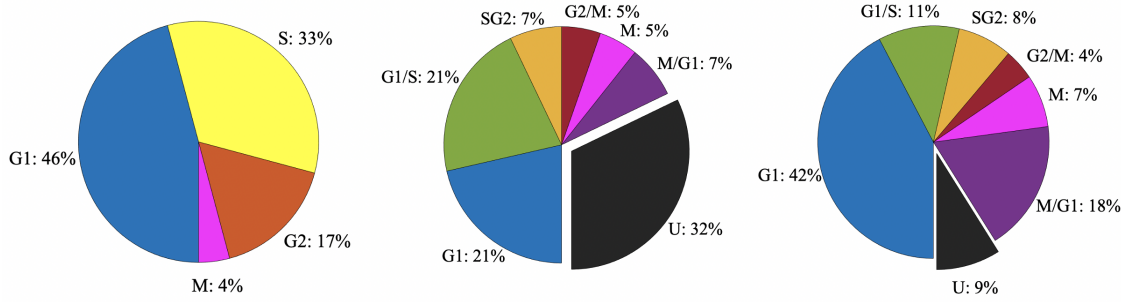


FIGURE 2.10 – Proportions du temps passé dans chaque phase du cycle cellulaire. Gauche : proportions calculées à partir des données expérimentales de [21] ; Milieu : proportions calculées à partir du nombre d’états dans chaque ensemble de la partition $\mathcal{P}$; droite : proportions calculées à partir de la distribution stationnaire de la matrice de transition (2.5). Notons que les six “phases” des deux camemberts de droite ne correspondent pas exactement aux quatre phases biologiques classiques $G1$, S , $G2$ et M (à gauche), en particulier les transitions $G1/S$, $G2/M$ et $M/G1$ y sont implicites (voir section 2.2.1).

Dans [21] est estimée la durée des phases $G1$, S , $G2$ et M dans le cycle de cellules humaines. Ces durées sont reproduites dans la (Fig. 2.10, gauche). En comparant ces données biologiques avec la distribution stationnaire, plusieurs constats peuvent être faits. Tout d’abord, nous remarquons que la phase $G1$ est majoritaire (autour de 45%), ce qui est également le cas dans le modèle (42% sans compter les phases de transitions $M/G1$ et $G1/S$). En revanche, les proportions des autres phases présentent des différences notables. En particulier, le temps de la phase S est clairement sous-estimé alors que celui de la mitose (M) est sur-estimé.

Il est relativement logique que les probabilités asymptotiques ne reflètent pas exactement les proportions expérimentales, puisque ces probabilités sont calculées par la formule (2.3) qui ne prend pas en compte les vitesses relatives des différents processus biologiques

en jeu. Dans ce qui suit, nous proposons une adaptation de la formule (2.3) qui prend en compte la nature des transitions en leur affectant des priorités différentes.

2.3.3 Prise en compte des vitesses relatives dans la chaîne de Markov

Dans la formule (2.3), nous avons considéré que toutes les transitions issues d'un même état sont équiprobables. Il s'agit bien sûr d'une hypothèse forte, loin de la réalité biologique. Afin de mieux représenter les différentes vitesses, nous allons affecter des priorités différentes aux transitions puis nous généraliserons la formule (2.3).

Considérons à nouveau A le graphe d'un attracteur asynchrone et $A(\mathcal{P})$ le graphe résumé de A sur la partition $\mathcal{P} = \{V_1, V_2, \dots, V_k\}$. Les transitions de l'attracteur étant asynchrones, on peut identifier précisément, pour chacune, l'unique variable qui passe de 0 à 1 ou de 1 à 0. Nous introduisons pour chaque variable x un poids ρ_x qui est un réel strictement positif. ρ_x représente la vitesse des processus impliquant x . Un poids $\rho_x > 1$ indique une variable rapide et un poids $\rho_x < 1$ une variable lente. Ainsi, nous pouvons associer un poids à chaque transition de l'attracteur. Partant du constat que la vitesse d'activation (0 à 1) et d'inhibition (1 à 0) d'une variable peuvent être très différentes, nous allons plutôt considérer deux poids ρ_x^+ et ρ_x^- pour chaque variable x , correspondant respectivement à sa vitesse d'activation et celle d'inhibition. Pour tout arc (V_i, V_j) de $A(\mathcal{P})$, nous posons $\mathbb{Q}_{ij}$ la somme des poids de toutes les transitions dans T_{ij} . La probabilité de transition associée à l'arc $V_i \rightarrow V_j$ est alors donnée par la formule (2.6) qui généralise la formule (2.3) :

$$P_{ij} = \frac{\mathbb{Q}_{ij}}{\sum_{l=1}^k \mathbb{Q}_{il}}. \quad (2.6)$$

L'objectif de l'introduction des poids est de paramétrer le modèle de [26] pour que les proportions du temps passé dans les phases du cycle cellulaire puissent s'approcher au mieux des données biologiques (Fig. 2.10, gauche). Nous introduisons ρ^+ et $\rho^- \in \mathbb{R}_+^{10}$, ρ^+ désigne le vecteur des poids d'activation et ρ^- , celui des poids d'inhibition. Cet espace de paramètres, avec 20 degrés de liberté, peut être réduit si des variables ont le même poids d'activation ou d'inhibition. Dans ce cas, les variables sont réparties en des groupes en fonction de la nature des processus qui les impliquent. Bien sûr, une variable peut être dans un groupe en activation et dans un autre en inhibition. Ceci correspond à la notion de classes de priorité introduite dans [26, 64], à la différence que nous conservons les transitions de faible priorité même lorsqu'elles sont en concurrence avec celles de priorité forte. Nous considérons le tableau 2.5 qui donne la répartition des variables du modèle en quatre groupes, proposée dans [26]. Cette répartition est faite selon que les variables sont régulées par un mécanisme similaire ou selon qu'elles partagent certaines propriétés dynamiques, telles qu'une vitesse de dégradation ou d'activation.

Nous allons à présent fixer des valeurs pour les ρ_i . Afin de nous approcher des temps biologiques, notre objectif est de choisir les poids pour avoir :

- 1- La durée de $G1$ majoritaire ($\omega_{G1} > \omega_i$),
- 2- La durée de $SG2$ supérieure à celle de la mitose et des phases de transitions ($\omega_{SG2} > \omega_M, \omega_{G1/S}, \omega_{G2/M}, \omega_{M/G1}$),

Type de poids(activation/inhibition)	poids
CycD,Rb,P27,Cdh1,E2F ↓, CycE ↓	ρ_1
CycA↓, Cdc20↓, Ubc↓, CycB↓	ρ_2
E2F↑, CycE↑, CycA↑, Cdc20↑	ρ_3
Ubc↑, CycB↑	ρ_4

TABLE 2.5 – Quatre classes de priorité de transition proposées dans[26]. A chaque classe i est associée un poids $\rho_i > 0$. Le symbol $\uparrow$ spécifie le poids d’activation, et $\downarrow$ spécifie le poids d’inhibition de la variable. Les variables qui ne sont pas succédé de flèche ont un poids d’activation égal à leur poids d’inhibition.

3- La durée de M inférieure à celle de $G1$ et de $SG2$ ($\omega_M < \omega_{G1}, \omega_{SG2}$).
Afin d’essayer de réaliser ces objectifs, nous calculons la distribution stationnaire, ω , en faisant varier chaque ρ_i de façon indépendante. Une partie de ces résultats est présentée dans la Fig. 2.11

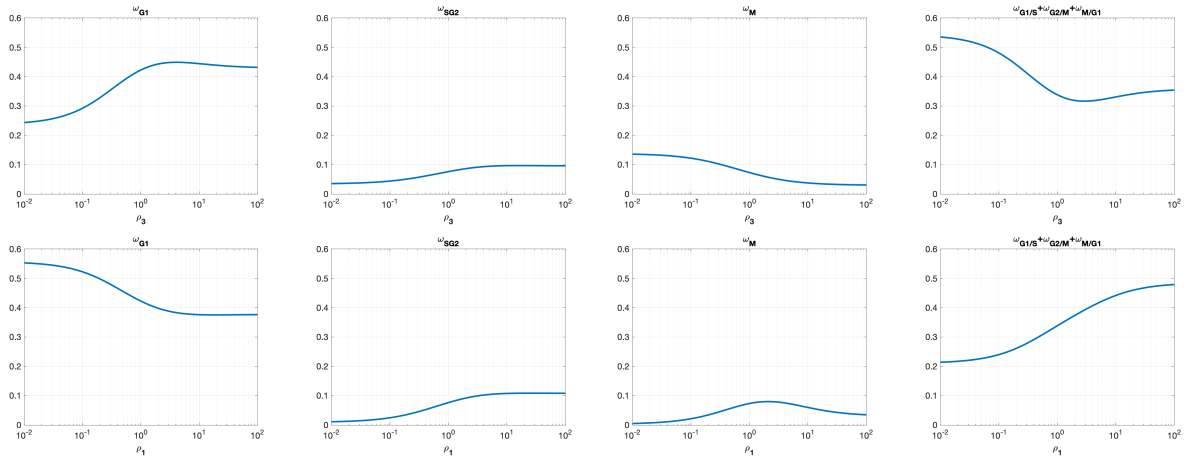


FIGURE 2.11 – Évolution de la proportion du temps passé dans chacune des trois phases ω_{G1} , ω_{SG2} et ω_M et les proportions cumulées ($\omega_{G1/S} + \omega_{G2/M} + \omega_{M/G1}$) des phases de transitions en fonction des poids ρ_1 ou ρ_3 . Chaque ω_i correspond à la i -ième composante de la distribution stationnaire ω obtenue en variant de façon indépendante ρ_1 ou ρ_3 .

Dans la Fig. 2.11, nous présentons l’évolution des proportions ω_{G1} , ω_{SG2} et ω_M et du cumul $\omega_{G1/S} + \omega_{G2/M} + \omega_{M/G1}$ en fonction de ρ_1 et de ρ_3 . Nous nous sommes concentrés sur ces deux poids car ils présentent des tendances plus pertinentes pour les objectifs que nous nous sommes fixés plus haut. D’après la Fig. 2.11, nous observons qu’augmenter la valeur de ρ_3 permet à la fois d’augmenter ω_{SG2} et de diminuer ω_M tout en maintenant ω_{G1} majoritaire (autour de 43%) et sans trop impacter les phases de transitions (autour de 35%). S’agissant de ρ_1 , une augmentation permet également d’augmenter ω_{SG2} et de diminuer ω_M (au moins à partir de $\rho_1 = 2$). Par contre, ρ_1 doit être moyennement élevé car il a tendance à augmenter le cumul des proportions des phases de transitions. Ainsi, pour satisfaire les objectifs précédents, nous choisissons un ρ_3 fort et un ρ_1 intermédiaire et nous posons $\rho_2 = \rho_4 = 1$ faibles. Par exemple, en choisissant le vecteur suivant : $\rho = (50, 1, 100, 1)$, nous obtenons la distribution stationnaire présentée dans la Fig.2.12. Par rapport au cas équiprobable, l’introduction des poids nous a permis de changer la

répartition des durées de phases en nous approchant des résultats expérimentaux. En particulier, $G1$ reste majoritaire, S et $G2$ ont augmenté et M a diminué. De façon tout à fait intéressante, le temps passé dans les états non classés U a fortement diminué.

Bien sûr, nos différentes tentatives sont loin d'être exhaustives, et il est clair que ces résultats pourraient être largement améliorés en appliquant une analyse de sensibilité plus systématique. Néanmoins, ces premiers résultats montrent l'intérêt, pour la modélisation, de combiner les chaînes de Markov à l'outil graphe résumé.

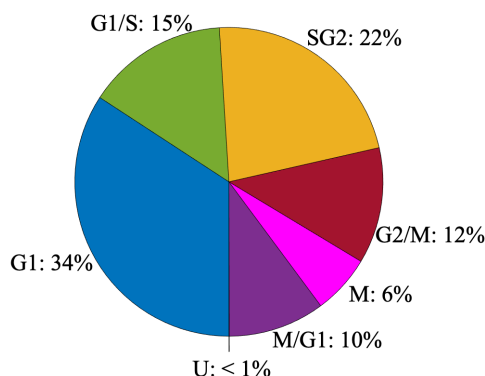


FIGURE 2.12 – Proportions du temps passé dans chaque phase du cycle, calculées à partir de la distribution stationnaire de la chaîne de Markov avec les poids $\rho = (50, 1, 100, 1)$.

Dans ce chapitre, nous avons introduit la notion de graphe résumé d'un attracteur asynchrone complexe. Si l'outil n'est pas nouveau du point de vue de la théorie des graphes, son utilisation dans l'analyse de modèles Booléens, particulièrement de modèles Booléens de systèmes biologiques, nous paraît importante. Si la détection d'attracteurs complexes dans des graphes de transition est relativement aisée, l'analyse de ces attracteurs quand ils dépassent la dizaine d'états devient rapidement délicate. Le plus souvent, les descriptions temporelles de phénomènes oscillant dans les cellules sont essentiellement qualitatives (succession d'évènements discrets), ce qui rend le graphe résumé pertinent pour pouvoir comparer un tel attracteur au phénomène modélisé et ainsi valider un modèle Booléen du point de vue biologique. Le graphe résumé, ainsi que les chaînes de Markov construites dessus, constituent donc un outil tout à fait intéressant dans l'étude de modèles discrets de réseaux de régulation biologique.

La recherche d'une partition adéquate est au cœur de l'utilisation de l'outil graphe résumé. Cette partition doit à la fois permettre une analyse pertinente du graphe, tout en intégrant le maximum d'informations biologiques sur le phénomène. Pour le cycle cellulaire, nous avons bien sûr tiré parti d'une littérature abondante décrivant les différentes phases du cycle, déjà intégrée pour la plupart dans le modèle de [26]. Dans le chapitre suivant, nous proposons un modèle Booléen d'un autre système oscillant au cœur des cellules mammifères : l'horloge circadienne. Dans la construction et l'analyse de ce nouveau modèle, nous mettons à profit la méthode du graphe résumé, à la fois pour proposer un

découpage en “phases” qualitatives du cycle circadien, ainsi que pour vérifier la correcte succession de ces phases dans notre modèle.

Chapitre 3

Construction et analyse d'un modèle du cycle circadien

Dans ce chapitre, nous mettons en œuvre la méthode du graphe résumé présentée dans le chapitre précédent sur un nouveau modèle, centré autour de la régulation de l'horloge circadienne. Comme pour le cycle cellulaire, le réseau de l'horloge est au cœur des cellules eucaryotes et de nombreux modèles existent, sous différentes formes. En particulier, [4] propose un modèle de 8 équations différentielles ordinaires reproduisant les régulations principales de l'horloge chez les cellules mammifères. Afin d'aborder ce système d'un point de vue qualitatif, comme nous l'avons fait au chapitre 2 pour le cycle cellulaire, nous proposons tout d'abord une version Booléenne de ce modèle, aussi fidèle que possible au modèle original. Pour cela, nous allons détailler les différentes hypothèses que nous sommes amenés à faire dans la construction de notre modèle Booléen. La présence d'un attracteur complexe, suivie de l'analyse de son graphe résumé nous permettent ensuite de le valider du point de vue biologique.

Contrairement au cycle cellulaire, qui a très tôt été décrit en terme de succession de phases qualitatives, le découpage en phases du cycle circadien est moins standard. Les articles [46, 84] proposent un tel découpage sur la base de données exclusivement transcriptionnelles. En nous basant sur ce découpage et en l'adaptant aux variables de notre modèle, nous proposons une partition en cinq phases qui nous permet de construire un graphe résumé pertinent. Cette façon de construire une partition est légèrement différente de celle proposée au chapitre 2, et nous permet de montrer que la construction d'une bonne partition est au cœur même du processus de modélisation. Grâce au graphe résumé, nous pouvons décrire qualitativement les oscillations de notre modèle et ainsi valider notre découpage. Nous utilisons également une chaîne de Markov sur le graphe résumé afin d'analyser plus particulièrement un certain type de déphasage dans notre modèle, relié au syndrome d'avance de phase.

Nous présentons la construction du modèle Booléen dans la partie 3.1, suivie du calcul rapide de ses dynamiques. Dans la partie 3.2, nous construisons le graphe résumé de l'attracteur, en proposant notamment un découpage en phases qualitatives du déroulement du cycle circadien chez les cellules mammifères. Enfin, dans la partie 3.3, nous étendons notre analyse en utilisant les chaînes de Markov, en étudiant particulièrement l'impact de quatre paramètres du modèle Booléen sur le déphasage du cycle.

3.1 Construction d'un modèle Booléen de l'horloge circadienne

A l'instar du cycle cellulaire, l'horloge circadienne est observée chez la plupart des organismes eucaryotes tels que la drosophile [97], les plantes [83], les mammifères [3] mais aussi chez les procaryotes tels que les cyanobactéries [47]. Comme expliquée au chapitre 1, l'horloge circadienne se synchronise à l'alternance jour/nuit, permettant ainsi à l'organisme de s'adapter au caractère périodique de l'environnement terrestre. Au niveau cellulaire, les oscillations se caractérisent par une variation périodique de certains constituants sur une période d'environ 24h [36]. Plusieurs modèles existent dans la littérature notamment basés sur des EDO.

3.1.1 Modélisations continues de la dynamique de l'horloge

Comme expliqué dans le chapitre 1, le cœur des régulations cellulaires responsable des oscillations circadiennes est aujourd'hui bien connu, et se centre autour de trois boucles de rétroaction impliquant le complexe Clock:Bmal1 (voir Figure 1.5). L'un des premiers modèles de l'horloge circadienne a été proposé en 1965 par Goodwin [38], basé sur une simple boucle de rétroaction négative entre une protéine et son propre gène. En 2003, Leloup et Golbeter [51] ont montré la génération d'oscillations soutenues en obscurité constante, par un modèle de 16 EDO. Dans [14], trois formalismes, Booléen, linéaire par morceaux et EDO non linéaires sont proposés pour étudier l'oscillateur Kai ABC qui est au cœur de l'horloge circadienne de la cyanobactérie.

Le modèle récemment proposé par Almeida *et al.* [4], est l'un des modèles les plus complets des oscillations circadiennes des cellules mammifères. Il implique 8 variables principales : *BMAL1* et *PC* qui représentent les deux complexes Clock:Bmal1 et Per:Cry, *PER* et *CRY* qui désignent les protéines Per et Cry précédemment citées, et les variables *ROR*, *REV*, *DBP* et *E4* qui regroupent le reste du réseau. Les équations de ce modèle sont reproduites ici :

$$\frac{d}{dt}BMAL1 = Rbox - \gamma_{bp}BMAL1 \cdot PC, \quad (3.1)$$

$$\frac{d}{dt}ROR = Ebox + Rbox - \gamma_{ror}ROR, \quad (3.2)$$

$$\frac{d}{dt}REV = 2Ebox + Dbox - \gamma_{rev}REV, \quad (3.3)$$

$$\frac{d}{dt}DBP = Ebox - \gamma_{dbp}DBP, \quad (3.4)$$

$$\frac{d}{dt}E4 = 2Rbox - \gamma_{e4}E4, \quad (3.5)$$

$$\frac{d}{dt}CRY = Ebox + 2Rbox - M_{PC} - \gamma_cCRY, \quad (3.6)$$

$$\frac{d}{dt}PER = Ebox + Dbox - M_{PC} - \gamma_pPER, \quad (3.7)$$

$$\frac{d}{dt}PC = M_{PC} - \gamma_{bp}BMAL1 \cdot PC. \quad (3.8)$$

Le terme $M_{PC} = \gamma_{pc}PER \cdot CRY - \gamma_{cp}PC$ traduit la liaison et la dissociation dans le processus de formation du complexe PC .

Le modèle prend en compte les gènes dits contrôlés par l'horloge (CCE, clock controlled elements) $Rbox$, $Ebox$ et $Dbox$. Chacune de ces box possède un activateur et un inhibiteur, donnés dans le tableau 3.1. L'activation des CCEs est représentée par des combinaisons de termes de type Michaelis-Menten ou Hill :

$$Ebox = V_E \frac{[BMAL1]}{[BMAL1] + k_E + k_{Er}[BMAL1][CRY]}, \quad (3.9)$$

$$Rbox = V_R \frac{[ROR]}{[ROR] + k_R} \frac{k_{Rr}^2}{k_{Rr}^2 + [REV]^2}, \quad (3.10)$$

$$Dbox = V_D \frac{[DBP]}{[DBP] + k_D} \frac{k_{Dr}}{k_{Dr} + [E4]}. \quad (3.11)$$

CCEs	activateur	inhibiteur
$Ebox$	$BMAL1$	CRY
$Rbox$	ROR	REV
$Dbox$	DBP	$E4$

TABLE 3.1 – Activateurs et inibiteurs des trois CCEs (clock controlled elements)

Le modèle reproduit, d'une part, les oscillations circadiennes caractérisées par une opposition de phase entre les deux complexes Clock:Bmal1 et Per:Cry (voir Figure 1.7 et [4]). D'autre part, l'ordre des pics dans l'expression des protéines Bmal, Per, et Cry est bien conforme aux résultats expérimentaux de [46]. Afin d'appliquer nos outils et d'analyser les oscillations d'un point de vue qualitatif (comme enchaînement de phases), nous proposons dans la suite de construire une version discrète de ce système, sous la forme d'un réseau Booléen asynchrone. La traduction automatique de modèles continus en modèles Booléens se voit de plus en plus dans la littérature (voir par exemple [82, 57]). Dans notre cas, nous ne proposons pas une méthode automatique, mais nous traitons les équations au cas par cas pour trouver, pour chacune, une fonction Booléenne s'approchant au mieux des hypothèses du modèle original. Notre objectif principal est de retrouver des oscillations dans le modèle Booléen afin de les analyser grâce à nos outils.

3.1.2 Élaboration du modèle Booléen

Nous commençons par traduire l'activation des trois "box", donnée par (3.9)-(3.11), en choisissant les fonctions Booléennes suivantes :

$$Ebox = BMAL1 \wedge \overline{CRY} \quad (3.12)$$

$$Rbox = ROR \wedge \overline{REV} \quad (3.13)$$

$$Dbox = DBP \wedge \overline{E4}. \quad (3.14)$$

Dans ces fonctions (3.12), (3.13) et (3.14), nous considérons, pour chaque box, l'activateur et l'inhibiteur donnés dans le tableau 3.1, et nous supposons qu'elle est active en présence

de son activateur et en l'absence de son inhibiteur. Pour traduire chacune de ces équations différentielles en une règle Booléenne de transition locale, nous étudions au cas par cas chaque terme de synthèse et de dégradation. De manière générale, nous ne prenons pas en compte les termes de dégradation dans la construction des fonctions Booléennes car ils ne correspondent à aucune régulation biologique explicite. La seule exception ici sera le terme de dégradation de $BMAL1$, $-\gamma_{bp}BMAL1 \cdot PC$, qui correspond à la liaison des deux complexes Clock:Bmal1 et Per:Cry dans le noyau, entraînant une inhibition de $BMAL1$ par PC . Par ailleurs, les variables $BMAL1$, ROR , REV , DBP et $E4$ (3.1 à 3.5) sont régulées positivement par un ou deux CCEs. Pour celles qui sont régulées par deux CCEs, nous supposons qu'un seul suffit pour les activer et nous choisissons alors une disjonction pour obtenir les fonctions Booléennes suivantes :

$$BMAL1 = Rbox \wedge \overline{PC} \quad (3.15)$$

$$ROR = Ebox \vee Rbox, \quad (3.16)$$

$$REV = Ebox \vee Dbox, \quad (3.17)$$

$$DBP = Ebox, \quad (3.18)$$

$$E4 = Rbox. \quad (3.19)$$

Pour les variables PER , CRY et PC , les EDOs impliquent des termes de la loi d'action de masse. La traduction de ces termes en fonctions Booléennes est de manière générale assez délicate. En effet, le terme M_{PC} apparaissant dans les équations (3.6) à (3.8) induit des effets négatifs de PER sur CRY et *vice versa*, qui ne correspondent bien sûr à aucun processus d'inhibition biologique. De même, il induit un effet positif de PC sur PER et CRY qui ne correspond pas non plus à une activation explicite. Dans le cas Booléen, la formation du complexe PC peut être vue simplement comme un signal donné par la présence simultanée de PER et CRY . Ainsi, PER et CRY ont bien un effet positif sur PC , en revanche l'effet de PC sur PER et CRY est ignoré puisque la formation du complexe en Booléen n'affecte pas le niveau des deux variables impliquées. Les fonctions Booléennes sont les suivantes :

$$CRY = Ebox \vee Rbox, \quad (3.20)$$

$$PER = Ebox \vee Dbox, \quad (3.21)$$

$$PC = PER \wedge CRY. \quad (3.22)$$

Il est à noter que dans la construction de la fonction Booléenne de PC (3.22), nous avons ignoré le terme $-\gamma_{bp}BMAL1 \cdot PC$ de l'équation différentielle (3.8) correspondant à la dégradation de PC par $BMAL1$. Ce terme de dégradation a un sens dans le cas continue puisqu'il explique la liaison entre le complexe Per:Cry et le complexe Clock:Bmal1. Dans le contexte Booléen, la prise en compte de cette liaison dénature la boucle de rétroaction négative entre $BMAL1$ et PC de sorte que les oscillations circadiennes sont dégradées.

En remplaçant les expressions Booléennes des CCEs (3.12)-(3.14) dans les fonctions (3.15)-(3.21), nous obtenons ainsi un modèle Booléen de 8 variables. Cependant, ce modèle n'est pas encore satisfaisant car il ne présente pas d'oscillations. En regardant de plus près, les variables CRY , REV et ROR ont la particularité de réguler les CCEs tout en étant régulées par ces derniers. Afin de prendre en compte ceci dans le modèle booléen, nous décidons d'affecter à ces trois variables un niveau d'activité supplémentaire (0,1,2).

En utilisant la méthode proposée dans [15], nous représentons ces variables multivaluées en introduisant trois nouvelles variables Booléennes $CRY2$, $REV2$ et $ROR2$. Ainsi, pour représenter une variable $x \in \{0, 1, 2\}$, nous utilisons deux variables Booléennes $X, X2 \in \{0, 1\}$ de la façon suivante :

$$x = \begin{cases} 0 & \text{si } X = X2 = 0, \\ 1 & \text{si } X = 1, X2 = 0, \\ 2 & \text{si } X = X2 = 1. \end{cases}$$

Ce faisant, les états correspondant à $X2 = 1, X = 0$ sont à exclure car ils n'ont pas de sens biologique. Dans [15], une méthode est donnée pour construire les règles des variables X et $X2$ afin de s'assurer que ces états "interdits" ne soient pas atteints à partir d'états autorisés.

Maintenant, il s'agit de trouver des fonctions Booléennes pour les variables $CRY2$, $REV2$ et $ROR2$. Cela nous amène, par conséquent, à modifier toutes les fonctions Booléennes précédemment définies et faisant intervenir CRY , REV ou ROR . A commencer par les CCEs, à savoir $Ebox$ et $Rbox$, nous remplaçons ainsi le tableau 3.1 par le tableau 3.2. Nous supposons ainsi que l'inhibition de $Ebox$ par Cry se fait par la variable $CRY2$,

CCEs	activateur	inhibiteur
$Ebox$	$BMAL1$	$CRY2$
$Rbox$	ROR	REV
$Rbox2$	$ROR2$	REV
$Dbox$	DBP	$E4$

TABLE 3.2 – Inhibiteur et activateur des CCEs avec prise en compte des variables multivaluées.

c'est à dire lorsque Cry est à son niveau d'activité le plus élevé (niveau 2). Pour $Rbox$, étant régulée par deux variables multivaluées (ROR et REV), elle sera également multivaluée. Nous introduisons ainsi une nouvelle variable $Rbox2$ qui est activée en présence de $ROR2$. Ainsi, le tableau 3.2, nous donne les nouvelles expressions Booléennes de $Ebox$, $Rbox$ et $Rbox2$.

$$Ebox = BMAL1 \wedge \overline{CRY2} \quad (3.23)$$

$$Rbox = ROR \wedge \overline{REV} \quad (3.24)$$

$$Rbox2 = ROR2 \wedge \overline{REV}. \quad (3.25)$$

L'expression de $Dbox$ (3.14) est inchangée puisque $Dbox$ n'est régulée par aucune des variables CRY , REV ou ROR . Pour terminer la construction des fonctions Booléennes, il nous reste à expliciter certains choix de modélisation faits notamment après comparaison avec la solution du modèle d'EDO. Pour $BMAL1$, nous supposons que l'activation se fait à partir de $Rbox$ et non $Rbox2$, la fonction (3.15) restant inchangée. Pour CRY , nous supposons que l'activation se fait par $Ebox$, et que $Rbox$ joue sur $CRY2$:

$$CRY = (Ebox) \vee CRY2, \quad (3.26)$$

$$CRY2 = (Ebox \wedge Rbox) \wedge CRY. \quad (3.27)$$

De même pour REV , nous supposons que l'activation se fait par $Ebox$, et que $Dbox$ joue sur $REV2$:

$$REV = Ebox \vee REV2, \quad (3.28)$$

$$REV2 = Ebox \wedge Dbox \wedge REV. \quad (3.29)$$

Pour ROR , nous supposons que l'activation se fait par $Rbox2$:

$$ROR = (Rbox2) \vee ROR2, \quad (3.30)$$

$$ROR2 = (Ebox \wedge Rbox) \wedge ROR. \quad (3.31)$$

En remplaçant $Rbox2$ par son expression dans (3.30), nous obtenons :

$$\begin{aligned} ROR &= Rbox2 \vee ROR2, \\ &= (ROR2 \wedge \overline{REV}) \vee ROR2, \\ &= ROR2. \end{aligned}$$

Ceci veut dire qu'en considérant cette expression de ROR , l'inhibition de ROR par REV est négligée. Pour cette raison, nous considérons, au lieu de cette expression, que ROR est activée seulement par $Rbox2$:

$$ROR = Rbox2. \quad (3.32)$$

Pour récapituler, nous obtenons :

$$\begin{aligned} BMAL1 &= Rbox \wedge \overline{PC}, & (cf. \quad 3.15) \\ ROR &= Rbox2, & (cf. \quad 3.32) \\ REV &= Ebox \vee REV2, & (cf. \quad 3.28) \\ DBP &= Ebox, & (cf. \quad 3.18) \\ E4 &= Rbox, & (cf. \quad 3.19) \\ CRY &= Ebox \vee CRY2, & (cf. \quad 3.26) \\ PER &= Ebox \vee Dbox, & (cf. \quad 3.21) \\ PC &= PER \wedge CRY, & (cf. \quad 3.22) \\ REV2 &= Ebox \wedge Dbox \wedge REV, & (cf. \quad 3.29) \\ CRY2 &= Ebox \wedge Dbox \wedge CRY, & (cf. \quad 3.27) \\ ROR2 &= Ebox \wedge Dbox \wedge ROR, & (cf. \quad 3.30) \end{aligned}$$

avec

$$Ebox = BMAL1 \wedge \overline{CRY2}, \quad (cf. \quad 3.23)$$

$$Rbox = ROR \wedge \overline{REV}, \quad (cf. \quad 3.24)$$

$$Rbox2 = ROR2 \wedge \overline{REV}, \quad (cf. \quad 3.25)$$

$$Dbox = DBP \wedge \overline{E4}, \quad (cf. \quad 3.14).$$

Finalement, en remplaçant l'expression des box dans les fonctions Booléennes précédentes, nous obtenons un réseau Booléen de dimension 11 dont les règles sont données dans la Table 3.3. Le graphe d'interaction du modèle est donné dans la figure 3.1.

Variable	règle Booléenne
$BMAL1$	$ROR \wedge \overline{REV} \wedge \overline{PC}$
ROR	$ROR2 \vee \overline{REV}$
REV	$(BMAL1 \wedge \overline{CRY2}) \vee REV2$
DBP	$BMAL1 \wedge \overline{CRY2}$
$E4$	$ROR \wedge \overline{REV}$
CRY	$BMAL1 \vee CRY2$
PER	$(BMAL1 \wedge \overline{CRY2}) \vee (DBP \wedge \overline{E4})$
PC	$PER \wedge CRY$
$REV2$	$BMAL1 \wedge \overline{CRY2} \wedge DBP \wedge \overline{E4} \wedge REV$
$CRY2$	$BMAL1 \wedge \overline{CRY2} \wedge ROR \wedge \overline{REV} \wedge CRY$
$ROR2$	$BMAL1 \wedge \overline{CRY2} \wedge ROR \wedge \overline{REV}$

TABLE 3.3 – Règles Booléennes des 11 variables du modèle de l’horloge.

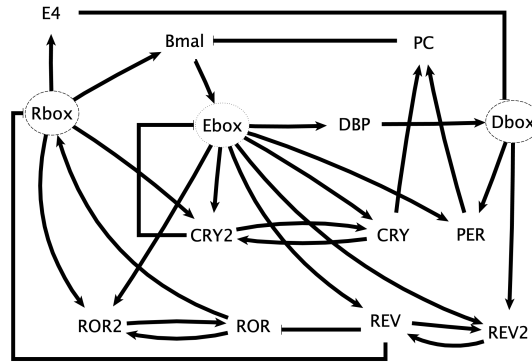


FIGURE 3.1 – Graphe d’interaction du modèle du modèle Booléen de l’horloge ; les CCEs sont les variables encadrées et elles ne figurent pas explicitement dans le modèle final.

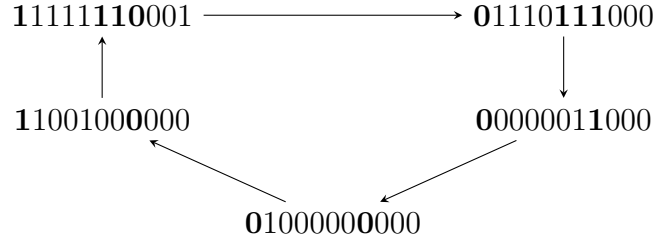


FIGURE 3.2 – Attracteur synchrone de 5 états du modèle de l’horloge ; l’ordre des variables est : *BMAL1*, *ROR*, *REV*, *DBP*, *E4*, *CRY*, *PER*, *PC*, *REV2*, *CRY2*, *ROR2*. Dans l’ordre, nous observons tout d’abord : Activation de *BMAL1* (11001000000), suivie de l’activation de *PER* et *CRY* (11111110001), suivie de l’activation de *PC* (01110111000 et 00000011000). L’état (01000000000) est l’intermédiaire entre la désactivation du complexe *PC* et la prochaine activation de *BMAL1*.

3.1.3 Analyse préliminaire de la dynamique du réseau

La simulation du modèle montre un unique attracteur synchrone de 5 états représenté dans la figure 3.2 et un unique attracteur asynchrone de 442 états et 1737 transitions représenté dans la figure 3.3.

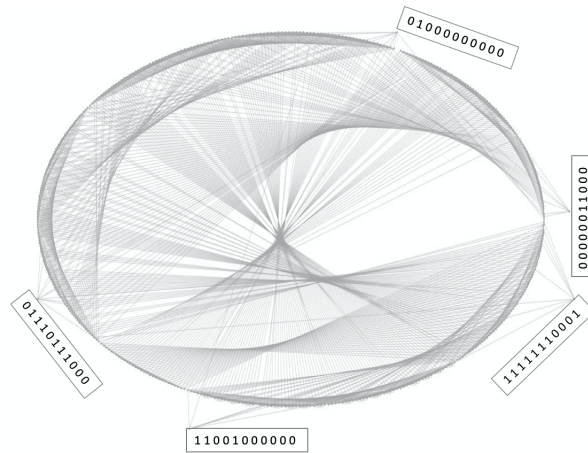


FIGURE 3.3 – Attracteur asynchrone complexe du modèle de l’horloge, contenant 442 états et 1737 transitions (graphe réalisé avec Matlab). Les 5 états de l’attracteur synchrone sont présents dans cet attracteur complexe.

A première vue, le modèle a capturé, pour les deux stratégies de mise à jour, un seul phénotype caractérisé par des oscillations (un cycle en synchrone, un attracteur complexe en asynchrone) que nous allons essayer de comparer aux oscillations circadiennes.

L’attracteur synchrone résume en 5 étapes l’ordre attendu dans l’activation des gènes de l’horloge *BMAL1*, *PER*, *CRY*, conformément à ce qui est attendu. En effet, le cycle de l’attracteur commence par l’activation de *BMAL1* (11001000000), puis par celle de *PER* et *CRY* (11111110001), puis par celle du complexe *PC* (01110111000 et 00000011000). Enfin, l’état 01000000000 est un intermédiaire entre la désactivation du complexe *PC* et le début d’un nouveau cycle, c’est à dire la prochaine activation de *BMAL1*. Dans ce cycle, on peut observer que *BMAL1* et *PC* sont actives dans des “phases” différentes. Plus

précisément, *BMAL1* est active dans la “phase” $\{11001000000, 11111110001\}$ et *PC* est active dans la “phase” $\{01110111000, 00000011000\}$. On retrouve donc bien l’opposition de phase entre les complexes Clock:Bmal1 et Per:Cry ([4, 46]).

Contrairement à l’attracteur synchrone, qui n’est constitué que d’un seul cycle de 5 états, l’attracteur asynchrone est complexe et composé de nombreux cycles entrelacés. Ainsi, il n’est pas évident d’observer des propriétés telles que l’ordre activation des variables (*BMAL1*, puis *PER*, *CRY*) ou même l’opposition de phase entre *BMAL1* et *PC*. Pour cela, nous allons utiliser la notion de graphe résumé développée dans le chapitre précédent.

Dans la section suivante, nous identifions d’abord des phases qualitatives du cycle circadien dans l’attracteur asynchrone, puis nous utilisons un graphe résumé pour analyser notre modèle.

3.2 Analyse du modèle par un graphe résumé

Dans la suite, nous posons $A = (V \subset \Omega^{11}, E)$ le graphe de l’attracteur asynchrone du modèle de la figure 3.3. Nous avons $|V| = 442$ états, et $|E| = 1737$ transitions.

3.2.1 Identification de phases qualitatives du cycle circadien

Nous cherchons à identifier des étapes clefs dans le cycle circadien de façon à associer à chaque étape un sous ensemble d’états de l’attracteur. Ceci correspond à une recherche de partition dans le but d’obtenir un graphe résumé permettant de comparer la structure de l’attracteur avec les oscillations circadiennes. Pour identifier ces phases qualitatives, nous nous inspirons fortement de la description faite dans [46, 84]. Ces revues décrivent le cycle circadien en 5 grandes phases appelées respectivement *Poised*, *Derepression*, *Activation*, *Transcription* et *Repression*. Cette description tient compte essentiellement de données transcriptionnelles, notamment des gènes *Bmal1*, *Per*, *Cry*, tout au long du cycle circadien. Ces phases sont décrites dans la Fig. 3.4, et positionnées temporellement le long d’un axe *Circadian time* (CT) qui indique le temps pour un organisme diurne. Les phases *Poised* et *Derepression* se situent juste avant le début du cycle. La phase *Poised* correspond à une phase dans laquelle le complexe Clock:Bmal1 se lie à Cry1 en formant un tridimère Clock:Bmal1:Cry1. Ce tridimère réprime la transcription de *Bmal1* empêchant ainsi l’activité de Clock:Bmal1. La phase *Derepression* suit la phase *Poised* et elle correspond à une étape dans laquelle le niveau de *Cry1* décroît, favorisant le relâchement de la répression du complexe Clock:Bmal1. Ces deux premières phases se situent temporellement entre la fin de nuit et le début de journée. Après la phase *Derepression*, il s’en suit la phase d’*Activation* qui voit l’initiation de la transcription de *Bmal1*. Cette phase a lieu durant la journée. Plus tard dans la journée, la transcription des gènes *Per* et *Cry* démarre dans la phase appelée *Transcription*. Les protéines *Per* et *Cry* correspondantes, une fois suffisamment produites, entrent dans le noyau pour réprimer le complexe Clock:Bmal1 sous la forme du complexe Per:Cry. C’est la phase de *Repression*. Cette phase s’enchaîne avec une autorépression de *Per* et *Cry* entraînant par conséquent une dissociation du complexe Per:Cry.

Cette description biologique des phases du cycle circadien en 5 grandes étapes permet d’avoir une idée des différentes zones d’activation et d’inhibition des variables majeures

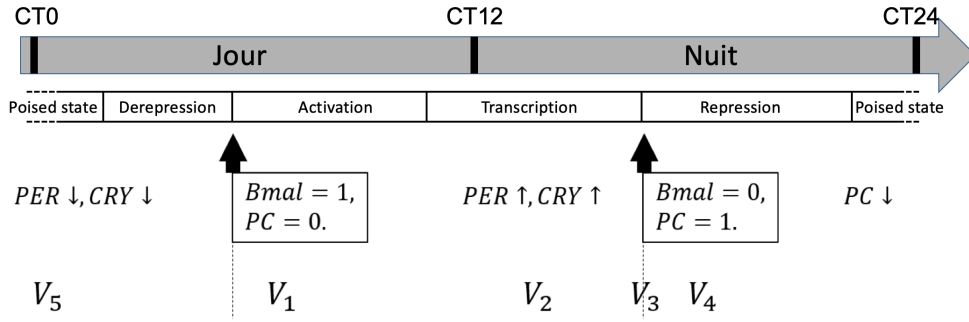


FIGURE 3.4 – Description des principales étapes qualitatives de l’horloge circadienne. CT signifie *Circadian Time* et est un marqueur standard du temps commençant arbitrairement par CT0 : le début de l’activité d’un organisme diurne. En haut : succession temporelle des 5 phases biologiques telles que décrites dans [46, 84] ; au milieu : projection des principaux événements de régulation sur les variables du modèle ; en bas : phases qualitatives correspondantes V_i (3.33).

BMAL1, *PER*, *CRY* et *PC* de notre modèle. Ainsi, l’entrée dans la phase *Activation* va correspondre, dans le modèle, au moment où la variable *BMAL1* passe à 1. Pareillement, l’entrée dans la phase *Repression* peut être associée au moment où *BMAL1* passe à 0. La phase *Activation* se passe durant la journée, donc en l’absence de Per:Cry, elle correspond à $BMAL1 = 1, PC = 0$. La phase *Repression* se passe elle durant la nuit, quand le complexe Per:Cry est actif, et donc elle correspond à $BMAL1 = 0, PC = 1$. Par définition d’un attracteur asynchrone, nous savons qu’il doit exister des étapes intermédiaires, c’est à dire pour lesquelles $BMAL1 = 1, PC = 1$ ou $BMAL1 = 0, PC = 0$. Nous allons donc considérer les valeurs prises par *PER* et *CRY* pour décrire ces étapes intermédiaires. Dans la description de [46], la phase *Transcription* est l’intermédiaire entre la phase *Activation* et la phase *Repression*. Cette phase correspondant à la transcription des gènes *Per* et *Cry*, nous l’associons aux états de l’attracteur tels que $BMAL1 = 1, PC = 0$ et $PER = CRY = 1$. C’est à dire, les deux variables *PER* et *CRY* sont à 1 mais le complexe *PC* n’est pas encore présent. Ceci peut être interprété comme la pré-formation du complexe Per:Cry dans le cytosol [31]. Cette étape est suivie de la formation du complexe dans le cytosol, c’est à dire de l’ensemble des états tels que $BMAL1 = 1, PER = 1, CRY = 1$ et $PC = 1$. Cette étape se trouve entre la fin de la journée et le début de la nuit[74] et se situe avant la translocation de Per:Cry dans le noyau et donc avant la répression de *Bmal1*. Par rapport aux phases de [46], elle correspond à un intermédiaire entre la phase *Transcription* et la phase *Repression*. Dans [46], la phase *Repression*, est suivie des phases *Poised* et *Derepression*, dans lesquelles le complexe Clock:Bmal1 est inhibé et le complexe Per:Cry s’est dissocié dû à l’absence de *Per* ou de *Cry* ou des deux. Étant donné que nous ignorons l’ordre dans lequel *Per* et *Cry* se désactivent, nous regroupons les phases *Poised* et *Derepression* par l’ensemble des états de l’attracteur tels que $BMAL1 = 0, PC = 0, PER = 0, CRY = 0$.

Ainsi, en tenant compte des variables du modèle, nous définissons 5 étapes clefs de

l'attracteur (voir aussi Fig. 3.4) :

$$\begin{aligned}
V_1 &= \{x \in V / x_{BMAL1} = 1, x_{PC} = 0, (x_{PER}, x_{CRY}) \neq (1, 1)\}, \\
V_2 &= \{x \in V / x_{BMAL1} = 1, x_{PC} = 0, x_{PER} = 1, x_{CRY} = 1\}, \\
V_3 &= \{x \in V / x_{BMAL1} = 1, x_{PC} = 1, x_{PER} = 1, x_{CRY} = 1\}, \\
V_4 &= \{x \in V / x_{BMAL1} = 0, x_{PC} = 1\}, \\
V_5 &= \{x \in V / x_{BMAL1} = 0, x_{PC} = 0, x_{PER} = 0, x_{CRY} = 0\}.
\end{aligned}$$

V_1 et V_4 sont associées respectivement à la phase *Activation* et à la phase *Repression* de [46]. L'étape V_2 se situe pendant la phase *Transcription* et l'étape V_3 correspond à la transition vers la phase *Repression*. Quant à V_5 , elle regroupe les deux phases *Poised* et *Derepression*. Les ensembles V_i ainsi définis peuvent être représentés de manière plus compacte avec les formules Booléennes suivantes :

$$\left\{ \begin{array}{ll} V_1 := (BMAL1 \wedge \overline{PC}) \wedge (\overline{PER} \vee \overline{CRY}) & \text{(journée),} \\ V_2 := (BMAL1 \wedge \overline{PC}) \wedge PER \wedge CRY & \text{(fin journée),} \\ V_3 := BMAL1 \wedge PC \wedge PER \wedge CRY & \text{(transition jour/nuit),} \\ V_4 := \overline{BMAL1} \wedge PC & \text{(nuit),} \\ V_5 := \overline{BMAL1} \wedge \overline{PC} \wedge \overline{PER} \wedge \overline{CRY} & \text{(fin nuit/début journée).} \end{array} \right. \quad (3.33)$$

Ces formules Booléennes sont une façon de caractériser les ensembles d'états V_i . Ainsi, un état de l'attracteur est classé dans V_i si sa configuration vérifie la formule Booléenne correspondante. Par exemple, la formule $V_4 = \overline{BMAL1} \wedge PC$ veut dire que tous les états de l'attracteur tels que $BMAL1 = 0$ et $PC = 1$ sont classés dans l'ensemble V_4 .

3.2.2 Construction et analyse du graphe résumé

En appliquant les formules Booléennes (3.33) aux états de l'attracteur, nous obtenons un classement dans lequel chaque phase de l'horloge circadienne est effectivement représentée dans l'attracteur : $|V_1| = 77$ états, $|V_2| = 31$ états, $|V_3| = 31$ états, $|V_4| = 136$ états et $|V_5| = 30$ états. Il reste également $|U| = 137$ états non classés. En utilisant la partition $\mathcal{P} = \{V_1, \dots, V_5, U\}$, nous construisons le graphe résumé de l'attracteur reproduit dans la figure 3.5.

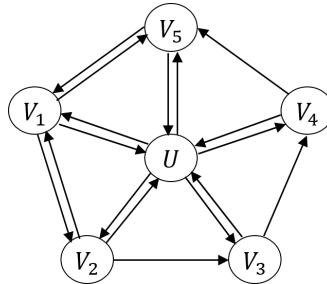


FIGURE 3.5 – Graphe résumé de l'attracteur sur la partition $\mathcal{P} = \{V_1, \dots, V_5, U\}$, les V_i étant définis par (3.33).

Le graphe résumé de la figure 3.5 montre bien l'ordre attendu des 5 étapes clés décrites dans la section précédente dans l'attracteur. Ainsi, l'activation de $BMAL1$ en V_1

est suivie de l'accumulation de *PER* et *CRY* en V_2 . Cette dernière, correspondant à la pré-formation du complexe *Per:Cry*, est suivie de la formation du complexe en V_3 . L'étape V_3 , dans laquelle *BMAL1* et *PC* sont toutes deux à 1, est suivie de V_4 qui correspond à la répression de *BMAL1* par *PC*. A la suite de cette étape, nous avons l'étape V_5 qui est un intermédiaire avant la prochaine activation de *BMAL1* et l'initiation d'un nouveau cycle. Ainsi, le cycle $V_1 \rightarrow V_2 \rightarrow V_3 \rightarrow V_4 \rightarrow V_5 \rightarrow V_1$ décrit bien qualitativement les étapes clés du cycle circadien conformément à la description de [46, 84].

Par ailleurs, le graphe résumé fait apparaître trois transitions unidirectionnelles (*i.e* sans arc retour) montrant ainsi une progression irréversible entre certaines étapes. La première, $V_2 \rightarrow V_3$, correspond à l'accumulation de *PER* et de *CRY* et à la formation du complexe *PC*. Cette transition permet de voir que, dans l'attracteur, la formation du complexe *PC* est irréversible (absence d'arc $V_3 \rightarrow V_2$). Cette irréversibilité confirme que la désactivation du complexe *PC* n'a lieu qu'après qu'il ait inhibé *BMAL1*, conformément à ce qui est évoqué dans [76]. La deuxième transition irréversible est $V_3 \rightarrow V_4$, qui correspond à la transition entre la formation du complexe *PC* et la répression de *BMAL1*. Cette transition peut être associée au moment de la translocation du complexe *Per:Cry* dans le noyau pour bloquer l'activité de *Clock:Bmal1*. La troisième transition irréversible est $V_4 \rightarrow V_5$ et correspond à la désactivation du complexe *PC* comme conséquence de l'auto-inhibition de *PER* et *CRY* [31].

Enfin, on peut constater que l'ensemble des états non classés U est fortement interconnecté avec tous les V_i . Ceci implique de nombreux contournements possibles du cycle. Ce problème sera traité plus spécifiquement dans le chapitre 4.

3.2.3 Validation des mutations à l'aide du graphe résumé

Comme nous l'avons fait dans le chapitre précédent, nous vérifions à présent le comportement de notre modèle par rapport à six mutations de gènes connues. Les trois premières, délétions de *Bmal1*, *Ror* et surexpression de *Rev-Erb α* mènent à la disparition de l'attracteur complexe et à l'apparition d'un point fixe. Quant aux délétions de *Per*, de *Cry*, *Rev-Erb α* , elles conservent un attracteur complexe, mais sous forme altérée. Les résultats, décrits ci-après, sont résumés dans la table 3.4.

Mutation	Comportement asymptotique du modèle
BMAL1=0	01001000000 $\in V_5$
REV=1	00100000000 $\in V_5$
ROR=0	00000000000 $\in V_5$
REV =0	attracteur de 80 états
PER =0	attracteur de 114 états
CRY =0	attracteur de 106 états

TABLE 3.4 – Effet de quelques mutations sur la dynamique du modèle Booléen de l'horloge.

Mutations avec arrêt du cycle circadien.

BMAL1=0. Selon [75], le mutant $Bmal1^{-/-}$ conduit à une absence totale d'oscillations circadiennes en obscurité constante. Cette perte d'oscillations se traduit dans le modèle par la disparition de l'attracteur complexe et l'apparition du point fixe (01000000000). Ce point fixe dont toutes les variables sont à 0 sauf ROR , qui est un activateur de $BMAL1$, se situe dans l'ensemble V_5 du graphe résumé. Selon notre définition de V_5 , le cycle circadien n'a donc pas démarré, confirmant le rôle clef de $BMAL1$ dans le déclenchement du cycle.

REV=1. Cette mutation mène aussi à la disparition de l'attracteur complexe et à l'apparition du point fixe (00100000000), dans lequel seul REV est à 1. Là encore, le cycle circadien s'est arrêté en V_5 . D'après [49], la surexpression de $Rev - Erba$ permet d'inhiber complètement l'activité de $Bmal1$ empêchant ainsi la transcription des autres gènes de l'horloge. Le modèle est donc en phase avec les résultats expérimentaux [49, 75, 71].

ROR=0. Là encore, l'attracteur disparaît au profit du point fixe (00000000000). Ce résultat ne correspond pas à [75], qui semble indiquer que des oscillations seraient maintenues. Néanmoins, ceci n'est pas surprenant car dans notre modèle, ROR est un activateur essentiel de $BMAL1$, via la $Rbox$.

Mutations avec oscillations dégradées : $REV = 0$, $PER = 0$ et $CRY = 0$. Les mutations $Rev - erba^{op}$, $Per2^{-/-}$, $Cry1^{-/-}$ ont la particularité de maintenir des oscillations circadiennes [49, 71] ce qui se traduit dans notre modèle par la conservation d'un attracteur complexe. Nous allons donc utiliser le graphe résumé pour comprendre plus précisément la nature de ces oscillations.

REV=0. Ce mutant admet un attracteur de 80 états dont le graphe résumé (voir Fig. 3.6) décrit bien les 5 étapes clefs du cycle circadien. Le modèle conserve donc bien des oscillations normales ce qui est en accord avec les observations dans [71]. De plus, par rapport au graphe résumé initial, les transitions où $BMAL1$ passe de 1 à 0 sont supprimées ($V_1 \rightarrow U$, $V_1 \rightarrow V_5$, $V_2 \rightarrow U$) sauf celles qui sont associées à l'arc $V_3 \rightarrow V_4$. Ceci correspond à l'accumulation $BMAL1$ observée dans ce même article, faisant le lien avec le raccourcissement de la période des oscillations.

PER=0, CRY=0. Ces mutants $PER = 0$ et $CRY = 0$ admettent un attracteur complexe de 114 et 106 états, respectivement. Les graphes résumés de ces deux attracteurs sont représentés dans la Fig. 3.6. Cette fois, on note que le cycle se limite aux ensembles V_1 , V_5 et U , indiquant des oscillations dégradées. Ceci correspond bien à ce qui est connu expérimentalement [99, 75]. Dans le chapitre 4, nous étudierons de nouveau ces deux attracteurs mutants pour comprendre le rôle des transitions associées aux arcs reliant ces trois ensembles.

3.3 Analyse temporelle des oscillations circadiennes

Nous proposons à présent une étude de l'attracteur du modèle, en utilisant notamment l'outil des chaînes de Markov introduit au chapitre 2. Comme pour le cycle cellulaire, cet outil va nous permettre de quantifier le temps passé dans les différentes phases et ainsi comparer de façon plus précise l'attracteur avec les étapes temporelles du cycle circadien (voir Fig. 3.4). Dans un second temps, nous utiliserons la chaîne de Markov pour analyser l'effet de perturbations du réseau sur le déphasage du cycle.

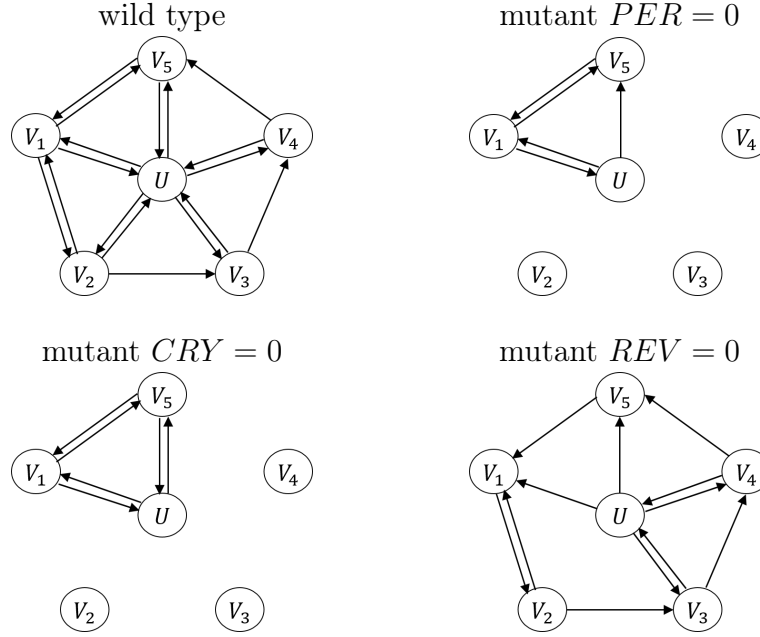


FIGURE 3.6 – Graphe résumé du modèle Booléen de l’horloge ainsi que des trois mutants $PER = 0$, $CRY = 0$ et $REV = 0$.

3.3.1 Estimation du temps passé dans les phases du cycle

Considérons le graphe résumé de la Fig. 3.5. En appliquant la formule (2.3), nous construisons la chaîne de Markov sur ce graphe résumé, dont la matrice de transition P est donnée par (3.34).

$$\begin{array}{c}
 \begin{array}{cccccc}
 & V1 & V2 & V3 & V4 & V5 & U \\
 V1 & \begin{bmatrix} 0.6710 & 0.1738 & 0 & 0 & 0.0431 & 0.1121 \end{bmatrix} \\
 V2 & \begin{bmatrix} 0.0478 & 0.5516 & 0.2613 & 0 & 0 & 0.1392 \end{bmatrix} \\
 V3 & \begin{bmatrix} 0 & 0 & 0.6258 & 0.3263 & 0 & 0.0478 \end{bmatrix} \\
 V4 & \begin{bmatrix} 0 & 0 & 0 & 0.7947 & 0.0740 & 0.1313 \end{bmatrix} \\
 V5 & \begin{bmatrix} 0.1206 & 0 & 0 & 0 & 0.7978 & 0.0817 \end{bmatrix} \\
 U & \begin{bmatrix} 0.0999 & 0.0226 & 0.0356 & 0.1016 & 0.1147 & 0.6256 \end{bmatrix}
 \end{array}
 \end{array}
 \quad . \quad (3.34)$$

Cette chaîne de Markov est bien sûr ergodique grâce au théorème 1. De plus, elle est régulière puisque les coefficients diagonaux P_{ii} de la matrice de transition sont tous strictement positifs [9]. Donc, elle admet une unique distribution stationnaire :

$$\begin{aligned}
 \omega &= (\omega_{V1}, \quad \omega_{V2}, \quad \omega_{V3}, \quad \omega_{V4}, \quad \omega_{V5}, \quad \omega_U) \\
 &= (0.1659, \quad 0.0753, \quad 0.0734, \quad 0.2249, \quad 0.2417, \quad 0.2187),
 \end{aligned}$$

que nous représentons dans la Fig. 3.7 (gauche). Rappelons que la valeur ω_{V_i} estime la proportion du temps passé dans la phase V_i . En observant le vecteur ω , on peut voir que l'écart entre le temps passé dans V_1 (*Activation*) et le temps passé dans V_4 (*Repression*) n'est pas très important, indiquant une certaine symétrie temporelle entre les deux phases, avec ω_{V_4} légèrement supérieur à ω_{V_1} . Ceci correspond à ce qui est observé dans [46]. On observe de plus une certaine similarité entre la phase de *Transcription* et la somme des deux étapes V_2 et V_3 , ainsi qu'entre les phases *Poised states* et *Derepression* et l'étape V_5 . On note également que la proportion du temps passé dans l'étape V_5 est importante (24%), surtout par rapport au nombre d'états qu'elle contient (autour de 7% des états de l'attracteur). Globalement, le vecteur ω respecte donc certaines propriétés connues de l'horloge circadienne, telles que décrites dans les revues [46, 84]. Enfin, par rapport au modèle du cycle cellulaire (chapitre 2), la proportion du temps passé dans l'ensemble des états non classés U reste importante (autour de 22%), mais tout de même inférieure à la proportion d'états de U (30%).

Dans notre analyse précédente, nous avons vu que le mutant $REV = 0$ a des oscillations qualitativement très proches du modèle wild type (voir Fig. 3.6, en bas à droite). Grâce à la chaîne de Markov, nous pouvons à présent étudier cette mutation d'un point de vue quantitatif. En calculant la distribution stationnaire de ce mutant, nous obtenons l'histogramme présenté dans la Fig. 3.7 (droite). Cette figure nous apprend que la proportion du temps passé dans V_5 a considérablement diminué au profit du temps passé dans V_1, V_2 et V_3 . Ceci veut dire que la mutation a réduit les phases *Poised* et *Derepression* qui sont situées en fin de nuit. Ceci est à mettre directement en parallèle avec ce qui est observé chez des souris $Rev^{-/-}$ dans [71], qui voient un déphasage de leur horloge circadienne en fin de cycle. De plus, on observe chez ces souris une accumulation de *Bmal1*, qui peut être rapprochée de l'augmentation des étapes V_1, V_2 , et V_3 .

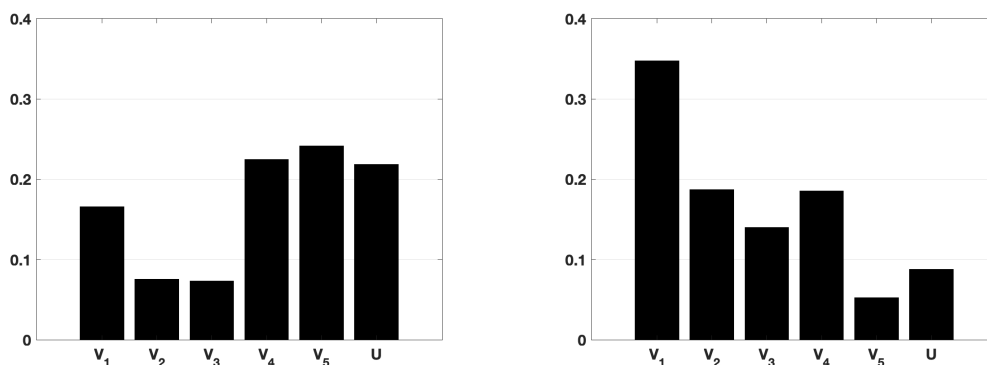


FIGURE 3.7 – Distributions stationnaires ω des chaînes de Markov du modèle wild type (gauche) et du mutant $REV = 0$ (droite).

3.3.2 Extension : étude du déphasage du cycle

Nous nous intéressons à présent à l'étude de perturbations liées au déroulement du cycle circadien, et notamment au déphasage du cycle entraînant un raccourcissement ou un allongement de la période des oscillations. L'un des déphasages les plus connus est

le syndrome d’avance de phase (FASPS, Familial Advanced Sleep Phase Syndrome). Les personnes atteintes de ce syndrome ont un cycle veille/sommeil normal mais la phase nocturne arrive plus tôt et s’interrompt plus tôt, par exemple autour de 19 h 30 et 4 h 30 [35, 93]. D’après [92], FASPS est associé à une déphosphorylation de la protéine *Per2* qui est impliquée dans le complexe Per:Cry. Par ailleurs, ce même syndrome est observé suite à la mutation de la caséine kinase *Iε* (mutation *tau*), enzyme responsable de la modification de *Per2*. De façon similaire, cette mutation diminue la vitesse de phosphorylation de la protéine *Per2*. Dans le modèle d’EDO de [4], la mutation *tau* est testée en variant le paramètre de dégradation de la variable *PER* et les résultats concluent à un changement asymétrique au niveau des phases du cycle : le temps d’activation de *BMAL1* et *CRY* restant quasiment constant alors que la phase d’activation de *PER* et celle de *PC* décroissent en fonction de la dégradation de *PER*.

Dans [92], l’effet de cette mutation est directement lié à des sites de phosphorylation de Per2, pouvant avoir des effets indirects sur la stabilisation de la protéine Per comme du complexe Per:Cry. Dans notre modèle booléen, nous décelons quatre types de transitions possibles pouvant être affectées par une phosphorylation de Per : $PER \downarrow$, $PER \uparrow$, $PC \downarrow$ et $PC \uparrow$. Nous allons donc jouer sur quatre poids ρ_{PER}^+ , ρ_{PER}^- , ρ_{PC}^+ et ρ_{PC}^- pour recalculer la matrice de la chaîne de Markov selon la formule (2.6) et ainsi déterminer quels effets ont le plus d’impact dans notre modèle. Pour cela nous posons, pour chaque distribution stationnaire ω , $\tau_J = \omega_{V_1} + \omega_{V_2} + \omega_{V_3}$ regroupant le temps passé dans les étapes de la journée et $\tau_N = \omega_{V_4} + \omega_{V_5}$ regroupant le temps passé dans les étapes de la nuit. On rappelle que $\tau_J + \tau_N = 1 - \omega_U$, où ω_U est la proportion des états passé dans les états non classés.

Nous effectuons à présent N expériences indépendantes qui consistent à choisir deux vecteurs de poids ρ^+ et ρ^- . Tous les poids sont fixés à 1, sauf les poids ρ_{PER}^+ , ρ_{PER}^- , ρ_{PC}^+ et ρ_{PC}^- sont choisis aléatoirement. Plus précisément, ces poids sont tirés au hasard parmi p valeurs réparties uniformément (sur une échelle logarithmique) entre $\rho_{min} = 10^{-3}$ et $\rho_{max} = 10^3$. Pour chacune des N expériences, nous recalculons la chaîne de Markov et sa distribution stationnaire, avant de calculer les valeurs τ_J et τ_N . Les résultats de cette simulation sont reproduits dans la Fig. 3.8 dans laquelle, nous projetons τ_J et τ_N sur ρ_{PER}^+ , ρ_{PER}^- , ρ_{PC}^+ et ρ_{PC}^- . La proportion du temps des phases de jour est en rouge, la proportion du temps des phases de nuit est en bleu et nous représentons en noir la proportion du temps passé dans les états non classés ω_U .

D’après la Fig. 3.8, nous voyons que les quatre poids ont une influence sur la durée des phases, mais que c’est clairement la dégradation du complexe *PC* qui a l’impact le plus marqué. L’impact de la dégradation du complexe a déjà été repéré dans [92] : la mutation, jouant sur la phosphorylation de Per2, est mise en relation avec une déstabilisation du complexe, correspondant à une augmentation de ρ_{PC}^- . Dans notre modèle, l’augmentation de ce paramètre conduit à une diminution des phases de “nuit” au profit des phases de “jour”, induisant un décalage dans les phases de l’horloge circadienne.

Nous voyons ici une nouvelle utilisation de la chaîne de Markov sur le graphe résumé. En faisant varier les vitesses relatives des variables Booléennes, nous pouvons repérer quels sont les processus qui ont le plus d’impact sur les durées des phases du cycle. Une

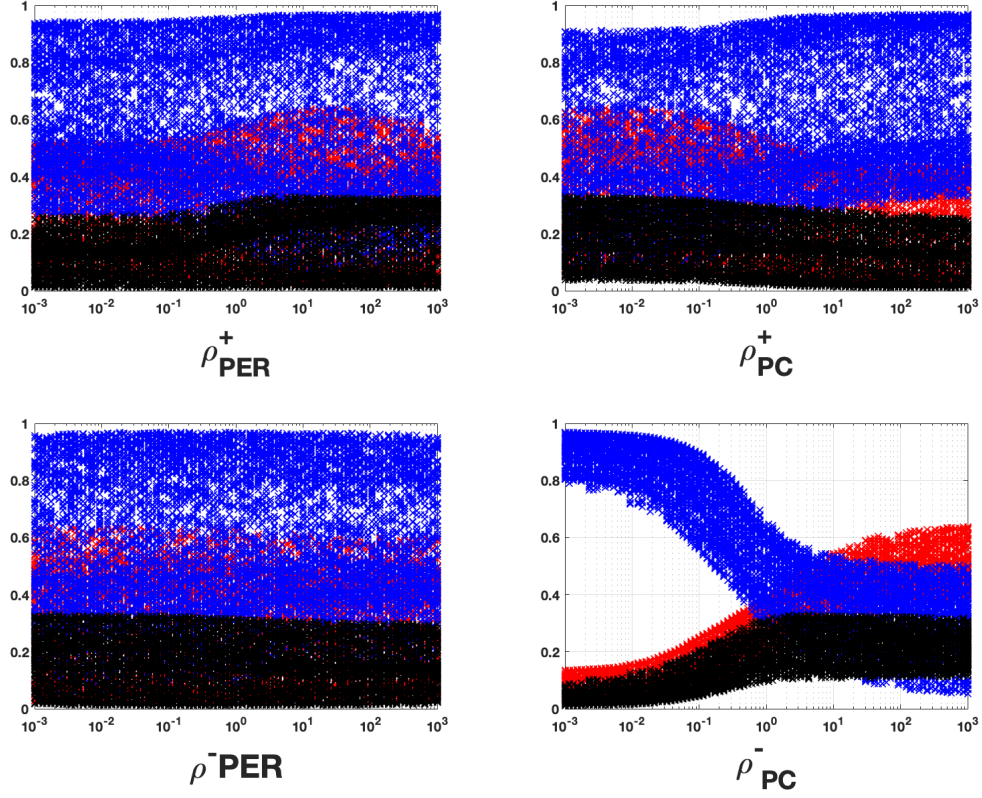


FIGURE 3.8 – Résultat de $N = 10000$ expériences indépendantes dans lesquelles les poids ρ_{PER}^+ , ρ_{PER}^- , ρ_{PC}^+ et ρ_{PC}^- sont modifiés (tirés au hasard parmi $p = 100$ valeurs, voir texte). En bleu est représenté le temps des phases de “nuit” (τ_N) et en rouge le temps des phases de jour (τ_J). Les points noirs correspondent au temps passé dans les états non classés ω_U (sachant que $\tau_J + \tau_N + \omega_U = 1$).

étude plus détaillée nous permettrait certainement d'affiner le modèle Booléen, et ainsi d'approcher au plus près les durées relatives de la figure 3.7 des phases de l'horloge.

Les chapitres 2 et 3 montrent tous les deux l'utilité du graphe résumé pour étudier les réseaux de régulation cellulaires présentant des oscillations. Que ce soit dans le cadre du cycle cellulaire ou de l'horloge circadienne, les descriptions temporelles de telles oscillations sont le plus souvent qualitatives, sous forme d'enchaînement d'événements discrets. Les modèles Booléens asynchrones sont particulièrement bien adaptés pour modéliser de tels réseaux de régulation, et l'ajout du graphe résumé dans l'arsenal du modélisateur est idéal pour capturer la notion de phase qualitative et ainsi valider un modèle Booléen présentant un ou plusieurs attracteurs complexes.

Outre l'utilisation directe du graphe résumé pour comparer un attracteur à un phénotype biologique, nous avons vu jusqu'ici deux types d'utilisation du graphe résumé : dans l'analyse de modèles mutants tout d'abord, puis dans le calcul du temps passé dans les phases à partir d'une chaîne de Markov. Ces deux types d'utilisation permettent notamment de valider un réseau Booléen. Une fois un modèle validé, nous pouvons pousser l'analyse du graphe résumé afin d'étudier certaines propriétés clés de la dynamique asymptotique d'un réseau, comme la présence de court-circuits entre certaines phases par exemple, ou encore la présence de goulots d'étranglement. Nous montrons dans le chapitre suivant comment utiliser le graphe résumé pour repérer, et éventuellement modifier localement ce type de propriétés dynamiques.

Chapitre 4

Utilisation du graphe résumé pour ajuster la dynamique d'un attracteur complexe

Nous avons vu aux chapitres précédents que la construction du graphe résumé d'un attracteur offre l'opportunité de comparer, du point de vue qualitatif, un attracteur asynchrone avec un phénotype biologique complexe, comme le cycle cellulaire ou le cycle circadien par exemple. Une fois qu'un modèle Booléen est validé du point de vue biologique, le graphe résumé nous fournit également une description simplifiée de la dynamique asymptotique du modèle. A ce titre, il est particulièrement bien adapté pour analyser le modèle en repérant par exemple les transitions de phases. Nous avons vu au chapitre 2 que chaque transition entre phases (dans le graphe résumé) est associée de façon univoque à un ensemble de transitions dans l'attracteur, nous permettant ainsi d'avoir un accès direct aux parties des trajectoires de l'attracteur qui permettent ces changements de phase. Ceci nous permet d'envisager des altérations locales de la dynamique du modèle, afin de favoriser ou au contraire d'empêcher tel ou tel changement de phase.

En particulier, dans les deux modèles précédents nous avons remarqué la présence de "court-circuits" entre certaines phases du cycle cellulaire ou du cycle circadien. La plupart du temps, ces court-circuits passent par l'ensemble U des états non classés, et à ce titre ils sont particulièrement difficiles à comparer à des trajectoires biologiques. Dans ce chapitre, nous proposons d'éliminer ces court-circuits afin de respecter l'ordre strict des phases. En combinant cela à un algorithme d'inférence Booléenne, nous montrons que nous pouvons relier ces court-circuits à certaines parties des règles du réseau. Ce lien entre la dynamique et la topologie d'un réseau Booléen était déjà exploité dans [87] pour retrouver des interactions opérationnelles dans un attracteur. Couplée avec le graphe résumé, cette méthode offre un moyen puissant pour analyser les modèles de réseaux biologiques présentant des oscillations.

Dans la partie 4.1, nous proposons d'utiliser cette méthode pour enlever les court-circuits du modèle de cycle cellulaire. Nous déduisons de cette étude une version réduite du modèle, regroupant seulement quatre variables, qui reproduit fidèlement l'enchaînement des phases du cycle (partie 4.2). Enfin, nous utilisons la méthode pour étudier le modèle du cycle circadien dans la partie 4.3, et nous montrons notamment qu'une partie des court-circuits est nécessaire à la robustesse de l'attracteur du modèle.

4.1 Analyse avancée de l'attracteur du cycle cellulaire

Dans le chapitre 2, l'analyse des graphes résumés des figures 2.7 et 2.8 révèle la présence de chemins qui “court-circuitent” le cours normal du cycle cellulaire dans l'attracteur. Nous avons vu que ces chemins passent par les états non classés U . Dans cette partie, nous proposons de supprimer ces court-circuits pour construire une version du modèle avec un attracteur décrivant le plus fidèlement possible le cycle cellulaire. Nous allons voir que la suppression totale de ces court-circuits n'est pas complètement satisfaisante, et nous proposons dans la suite un traitement séparé des court-circuits afin de comprendre leur raison d'être dans l'attracteur.

4.1.1 Vers un modèle sans court-circuits

Dans le graphe résumé de la figure 2.7 nous distinguons deux types de court-circuits, ceux qui passent par des états non classés (en pointillés) ainsi que des transitions directes de $SG2$ vers $M/G1$ (en traits plein). Nous rappelons que chaque arc $V_i \rightarrow V_j$ est associé à un ensemble de transitions de l'attracteur, noté T_{ij} . Pour supprimer tous les court-circuits, il nous suffit de supprimer dans l'attracteur les 5 transitions associées à $SG2 \rightarrow M/G1$ ainsi que les 6 transitions associées à $G1 \rightarrow U$, et les 6 transitions associées à $G1/S \rightarrow U$. Ainsi, seulement 17 transitions de l'attracteur sur 338 sont à l'origine de l'ensemble des court-circuits, et les supprimer nous permet d'obtenir un graphe résumé sans court-circuits.

Dans la suite nous étudions l'impact de cette suppression sur les règles du réseau. Afin de visualiser cet impact, nous allons reconstruire les règles du réseau à partir de l'attracteur modifié. Pour reconstruire des règles à partir d'un attracteur asynchrone, nous utilisons la même technique que celle décrite dans [87] dans le cadre de la recherche d'interactions opérationnelles. Ci-après, nous décrivons brièvement cette technique, plus de détails peuvent être trouvés dans [87].

Considérons le graphe $A = (V \subset \{0,1\}^n, E)$, sous graphe du graphe de transition asynchrone d'un réseau Booléen de dimension n . Nous supposons que l'ensemble V est un invariant positif¹, c'est à dire que les successeurs asynchrones des états de V restent dans V . En d'autres termes,

$$\forall x \in V, Succ(x) \subset V. \quad (4.1)$$

Pour reconstruire partiellement le réseau qui génère le graphe de transition asynchrone A , nous commençons par calculer, pour chaque état $x \in V$, son successeur synchrone $y = F(x)$ en utilisant l'ensemble de ses successeurs asynchrones $Succ(x)$. Nous rappelons qu'un successeur asynchrone de x est un vecteur $\tilde{x}^i = (x_1, \dots, \bar{x}_i, \dots, x_n)$, où $i \in \{1, \dots, n\}$. Introduisons l'ensemble $I(x) = \{i \in \{1, \dots, n\} / \tilde{x}^i \in Succ(x)\}$, nous avons alors :

$$\forall i \in \{1, \dots, n\}, y_i = \begin{cases} \bar{x}_i, & \text{si } i \in I(x), \\ x_i, & \text{si } i \notin I(x) \end{cases}$$

En procédant ainsi pour chaque $x \in V$, nous reconstruisons la fonction partielle, sur V , du réseau : $F|_V$. Il nous suffit alors d'utiliser un algorithme d'inférence Booléenne

1. c'est typiquement le cas si A est un attracteur

pour reconstruire des règles qui génèrent la dynamique A . Généralement, ces algorithmes sont basés sur un principe de parcimonie : ils renvoient des fonctions avec le moins de dépendances possibles (fonctions à support minimal). Dans [87], l'algorithme utilisé est l'algorithme Reveal [54]. On peut également citer l'algorithme de [60] utilisé dans [88].

Revenons à présent à l'attracteur $A = (V, E)$ du modèle du cycle cellulaire. En supprimant les 17 transitions précédemment citées, nous obtenons un graphe $A' = (V, E')$ qui est un graphe partiel de A où E' contient $338 - 17 = 321$ transitions. Comme nous n'avons fait qu'ôter des transitions à une composante fortement connexe, la propriété d'invariance (4.1) est conservée par A' . Cependant, le graphe A' n'est pas fortement connexe. Par conséquent A' ne peut être considéré comme un attracteur. Ceci nous amène à décomposer A' en composantes fortement connexe. Nous obtenons alors une SCC terminale de 53 états, notée $A'' = (V'', E'')$. En appliquant la méthode précédente sur A'' , nous obtenons une fonction $F'' : V'' \rightarrow \Omega^{10}$ sur laquelle nous appliquons l'algorithme d'inférence. Nous obtenons alors le modèle de la Table 4.1.

Variable	Logical rule
CycD	$CycD$
Rb	$(\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge \overline{CycD} \wedge \overline{CycB})$
E2F	$(\overline{Rb} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge \overline{Rb} \wedge \overline{CycB})$
CycE	$E2F \wedge \overline{Rb}$
CycA	$(E2F \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge (\overline{Cdh1} \wedge \overline{UbcH10})) \vee$ $(\overline{CycA} \wedge \overline{Rb} \wedge \overline{\mathbf{CycB}} \wedge \overline{\mathbf{Cdc20}})$
p27	$(\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB}) \vee$ $(p27 \wedge (\overline{CycE} \wedge \overline{CycA}) \wedge \overline{CycB} \wedge \overline{CycD})$
Cdc20	$CycB$
Cdh1	$(\overline{CycA} \wedge \overline{CycB}) \vee Cdc20 \vee (p27 \wedge \overline{CycB})$
UbcH10	$\overline{Cdh1} \vee (Cdh1 \wedge \overline{UbcH10} \wedge (Cdc20 \vee \overline{CycA} \vee \overline{CycB}))$
CycB	$\overline{\mathbf{E2F}} \wedge \overline{\mathbf{CycE}} \wedge \overline{Cdc20} \wedge \overline{Cdh1}$

TABLE 4.1 – Nouvelle version du modèle de [26], inférée à partir du nouvel attracteur A'' . Les différences avec le modèle original, suggérées par l'algorithme d'inférence, sont en gras.

Il est important de signaler qu'à partir de F'' , l'algorithme d'inférence nous donne un ensemble de règles telles que celles de $CycD$, Rb et $P27$ sont toutes constantes. Ceci est lié au fait que ces trois variables sont constantes dans tous les états de l'attracteur. Ainsi, pour écrire le modèle de la table 4.1, nous avons adapté les règles dépendant de ces variables, afin de mieux mettre en évidence les changements effectifs avec le modèle initial (en gras).

De façon surprenante, les règles du tableau 4.1 sont très proches du modèle initial, sauf pour celles de deux variables, CycA et CycB, pour lesquelles l'inférence suggère de nouvelles interactions, absentes du modèle initial et résumées dans la figure 4.1.

La figure 4.1 montre une inhibition de la variable CycB sur la variable CycA et une inhibition des deux variables E2F et CycE sur CycB. L'inhibition de CycB sur CycA peut être vue comme un effet indirect de la variable Cdc20 dans le modèle initial. En effet, le modèle initial considère que CycB active Cdc20 qui, à son tour, inhibe CycA.

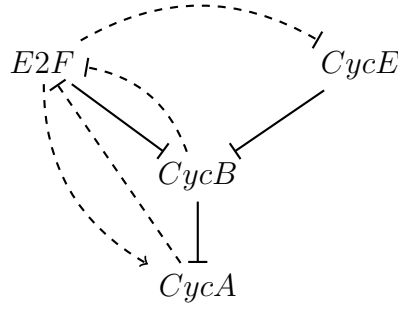


FIGURE 4.1 – Graphe des dépendances entre les variables $E2F$, $CycE$, $CycA$ et $CycB$ dans le modèle initial (pointillés); et les trois nouvelles inhibitions suggérées par l’algorithme d’inférence après suppression des 17 transitions (traits pleins).

Ces interactions entre la cycline B, $cdc20$ et la cycline A sont connues dans la littérature et elles interviennent durant la phase de transition G2/M [41]. De plus, cette inhibition de CycA par CycB était observée par Gérard et Goldbeter [32] lorsqu’ils considèrent le réseau de régulation des cyclines comme un ensemble de quatre modules centrés autour de CycD (ou E2F), CycE, CycA, et CycB. Dans [32], l’inactivation du module de CycA par le module de CycB se fait par l’intermédiaire de Cdc20. Les inhibitions de E2F et CycE sur CycB, en revanche, ne correspondent à aucune régulation connue. Cela signifie donc certainement que le fait de couper la totalité des court-circuits était sûrement une hypothèse trop forte.

Au niveau de la dynamique globale, le nouveau modèle reste proche du modèle initial puisqu’il conserve le même point fixe G_0 lorsque $CycD = 0$, en plus de l’attracteur de 53 états précédent lorsque $CycD = 1$. Le graphe résumé de cet attracteur est représenté dans la figure 4.2. En plus de la suppression de tous court-circuits, on remarque que les arcs retours $G1 \rightarrow M/G1$ et $G1/S \rightarrow G1$ ont disparu. Cette disparition peut être mise en rapport avec l’avènement du point de restriction à la fin de la mitose [81], durant la phase G1 [55] ou à la fin de la phase G1 [10, 69].

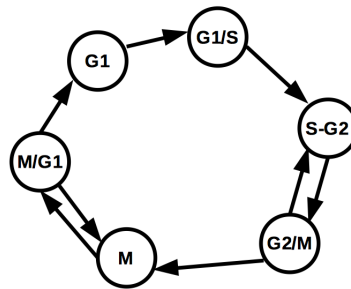


FIGURE 4.2 – Graphe résumé de l’attracteur du modèle du tableau 4.1. La partition utilisée est construite en utilisant le classement décrit dans la Table 2.2 (page 40).

Ainsi, grâce au graphe résumé, il a été possible d’ajuster localement la dynamique asymptotique du réseau. Malgré la suppression de plusieurs transitions dans l’attracteur, l’impact sur les règles du réseau reste limité ce qui semble indiquer que les transitions ont été judicieusement sélectionnées. Du point de vue biologique, la méthode fournit

ainsi des prédictions qu'on peut ensuite comparer avec les connaissances biologiques. Ces prédictions peuvent mettre en évidence de nouvelles régulations à tester ou encore des liens indirects qui peuvent être manquant dans le modèle. Cet exemple d'application montre l'intérêt de la méthode du point de vue de l'analyse d'attracteur : en combinant le graphe résumé à l'inférence, il devient possible d'ajuster localement la dynamique asymptotique d'un réseau Booléen asynchrone. Si la suppression simultanée de tous les court-circuits du graphe résumé peut sembler un peu drastique, il est également possible de procéder par étapes. Dans la partie suivante, nous étudions ces suppressions de manière séparée, afin de repérer quels sont leurs effets spécifiques sur la dynamique du modèle.

4.1.2 Suppression séparée des court-circuits dans l'attracteur

Nous allons à présent traiter chacun des court-circuits $G1 \rightarrow U$, $G1/S \rightarrow U$ et $SG2 \rightarrow M/G1$ de manière séparée. Dans chaque cas, nous appliquons le même traitement : tout d'abord la suppression des transitions correspondantes dans l'attracteur, suivie d'une éventuelle redécomposition du graphe en composantes fortement connexes, puis nous appliquons la méthode décrite dans la partie précédente pour inférer les nouvelles règles du modèle.

Nous commençons par l'arc $SG2 \rightarrow M/G1$ qui a la particularité de ne pas passer par les états non classés U . Dans les 5 transitions associées à cet arc, nous constatons que seule la variable $CycA$ bouge, et plus précisément passe de 1 à 0. Dans tous ces cas, on observe donc une désactivation prématurée de $CycA$ (elle est sensée se produire après la transition $G2/M$). Ce phénomène a déjà été observé dans [89], où il est montré, grâce à une technique de vérification formelle (CTL, Computational Tree Logic), que l'ordre d'inhibition des cyclines à la fin du cycle n'est pas respecté. En appliquant notre traitement, nous n'obtenons qu'une seule règle modifiée, qui est naturellement celle de $CycA$:

$$CycA = (\overline{\mathbf{E2F}} \wedge \overline{CycE} \wedge CycA \wedge \overline{\mathbf{CycB}}) \vee (CycA \wedge \overline{Cdh1} \wedge \overline{Ubc}) \vee (E2F \wedge \overline{Cdh1} \wedge \overline{Ubc}) \\ \vee (CycA \wedge \overline{Cdh1} \wedge \overline{Cdc20}) \vee (E2F \wedge \overline{Cdh1} \wedge \overline{Cdc20}). \quad (4.2)$$

En ne s'intéressant qu'aux dépendances entre les variables, nous remarquons que seules deux régulations se sont rajoutées par rapport au modèle initial :

$$\mathbf{E2F} \multimap CycA \vdash \mathbf{CycB}$$

Comme évoqué plus haut, l'inhibition de $CycA$ par $CycB$ est déjà connue, de façon indirecte [32]. Quant à l'inhibition de $E2F$, elle est plus discutable, mais peut être liée à la complexité de la règle ($E2F$ a déjà un effet positif sur $CycA$ au sein de la même règle).

Considérons à présent les arcs $G1 \rightarrow U$ et $G1/S \rightarrow U$. Ces arcs correspondent tous à une activation de $CycB$ dès la phase $G1$. Les transitions de $G1 \rightarrow U$ montrent même une activation de $CycB$ avant $CycA$. La cycline B étant un marqueur de la phase M , ces court-circuits révèlent, donc là encore un ordre d'activation des cyclines erroné. La suppression de ces court-circuits impacte la règle de la variable $CycB$.

— Lorsque l'on supprime les 6 transitions de $G1 \rightarrow U$, on obtient :

$$CycB' = [(\overline{\mathbf{E2F}} \wedge \overline{\mathbf{CycE}}) \vee \mathbf{CycA}] \wedge (\overline{Cdc20} \wedge \overline{Cdh1}); \quad (4.3)$$

— Lorsque l'on supprime les 6 transitions de $G1/S \rightarrow U$, on obtient :

$$CycB' = [(\overline{\mathbf{E2F}} \wedge \overline{\mathbf{CycE}}) \vee \overline{\mathbf{CycA}}] \wedge (\overline{Cdc20} \wedge \overline{Cdh1}); \quad (4.4)$$

— Lorsque l'on supprime les 12 transitions ensemble, on obtient :

$$CycB' = (\overline{\mathbf{E2F}} \wedge \overline{\mathbf{CycE}}) \wedge (\overline{Cdc20} \wedge \overline{Cdh1}). \quad (4.5)$$

Toutes les modifications suggérées sont représentées graphiquement par :

$$\begin{array}{ccc} \mathbf{E2F} \multimap CycB \vdash \mathbf{CycE} & \mathbf{E2F} \multimap CycB \vdash \mathbf{CycE} & \\ \uparrow & \top & \\ \mathbf{CycA} & \mathbf{CycA} & \mathbf{E2F} \multimap CycB \vdash \mathbf{CycE} \end{array}$$

Ces nouvelles interactions prédites par notre méthode n'ont pas d'explication biologique claire. Ceci signifie que la présence des transitions associées à $G1 \rightarrow U$ et $G1/S \rightarrow U$ dans l'attracteur semble inévitable, en tout cas dans la version actuelle du modèle. Comme elles correspondent à une activation anticipée de la cycline B, cela peut signifier que la règle de cette cycline est à revoir, notamment en ce qui concerne son terme d'activation. Une deuxième explication possible, qui n'est bien sûr pas à exclure, est que la partition que nous avons utilisée pour construire le graphe résumé doit être encore raffinée, certains états de U pouvant être mal classés.

Dans tous les cas, nous voyons bien ici l'intérêt de notre méthode pour analyser les attracteurs complexes de réseaux Booléens. En simplifiant un attracteur contenant des centaines d'états et de transitions, elle permet notamment de visualiser simplement la dynamique sous-jacente de l'attracteur, afin de mieux la comparer avec le phénomène biologique. Elle permet également de cibler, dans le modèle, des potentiels court-circuits entre des phases et de remonter aux règles directement responsables de ces court-circuits. Par cet aller-retour entre le comportement dynamique de l'attracteur et les règles du réseau de régulation, notre méthode fournit ainsi un outil particulièrement efficace dans le design de modèles Booléens de réseaux biologiques.

4.2 Vers un modèle simplifié du cycle cellulaire

Durant la construction du graphe résumé de la figure 2.7, nous avons été amenés à classer 76 états de l'attracteur (sur 112) selon des considérations biologiques sur les phases du cycle. Ce faisant, nous remarquons que seulement quatre variables $E2F$, $CycE$, $CycA$ et $CycB$ sont nécessaires pour classer ces états. Ainsi, la construction du graphe résumé nous a permis de repérer que ces variables jouent un rôle central dans la description qualitative des oscillations. Nous décidons à présent de construire un modèle Booléen simple, autour de ces quatre variables, reproduisant à l'identique le graphe résumé et donc la succession des phases du cycle. La construction de ce modèle réduit se fait en deux étapes principales. Tout d'abord, nous projetons les états de l'attracteur sur les 4 variables, et nous construisons les transitions entre ces états projetés. Ensuite, nous utilisons la méthode décrite dans la section 4.1.1 pour obtenir les règles de ce nouveau modèle.

4.2.1 Projection des états et construction d'un graphe de transition réduit

Pour chacun des 112 états $x = (*, *, E2F, CycE, CycA, *, *, *, *, CycB) \in \Omega^{10}$ de l'attracteur du modèle de [26], nous considérons l'état $\tilde{x} = (E2F, CycE, CycA, CycB) \in \Omega^4$ qui est la projection de x sur les quatre variables $E2F$, $CycE$, $CycA$ et $CycB$. Nous obtenons alors la table 4.2.

$E2F$	$CycE$	$CycA$	$CycB$	ensembles
0	0	0	0	$M/\tilde{G}1$
0	0	0	1	$\tilde{M}$
0	0	1	0	$S\tilde{G}2$
0	0	1	1	$G2/\tilde{M}$
0	1	0	0	$\tilde{G}1$
0	1	0	1	$\tilde{U}$
0	1	1	0	$G1/\tilde{S}$
0	1	1	1	$\tilde{U}$
1	0	0	0	$\tilde{G}1$
1	0	0	1	$\tilde{U}$
1	0	1	0	$G1/\tilde{S}$
1	0	1	1	$\tilde{U}$
1	1	0	0	$\tilde{G}1$
1	1	0	1	$\tilde{U}$
1	1	1	0	$G1/\tilde{S}$
1	1	1	1	$\tilde{U}$

TABLE 4.2 – Les 16 états projetés de l'attracteur de [26] sur les quatre variables $E2F$, $CycE$, $CycA$ et $CycB$.

La partition $\mathcal{P}$ se projette alors naturellement sur la partition $\tilde{\mathcal{P}}$ suivante :

$$\left\{ \begin{array}{l} \tilde{G}1 = \{1000, 1100, 0100\}, \\ G1/\tilde{S} = \{1110, 1010, 0110\}, \\ S\tilde{G}2 = \{0010\}, \\ G2/\tilde{M} = \{0011\}, \\ \tilde{M} = \{0001\}, \\ M/\tilde{G}1 = \{0000\}, \\ \tilde{U} = \{0101, 0111, 1001, 1011, 1101, 1111\}. \end{array} \right.$$

Les sous ensembles $\tilde{G}1$, $G1/\tilde{S}$, $S\tilde{G}2$, $G2/\tilde{M}$, $\tilde{M}$, $M/\tilde{G}1$ correspondent donc à une description simple des six phases biologiques du cycle cellulaire. L'ensemble $\tilde{U}$ correspond bien, quant lui, à des situations non biologiques.

Nous allons à présent construire un graphe de transition sur ces 16 états projetés. Pour cela, considérons une transition $x \rightarrow y$ de l'attracteur initial. Comme nous considérons

une stratégie asynchrone, x et y ne diffèrent que sur une variable. S'ils diffèrent sur l'une des quatre variables $E2F$, $CycE$, $CycA$ ou $CycB$, nous ajoutons la transition $\tilde{x} \rightarrow \tilde{y}$. Nous obtenons le graphe de la figure 4.3 qui est fortement connexe.

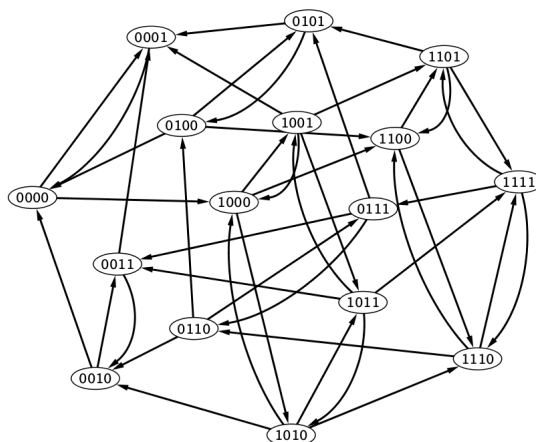


FIGURE 4.3 – Projection de l'attracteur initial sur les quatre variables $E2F$, $CycE$, $CycA$ et $CycB$: 16 états et 44 transitions asynchrones.

En utilisant la méthode de [87] citée précédemment, nous pouvons alors reconstruire les règles de ce nouveau réseau ainsi que le graphe des dépendances entre les variables, présentés dans la Fig. 4.4. Nous retrouvons des interactions du réseau initial, ainsi que deux nouvelles interactions, l'autoinhibition de $CycA$ et de $CycB$, qui étaient présentes uniquement de façon indirecte.

4.2.2 Analyse du modèle réduit

Analysons à présent le modèle décrit dans la figure 4.4. En utilisant la partition $\tilde{P}$ définie plus haut, nous construisons le graphe résumé de l'attracteur de la figure 4.3, présenté dans la figure 4.5.

Ce graphe correspond exactement au graphe résumé du modèle initial (Fig. 2.7),

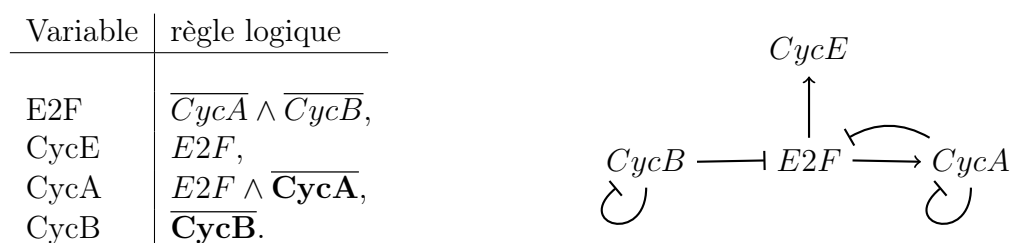


FIGURE 4.4 – Gauche : Règles logiques du modèle reconstruites à partir du graphe de la figure 4.3. Droite : graphe d'interaction correspondant, les seules interactions nouvelles (par rapport au modèle initial) sont les autoinhibitions de $CycA$ et de $CycB$

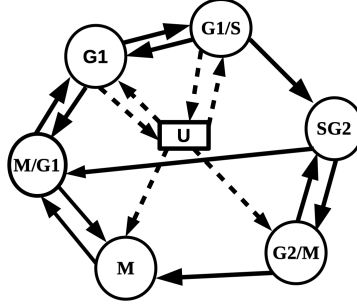


FIGURE 4.5 – Graphe résumé du modèle réduit sur la partition $\tilde{P}$.

confirmant ainsi que le modèle réduit capture bien les oscillations qualitatives du modèle initial. Afin de montrer le rapprochement entre le modèle réduit et le modèle initial, nous effectuons le traitement précédent consistant à couper les court-circuits. Le court-circuit $SG2 \rightarrow M/G1$ consiste en une transition qui, si elle est enlevée, mène à la nouvelle règle

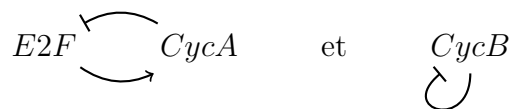
$$CycA' = (\overline{E2F} \wedge \overline{CycE} \wedge CycA \wedge \overline{CycB}) \vee (E2F \wedge \overline{CycA}).$$

Les court-circuits $G1 \rightarrow U$ et $G1/S \rightarrow U$, quant à eux, consistent en six transitions qui, une fois enlevées, conduisent à la nouvelle règle

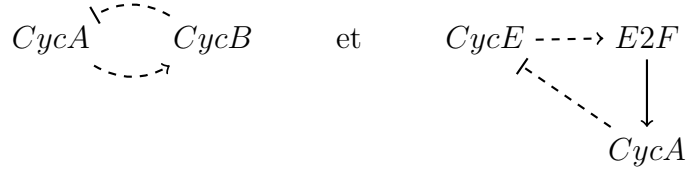
$$CycB' = \overline{E2F} \wedge \overline{CycE} \wedge \overline{CycB}.$$

Ces deux modifications de règles correspondent à ce qui a été trouvé dans la partie précédente sur le modèle initial. Néanmoins, le modèle réduit ne comportant que les variables $E2F$, $CycE$, $CycA$ et $CycB$, les suggestions de règles sont beaucoup plus simples et nous permettent de repérer directement, sur ces quatre variables, les effets (directs ou indirects) générant ces court-circuits dans le modèle initial. A ce titre, notre modèle réduit est particulièrement utile pour décrire le “cœur” des oscillations du réseau de régulation, et ainsi repérer plus facilement les modifications éventuelles à apporter pour ajuster au mieux les oscillations qualitatives du modèle.

Un autre intérêt du modèle réduit est d’offrir la possibilité de comparer plus facilement un modèle Booléen à un système biologique oscillatoire. Dans le cas du cycle cellulaire, nous avons isolé un cœur de quatre variables, trois cyclines et leur facteur de transcription, permettant de reproduire l’enchaînement des phases du cycle. Dans [32], par une approche complètement différente (basée sur des EDOs), Gérard et Goldbeter proposent une décomposition du cycle cellulaire en quatre modules très proches. Le premier module est centré sur le facteur de transcription $E2F$, le deuxième sur la cycline E et les deux derniers, commandant la division mitotique, sont centrés sur les cyclines A et B . Cet article recense les quatre boucles de rétroaction négatives principales entre ces variables. Nous retrouvons bien deux d’entre elles :



En revanche, les deux autres sont incomplètes (en pointillés) :



Certaines interactions sont manquantes dans le modèle initial (comme l’inhibition de *CycE* par *CycA*) et d’autres sont bien présentes, mais de manière indirecte. Cette observation nous indique des pistes à explorer dans le modèle réduit, afin de modifier le modèle global. Ceci confirme l’utilité du modèle réduit comme modèle “jouet” pour analyser le réseau de régulation des cyclines.

4.3 Utilisation du graphe résumé comme aide à la modélisation, application au réseau du cycle circadien

Dans le chapitre 3, nous avons construit un modèle Booléen de l’horloge circadienne en nous basant sur un modèle EDO existant [4]. Afin de valider ce modèle, nous avons notamment construit une partition de l’attracteur en cinq étapes qualitatives V_i du cycle (voir Fig. 3.4), 137 états de l’attracteur restant dans l’ensemble U des états non-classés. Nous avons ensuite construit le graphe résumé de la figure 3.5 (page 63), montrant que l’attracteur reproduit bien l’enchaînement des étapes, dans l’ordre voulu. Néanmoins, l’ensemble U reste hyperconnecté à toutes les étapes V_i , autorisant ainsi de multiples court-circuits entre les phases du cycle. Pour compléter cette validation, nous avons enfin étudié l’effet d’un petit nombre de mutations en comparant l’attracteur avec les résultats expérimentaux.

A présent, nous approfondissons l’analyse de notre modèle en explorant, comme pour le modèle du cycle cellulaire, les court-circuits possibles entre les phases. Pour cela, nous commençons par identifier plus précisément les états de U (non-classés biologiquement). Nous verrons qu’une partie de ces états peut être aisément supprimée de l’attracteur, suggérant des modifications de règles tout-à-fait pertinentes. Comme pour le modèle du cycle cellulaire, nous verrons également qu’une autre partie de U est plus difficilement supprimable, et qu’elle peut même être liée à la robustesse de l’attracteur, notamment dans les modèles des mutants.

4.3.1 Elimination de certains états non classés

Nous rappelons que $A = (V, E)$ désigne l’attracteur du modèle du chapitre 3, et que $U := V \setminus (\bigcup_i V_i) \subseteq \{0, 1\}^{11}$ est l’ensemble des 137 états non pris en compte dans notre description des phases de l’horloge circadienne (voir section 3.2.1). Afin de manipuler plus aisément cet ensemble U , nous allons le caractériser par une formule Booléenne. Pour cela, en affectant la valeur 1 aux états de U et la valeur 0 aux états de $V \setminus U$, nous construisons une fonction Booléenne partielle (définie sur l’ensemble $V \subset \{0, 1\}^{11}$), sur laquelle nous pouvons appliquer un algorithme d’inférence Booléenne (comme [54] évoqué

dans la partie 4.1.1). Nous obtenons la formule suivante :

$$U := \underbrace{\overline{BMAL} \wedge \overline{PC} \wedge (PER \vee CRY)}_{U_1} \vee \underbrace{BMAL1 \wedge PC \wedge \overline{PER}}_{U_2}, \quad (4.6)$$

que nous notons encore U par abus de notation. Dans cette formule disjonctive, on peut remarquer deux parties distinctes U_1 et U_2 caractérisant deux sous ensembles d'états disjoints et recouvrant U . L'ensemble U_1 contient 106 états, dans lesquels les variables $BMAL1$ et PC sont à 0, alors que U_2 contient 31 états, dans lesquels $BMAL1$ et PC sont à 1.

En considérant explicitement les deux ensembles U_1 et U_2 , le graphe résumé de la figure 3.5 est légèrement modifié : voir figure 4.6 (gauche). Dans ce graphe, on peut voir que l'arc $V_3 \rightarrow U_2$ est le seul lien connectant l'ensemble U_2 vers les autres ensembles V_i . En d'autres termes, supprimer les 6 transitions de l'attracteur associées à cet arc est suffisant pour déconnecter complètement U_2 . Ces transitions voient toutes la variable PER passer de 1 à 0 alors que $BMAL1$ et PC sont à 1, indiquant une inhibition de PER ayant lieu avant la répression de $BMAL1$ par PC (donc après V_4) (voir [31, 76]). Le graphe résumé nous montre donc que l'entrée en U_2 trahit le déroulement normal du cycle circadien. Pour ces raisons, nous avons supprimé les 6 transitions qui sont associées à cet arc et nous appliquons le traitement que nous avons décrit plus haut. Ce faisant, nous obtenons une nouvelle SCC terminale de 393 états, ne contenant plus aucun des états de U_2 . L'inférence de règles Booléennes sur ce nouvel attracteur voit une seule règle modifiée, celle de PER qui est donnée par l'équation (4.7).

$$PER' = (BMAL1 \wedge \overline{CRY2}) \vee (DBP \wedge \overline{E4}) \vee (\mathbf{BMAL1} \wedge \mathbf{PC}). \quad (4.7)$$

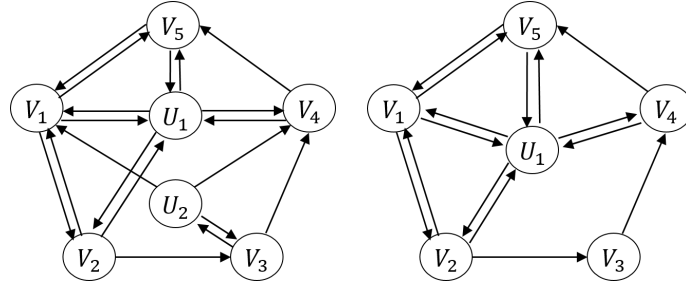


FIGURE 4.6 – Gauche : graphe résumé du modèle du cycle circadien avec U_1 et U_2 ; droite : nouveau graphe résumé après suppression des six transitions de l'arc $V_3 \rightarrow U_2$.

En observant la nouvelle règle (4.7), on peut remarquer que l'activation de PER par PC est la seule interaction qui s'ajoute (en gras), par rapport à la règle initiale. De façon intéressante, cette influence positive de PC sur PER pouvait déjà se voir dans le modèle d'EDO qui nous a servi de base pour construire le modèle, (3.1) à (3.8). L'influence provenait du terme de dissociation du complexe Per:Cry. Comme nous l'avons évoqué dans le chapitre 3, les termes dérivant de la loi d'action de masse sont souvent ignorés lors de la construction d'un modèle Booléen, car ils ne correspondent pas nécessairement

à des régulations biologiques explicites. Ici, la suppression des six transitions $V_3 \rightarrow U_2$ nous suggère une façon d'intégrer ce terme dans la règle de *PER*, en accord avec le bon déroulement du cycle. Bien qu'il ne s'agisse pas d'une méthode générale pour traduire tous les termes de loi d'action de masse dans un modèle Booléen, il est intéressant de noter que le graphe résumé peut fournir une aide précieuse lors de la conception de réseaux Booléens, en suggérant des modifications de règles permettant de raffiner la dynamique asymptotique du modèle.

4.3.2 Analyse des court-circuits du graphe résumé

Nous considérons désormais le modèle obtenu dans la section précédente, avec la nouvelle règle (4.7) et le graphe résumé de la figure 4.6 (droite). Dans ce graphe, le reste des états non classés U_1 est encore très connecté aux autres ensembles. Tenter de déconnecter U_1 du reste des états de l'attracteur revient à supprimer 117 transitions dans l'attracteur, correspondant aux quatre arcs $V_1, V_2, V_4, V_5 \rightarrow U_1$. Si nous faisons cela, nous obtenons un nouvel attracteur de 297 états qui ne contient plus que les étapes V_i du cycle. Cependant, ce nouveau modèle perd une propriété dynamique importante du modèle initial : les mutants $PER = 0$ et $CRY = 0$ ne présentent plus d'oscillations, et ont à la place un point fixe. Or, expérimentalement ces deux mutants sont capables d'osciller, bien que les oscillations soient dégradées (période plus courte). Ceci semble indiquer que, parmi ces 117 transitions dans l'attracteur permettant de relier U_1 , au moins une partie d'entre elles est nécessaire pour assurer un comportement adéquat du modèle. Comme pour le modèle du cycle cellulaire, nous proposons d'étudier les suppressions des court-circuits séparément, afin de mieux comprendre le rôle que chacun joue dans le maintien des oscillations.

Suppression de l'arc $V_4 \rightarrow U_1$. Cet arc est associé à 63 transitions dans l'attracteur, dans lesquelles la variable PC passe de 1 à 0 alors que PER et CRY ne sont pas toutes les deux à 0. D'un point de vue biologique, ceci correspond donc à une dégradation du complexe avant la disparition de PER et CRY . Ainsi, les transitions associées à $V_4 \rightarrow U_1$ peuvent être interprétées comme un temps de demi-vie court pour le complexe Per:Cry. En guise de comparaison, les 30 transitions associées à l'arc $V_4 \rightarrow V_5$ voient la variable PC passer de 1 à 0 après que PER et CRY aient totalement disparues. Ainsi, lorsque le modèle est dans la phase V_4 , deux types de trajectoires coexistent dans l'attracteur, selon que le complexe disparaît avant ou après Per et Cry. N'ayant pas trouvé d'information précise quant au temps de demi-vie du complexe (surtout par rapport à celui de PER et CRY), il nous est difficile de trancher. Cependant, il est intéressant de noter que la suppression des 30 transitions associées à $V_4 \rightarrow V_5$ conduit à une perte des oscillations du modèle (avec obtention d'un point fixe à la place de l'attracteur complexe) alors que la suppression des 63 transitions associées à l'arc $V_4 \rightarrow U_1$ n'affecte en rien la dynamique globale du modèle (l'attracteur est conservé).

Après suppression des transitions $V_4 \rightarrow U_1$, nous appliquons le traitement précédent et nous obtenons une nouvelle version du modèle avec un attracteur de 387 états décrivant bien les étapes clés du cycle circadien. Par rapport au modèle initiale, c'est la règle de PC qui est modifiée :

$$PC = (CRY \wedge PER) \vee \mathbf{PC} \wedge (\mathbf{PER} \vee \mathbf{CRY}), \quad (4.8)$$

ajoutant simplement, du point de vue du diagramme d'influence, une auto-régulation de PC . Forcer cette auto-activation de PC permet d'imposer la désactivation du complexe après la disparition de PER et CRY . Là encore, les modifications pointent vers les termes de formation et de dissociation du complexe $Per:Cry$, impliquant la loi d'action de masse dans le modèle EDO. Une meilleure connaissance quantitative des paramètres de ces termes nous permettraient ici de prendre une décision. Dans tous les cas, notre méthode nous permet de montrer que la suppression des 63 transitions $V_4 \longrightarrow U_1$ n'a pas d'impact majeur sur la dynamique sous-jacente.

Suppression de l'arc $V_5 \longrightarrow U_1$. Cet arc correspond à 9 transitions de l'attracteur, dans lesquelles PER se réactive (passe de 0 à 1) avant la complétion du cycle, donc en l'absence de $BMAL1$. En supprimant ces transitions, nous obtenons un modèle avec un attracteur de 385 états dont les propriétés dynamiques sont conservées. Cependant, cette suppression entraîne la modification suivante de la règle de PER :

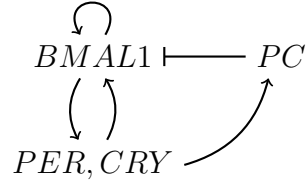
$$PER' = BMAL1 \wedge (\overline{CRY2} \vee PC) \vee [(DBP \wedge \overline{E4}) \wedge (PER \vee CRY \vee PC)] . \quad (4.9)$$

Dans (4.9), on note une auto-activation de PER ainsi qu'une activation de PER par CRY qui sont deux interactions *a priori* non documentées. Comme dans les cas précédents, la suppression de ces transitions n'a pas d'impact majeur sur la dynamique sous-jacente de l'attracteur. Même si cette modification semble plus difficile à justifier que les précédentes d'un point de vue biologique, il est intéressant de noter que les règles impactées sont à chaque fois PER et PC , indiquant que cette partie du réseau (complexation de Per et Cry) est sûrement à investiguer en premier, afin de mieux représenter le phénomène. En ce sens, notre méthode fournit donc une aide précieuse lors de la conception, ou de la correction d'un modèle.

Suppression des arcs $V_1 \longrightarrow U_1$ et $V_2 \longrightarrow U_1$. Ces deux arcs regroupent au total 45 transitions de l'attracteur, dans lesquels la variable $BMAL1$ disparaît dans la première partie du cycle. Dans les transitions associées à $V_1 \longrightarrow U_1$, $BMAL1$ s'éteint dès le début du cycle alors que dans $V_2 \longrightarrow U_1$, $BMAL1$ s'éteint au moment où PER et CRY se sont déjà accumulées. Ces transitions sont donc directement en désaccord avec ce qui est connu du régulateur $Clock:Bmal1$. En supprimant ces 45 transitions, on obtient encore un attracteur complexe, contenant 376 états. Cependant, si cet attracteur semble suivre les étapes clés du cycle circadien, les mutants $PER = 0$ et $CRY = 0$, perdent leurs oscillations. Ainsi, même si ces transitions ne sont pas correctes d'un point de vue biologique, leur suppression fait perdre au modèle une propriété clé de la dynamique de l'attracteur. Cette observation est confirmée lorsque nous inférons le modèle après la suppression des transitions, la règle modifiée est celle de $Bmal$:

$$BMAL1' = (ROR \wedge \overline{REV} \wedge \overline{PC}) \vee [(BMAL1 \wedge (PER \vee CRY) \wedge \overline{PC})] . \quad (4.10)$$

Au-delà de l'auto-activation de $BMAL1$, on note un effet positif de PER et de CRY sur $BMAL1$. Ces nouvelles interactions conduisent à générer deux boucles positives impliquant $BMAL1$, PER et CRY :



La présence de ces deux boucles vient directement interférer avec la boucle négative impliquant les complexes Clock:Bmal1 et Per:Cry.

Les analyses proposées dans ce chapitre ne sont bien sûr qu'un premier pas, mais elles permettent déjà d'envisager une technique tout à fait nouvelle pour l'analyse de réseaux Booléens. En combinant le graphe résumé, qui condense les dynamiques complexes d'un attracteur asynchrone, et l'inférence Booléenne, nous disposons d'un outil puissant pour analyser des comportements dynamiques complexes. En particulier, en pouvant remonter aisément aux règles du réseau, nous pouvons repérer plus facilement, dans un réseau de régulation biologique, quelle interaction ou quel groupe d'interactions est responsable de telle ou telle transition de phase. La déduction d'un modèle réduit de quatre variables clefs qui reproduit l'ensemble des transitions de phases constitue un bon exemple d'application, mais c'est loin d'être le seul. Pour le modèle de l'horloge, nous sommes notamment capables de proposer certaines modifications de règles directement responsables de certaines transitions. Si certaines sont directement explicables (loi d'action de masse), d'autres le sont moins et pointent ainsi vers des zones du réseau de régulation qui sont à retravailler. En ce sens, notre outil fournit une aide précieuse au processus de modélisation.

Conclusion et perspectives

Dans cette thèse, nous nous sommes intéressés à l'analyse dynamique des oscillations circadiennes et du cycle de division cellulaire par une approche de modélisation basée sur les réseaux Booléens asynchrones. Les réseaux Booléens asynchrones représentent une classe de systèmes dynamiques bien adaptée pour modéliser qualitativement des oscillateurs biologiques. Néanmoins, les oscillations dans un tel modèle sont capturées par un attracteur complexe qui est généralement difficile à appréhender, et notamment à comparer avec des phénotypes biologiques.

Dans ce cadre, nous avons proposé une nouvelle méthode d'analyse des attracteurs complexes, basée sur la construction d'un *graphe résumé* qui permet de comparer les trajectoires de l'attracteur avec les données biologiques. Pour construire un tel graphe, notre méthode s'appuie sur une étape stratégique qui est la recherche d'une partition. Nous avons montré avec nos deux exemples comment construire une partition adaptée aux informations biologiques disponibles. Cette phase nous a permis de confirmer que notre outil est un bon moyen pour montrer comment un attracteur complexe peut approcher un cycle biologique et ainsi aider à valider un modèle.

Combiné avec la théorie des chaînes de Markov, l'outil graphe résumé donne la possibilité d'ajouter de l'information au modèle Booléen pour faire une analyse semi-quantitative et d'estimer les proportions de temps dans les phases décrites par les oscillations. Lorsqu'il est couplé à une technique d'inférence Booléenne, il permet aussi d'ajuster localement et de façon précise la dynamique asymptotique d'un réseau Booléen. Ces exemples d'utilisation illustrent bien que l'outil graphe résumé est un moyen puissant à la disposition du modélisateur pour comprendre la structure interne d'un attracteur complexe, afin de faire une analyse plus avancée de ses propriétés dynamiques.

Ce travail génère plusieurs pistes de recherche intéressantes, notamment dans la modélisation de réseaux biologiques. En particulier, dans le chapitre 4, nous avons montré que quelques 45 transitions, dans lesquelles la variable *BMAL1* passe de 1 à 0, sont nécessaires au maintien de la robustesse de l'attracteur du modèle de l'horloge. En constatant que ces transitions disparaissent également de l'attracteur dans le mutant $REV = 0$, nous pensons que la présence de *REV* pourrait avoir un lien étroit avec ces transitions. Cette piste laisse penser au rôle de la boucle négative impliquant *REV* et *Bmal* qui garantit la robustesse des oscillations [74]. Cet exemple illustre la possibilité de se servir du graphe résumé pour préciser le rôle des boucles de rétroaction du graphe d'interaction, en les reliant à la présence de certaines transitions particulières dans l'attracteur.

D'un point de vue plus théorique, nous pouvons citer la construction d'un modèle Booléen réduit à l'aide d'un graphe résumé comme fait au chapitre 4. Un tel modèle réduit capture le cœur des oscillations du modèle initial, en se focalisant uniquement sur la succession des phases. Ce modèle peut servir de base pour une analyse plus fine des

oscillations, en construisant par exemple un système différentiel “minimal” à partir de ce modèle réduit. De manière générale, le graphe résumé constitue un outil intéressant pour chercher des propriétés qualitatives de la dynamique, spécifiques au cadre Booléen, afin de les transférer vers un formalisme plus continu.

Annexe A

Articles parus pendant la thèse

Nous joignons ci-après deux articles parus pendant la thèse :

Diop O., Tournier L., Fromion V. (2019) *Summarizing complex asynchronous Boolean attractors, application to the analysis of a mammalian cell cycle model*. 18th European Control Conference (ECC), IEEE, June 2019, Naples, Italy.

Diop, O., Chaves M., Tournier L. (2020) *Qualitative Analysis of Mammalian Circadian Oscillations : Cycle Dynamics and Robustness*. In International Conference on Computational Methods in Systems Biology (CMSB), p. 176-192. Springer, Cham.

Summarizing complex asynchronous Boolean attractors, application to the analysis of a mammalian cell cycle model

Ousmane Diop, Laurent Tournier and Vincent Fromion

Abstract—Biological regulatory networks can be efficiently modeled by asynchronous Boolean networks. In such networks, asymptotic dynamical behaviors are captured by asynchronous attractors, which are strongly connected graphs that may include multiple intertwined cycles. If it can be easily computed, comparing such a complex graph with a biological phenomenon may become rapidly cumbersome. In this paper, we propose to use the concept of summary graph to help the analysis of asynchronous attractors, by projecting them on partitions of the state space. The method is applied to a Boolean model of the mammalian cell cycle, for which we propose a suitable partition based on the cycle's phases. Thanks to our approach, we were able to isolate the expected succession of the phases.

Index Terms—Asynchronous Boolean networks; Biological regulatory networks; Cell cycle.

I. INTRODUCTION

Boolean networks are attracting much attention as models of biological systems, whether in a deterministic [4] or in a probabilistic context [13]. Their discrete nature is particularly well adapted to model gene regulation or signal transduction systems, where 0/1 variables catch the qualitative nature of such biochemical interaction networks. Asynchronous networks are particularly well adapted, as they can handle the large range of temporal timescales arising in biological systems [4], [19], [20]. From a graph theoretical point of view however, asynchronicity generates a lot of transitions and gives rise to complex transition graphs, with a high number of intertwined dynamical paths. In particular, an asynchronous attractor of a hundred states may be easy to compute (it is simply a terminal strongly connected component of the transition graph) but it can be quite hard to study and to compare to biological experiments, or more generally to any kind of biological knowledge about the phenomenon it is supposed to model. In this paper, we investigate some tools to represent a complex attractor, by summarizing it along a given partition of the state space.

When analyzing a model of a biological oscillator, as for instance the eukaryotic cell division cycle, such a *summarized* description of the attractor becomes particularly useful when comparing it with the sequence of events that is observed biologically. Regarding the cell cycle for instance, the overall progress is achieved through four successive phases: G1, S, G2 and M [6]. On the other hand, extensive studies have isolated the main regulators of the cycle (called cyclins) together with their main interactions. Several

dynamical models exist in the literature, as [7], [12] for yeast, [16] for plant or [8] for mammalian (see also [9] for a comparative study). In this paper, we study the Boolean model proposed in [8], which is a 10-dimensional network exhibiting a complex attractor of 112 states.

The rest of this paper is organized as follows. In Part II, we recall some preliminaries on asynchronous Boolean networks, focusing mainly on their attractors. In Part III, we introduce the summary graph, a new tool to analyze complex asynchronous attractors. In Part IV, we apply this tool to analyze the cell cycle model in [8], in particular to compare its attractor with the expected succession of cell cycle phases. Finally, in Part V we extend this analysis by proposing pointed modifications in the model to adjust the attractor in a prescribed way.

II. PRELIMINARIES: ASYNCHRONOUS BOOLEAN NETWORKS AND THEIR ATTRACTORS

We briefly recall the definitions of asynchronous Boolean networks (ABN) and asynchronous attractors. More details can be found for instance in [19], [20] or [4]. An n -dimensional Boolean network is characterized by its (exponential) state space $\Omega = \{0, 1\}^n$ and its transition function $F : \Omega \rightarrow \Omega$ (the coordinates f_i of the map F are sometimes called the logical rules of the network). Applying the asynchronous strategy consists in updating only one variable at a time. More precisely, consider a state $x \in \Omega$ and let $y = F(x)$; then for each $i \in \{1, \dots, n\}$ such that $y_i \neq x_i$ the state $x^{(i)} = (x_1, \dots, y_i, \dots, x_n)$ is an asynchronous successor of x and one draws an arc from x to $x^{(i)}$. By applying this strategy to every state $x \in \Omega$ one constructs the *asynchronous transition graph* that contains the asynchronous dynamics of the network. A first step to analyze such a graph is to compute its strongly connected components (SCC). Since the SCC graph is acyclic, it can be topologically sorted, isolating *terminal* SCC that cannot be escaped. Such terminal SCC are called *asynchronous attractors* of the network (see [19], [20], [4] for more information).

Example 1: To illustrate previous definitions, consider the four dimensional network given by the logical rules:

$$\begin{aligned} f_1(x_1, x_2, x_3, x_4) &= (x_1 \wedge x_3 \wedge x_4) \vee (\bar{x}_3 \wedge \bar{x}_4) \vee (\bar{x}_1 \wedge \bar{x}_2), \\ f_2(x_1, x_2, x_3, x_4) &= \bar{x}_4 \vee x_3 \vee x_1, \\ f_3(x_1, x_2, x_3, x_4) &= x_2 \wedge \bar{x}_4, \\ f_4(x_1, x_2, x_3, x_4) &= (x_1 \wedge \bar{x}_2 \wedge x_4) \vee x_3. \end{aligned}$$

We use the following notation: $\vee$ and $\wedge$ denote respectively *or* and *and* operators and $\bar{(\cdot)}$ denote the *not* operator. The transition graph of this network has $2^4 = 16$ vertices and only

*This work is supported by the ANR (French agency for research) through project ICycle ANR-16-CE33-0016-01
MaIAGE, INRA, University Paris-Saclay, 78350 Jouy-en-Josas, France.
ousmane.diop@inra.fr

4 SCCs. It has one attractor, which is the 12-state terminal SCC depicted in Figure 1.

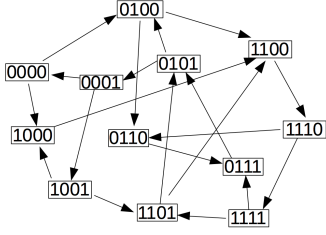


Fig. 1. Asynchronous attractor of the 4-d network of Example 1. This attractor is a terminal SCC with 12 states.

Since terminal SCC cannot be escaped, asynchronous attractors are important from a modeling point of view as they capture the network's asymptotic dynamical behaviors, that can then be compared to the phenomenon modeled by the network. The detection of attractors is therefore an essential step in the analysis of a Boolean model. Nevertheless, once an attractor has been computed it may comprise numerous intertwined cycles rendering its interpretation quite challenging. In particular, comparing such a SCC with biological experiments can be difficult. Thus, even for a network as simple as the previous one, the attractor in Fig. 1 is already relatively complex. Since it is a strongly connected component, we know that each state is reachable from any other state. In the following, we introduce some tools to go beyond and provide more practical, *summarized* descriptions of the behavior of the system in an attractor. For example, one may ask whether there exist underlying oscillations respecting a specific order between different regions of the attractor.

III. SUMMARY GRAPH OF AN ATTRACTOR

A. Summary graph of an attractor on a full partition

Let $G = (V, E)$ denote an asynchronous attractor, i.e. a strongly connected graph with $V \subseteq \Omega = \{0, 1\}^n$. Let $k \in \mathbb{N} \setminus \{0\}$ and $\mathcal{P} = \{V_1, V_2, \dots, V_k\}$ a k -partition of V , i.e. a collection of k subsets of V verifying:

- (P1) $\forall i = 1, \dots, k: V_i \neq \emptyset$,
- (P2) $\forall i, j = 1, \dots, k, i \neq j: V_i \cap V_j = \emptyset$,
- (P3) $\cup_{i=1}^k V_i = V$.

Note that a collection verifying only (P1) and (P2) is called a (k) -semi-partition of V . The idea is to use partition $\mathcal{P}$ to summarize the attractor G into a graph of size k as follows.

Definition 1 (Summary graph): The summary graph of G on partition $\mathcal{P} = \{V_1, V_2, \dots, V_k\}$ is the graph $G(\mathcal{P})$ with vertices V_i , $i = 1, \dots, k$ such that there exists an edge from V_i to V_j iff $i \neq j$ and there exists an edge in G from a state in V_i to a state in V_j .

Remark 1: Note that the same symbol V_i is used to designate the set of states V_i and the vertex V_i of $G(\mathcal{P})$. This classic abuse of notation is practical when using summary graph $G(\mathcal{P})$ to describe sets of trajectories in graph G .

Definition 1 has already been introduced (see e.g. [23]) in the more general context of graph summarization. From a

graph theoretical point of view, it corresponds to the classical notion of quotient graph [11]. From a dynamical point of view, the application of this notion to asynchronous Boolean networks provides practical ways to capture global dynamical properties of complex attractors.

As an illustration, consider again the 4d network defined in Example 1 and let $k = 4$. The idea is to summarize the 12 states attractor of Fig. 1 into a simple four-vertex graph. Introduce the two partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ as follows:

$$\mathcal{P}_1 \begin{cases} V_1 = \{0000, 1000, 0100, 1100\} \\ V_2 = \{1110, 0110, 1111\} \\ V_3 = \{1101, 0111\} \\ V_4 = \{0101, 0001, 1001\} \end{cases} \quad \mathcal{P}_2 \begin{cases} V_1 = \{0000, 0100\} \\ V_2 = \{1110, 1100, 0110, 1000, 0111\} \\ V_3 = \{1101, 1111, 0001\} \\ V_4 = \{1001, 0101\} \end{cases}$$

The two summary graphs, corresponding to partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ are depicted in Fig. 2 (respectively top left and right). As can be seen in this example, the choice of the partition may lead to drastically different summary graphs. Partition $\mathcal{P}_1$ seems to capture a relatively simple dynamical behavior that resembles the cycle $V_1 \rightarrow V_2 \rightarrow V_3 \rightarrow V_4 \rightarrow V_1$ (except for the transitions $V_3 \rightarrow V_1$ and $V_4 \rightarrow V_3$), whereas partition $\mathcal{P}_2$ generates a more complicated summary graph that is farther from simple oscillations. To explore this striking contrast, we decided to compute the summary graphs for all possible 4-partitions. The number of k -partitions of a set of size N is given by the Stirling number of the second kind σ_N^k , which grows very rapidly (see for instance [5]). In our case the number amounts to $\sigma_{12}^4 = 611501$ which remains manageable. To compare the summary graphs, we choose to use the number of edges, denoted by m , as a criterion. The minimal value obtained for m is 4, the summary graph is then a simple cycle (Fig. 2, bottom left). Such a simplification is extremely rare, as it is obtained for only three partitions. These three partitions are very close to one another: $\mathcal{P}_1^{\min} = \{V_1, V_2, V_3, V_4\}$, $\mathcal{P}_2^{\min} = \{V'_1, V_2, V_3, V_4\}$ and $\mathcal{P}_3^{\min} = \{V''_1, V_2, V_3, V_4\}$, with $V_2 = \{1110, 0110\}$, $V_3 = \{0111, 1111\}$ and

$$\begin{cases} V_1 = \{0100, 1100\}, & V_4 = \{0000, 1001, 0101, 1101, 1000, 0001\}, \\ V'_1 = \{0000, 0100, 1100, 1000\}, & V'_4 = \{1001, 0101, 1101, 0001\}, \\ V''_1 = \{0100, 1100, 1000\}, & V''_4 = \{0000, 1001, 0101, 1101, 0001\}. \end{cases}$$

Interestingly, even though they are close, none of them is finer than the other since $V_1 \subset V''_1 \subset V'_1$ and $V'_4 \subset V''_4 \subset V_4$. On the other hand, the maximal obtained value for m is 12, where the summary graph is the complete graph with 4 vertices (Fig. 2, bottom right). This configuration is less rare as it appears 5351 times, which is still relatively low compared to the total number of partitions.

As shown by this brief study, the problem of finding a “good” partition, that provides an efficient summarization of an asynchronous attractor can be brought back to a discrete optimization problem. Nevertheless, for bigger attractors the number of partitions rapidly explodes and such an exhaustive exploration becomes out of reach. In the rest of the paper, we propose to use the concept of summary graph to analyze the attractor of a higher dimensional Boolean model of a biological system. To cope with unavoidable combinatorial

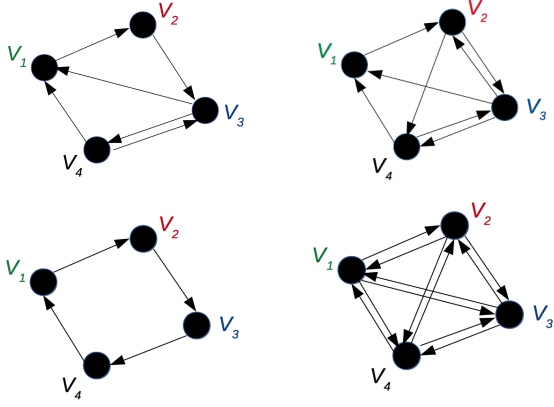


Fig. 2. Different summary graphs of the attractor of Fig. 1. Top left: partition $\mathcal{P}_1$, top right: partition $\mathcal{P}_2$, bottom: minimal (left) and maximal (right) summary graphs.

explosion, we notably show how to use biological information on the modeled phenomenon to help design efficient (and useful) partitions to summarize the attractor. Before describing the method, we propose in the following a slight generalization of the summary graph to semi-partitions, which will prove useful in a biological context.

B. Summary graph of an attractor on a semi-partition

In order to define a summary graph of an attractor $G = (V, E)$ one has to provide a partition $\mathcal{P}$ of V , i.e. a full classification of all the states in V . Sometimes it might be preferable to provide only a semi-partition of V , leaving some states unclassified. It is particularly important when considering a biological model, as some states in V may be difficult to interpret from a biological point of view and therefore hard to characterize.

Let $\mathcal{S} = \{V_1, V_2, \dots, V_k\}$ be a k -semi-partition of V , i.e. a collection of subsets of V verifying only properties (P1) and (P2) above. To generalize the notion of summary graph, the first possibility is to introduce the set

$$U := V \setminus \left(\bigcup_{i=1}^k V_i \right)$$

of unclassified states and to complete $\mathcal{S}$ into the partition $\mathcal{P}(\mathcal{S}) = \{V_1, \dots, V_k, U\}$, and then to compute the summary graph $G(\mathcal{P}(\mathcal{S}))$ of the completed partition using Definition 1. If one does not want to make the set U explicit, another possibility is to introduce the following definition.

Definition 2 (Direct path): Given a graph $G = (V, E)$ and a semi-partition $\mathcal{S} = \{V_1, V_2, \dots, V_k\}$ of V , there exists a direct path from V_i to V_j iff there exists a path in G from some $x \in V_i$ to some $y \in V_j$ that passes only through V_i , V_j and U the set of unclassified states.

Now we can generalize the definition of summary graph on semi-partitions.

Definition 3 (Summary graph): The summary graph of G on semi-partition $\mathcal{S} = \{V_1, V_2, \dots, V_k\}$ is the graph $G(\mathcal{S})$

with vertices V_i , $i = 1, \dots, k$ such that there exists an edge from V_i to V_j iff there exists a direct path from V_i to V_j . While relatively similarly defined, the two graphs $G(\mathcal{P}(\mathcal{S}))$ and $G(\mathcal{S})$ are different, as illustrated in Figure 3 for the semi-partition $\mathcal{S} = \{V_1, V_2, V_3, V_4\}$ with $V_1 = \{0000, 1101, 0111\}$, $V_2 = \{0100, 0101\}$, $V_3 = \{1100, 1110, 0110\}$ and $V_4 = \{1001, 1111\}$.

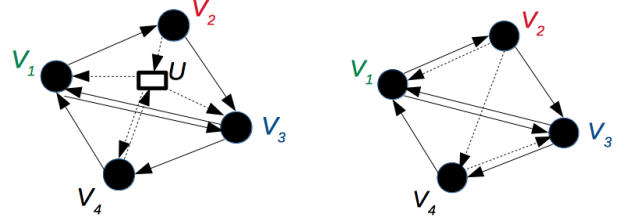


Fig. 3. Summary graphs on a semi-partition: on the left, the graph $G(\mathcal{P}(\mathcal{S}))$ (summary graph on the completed partition); on the right, the graph $G(\mathcal{S})$ (as defined by Def. 3). In both graphs dashed arrows indicate paths that cross the unclassified states U .

The main difference between the two graphs lies in the treatment of unclassified states. In $G(\mathcal{P}(\mathcal{S}))$ they appear explicitly in a class of their own, which makes easier the direct comparison with actual paths in the attractor. On the other hand, by making them implicit the graph $G(\mathcal{S})$ is easier to compare to the (biological) phenomenon that is described by the attractor. In the following part, we propose an application to a Boolean model of the mammalian cell cycle and we show how each summary graph brings specific useful information.

IV. APPLICATION TO THE ANALYSIS OF A BOOLEAN MODEL OF THE MAMMALIAN CELL CYCLE

To illustrate the concept of summary graph, we use a Boolean model of the eukaryote cell cycle originally introduced in [8]. The model has 10 variables representing principal regulators of the cycle and it includes all major known interactions between these regulators. In particular, it contains four cyclins $CycX$, $X \in \{A, B, D, E\}$ which are proteins that are essential for the cycle's progress. For completeness, the logical rules of the model are reproduced in Table I, for more details the reader is referred to [8], [9] and [21].

In this model, $CycD$ plays the role of a (constant) input: it models a growth factor whose presence is supposed to trigger the cycle. Therefore, the transition graph admits two separate regions: when $CycD = 0$, there is a unique attractor which is a single steady state representing the quiescent state of the cell; when $CycD = 1$, there is another attractor which is a complex terminal SCC comprising 112 states (out of $2^{10} = 1024$) and 338 transitions. This attractor, supposed to represent the actual cell cycle is already analyzed and discussed in [8]. Here, we propose to expand the analysis of this attractor with the help of summary graphs. First, we describe how to use biological information to construct a suitable semi-

TABLE I
LOGICAL RULES OF THE CELL CYCLE MODEL IN [8].

Variable	Logical rule
CycD	$CycD$
Rb	$(\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge \overline{CycD} \wedge \overline{CycB})$
E2F	$(\overline{Rb} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge \overline{Rb} \wedge \overline{CycB})$
CycE	$E2F \wedge \overline{Rb}$
CycA	$(E2F \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge (\overline{Cdh1} \wedge \overline{UbcH10})) \vee (CycA \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge (\overline{Cdh1} \wedge \overline{UbcH10}))$
p27	$(\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge (\overline{CycE} \wedge \overline{CycA}) \wedge \overline{CycB} \wedge \overline{CycD})$
Cdc20	$CycB$
Cdh1	$(\overline{CycA} \wedge \overline{CycB}) \vee Cdc20 \vee (p27 \wedge \overline{CycB})$
UbcH10	$\overline{Cdh1} \vee (Cdh1 \wedge \overline{UbcH10} \wedge (Cdc20 \vee CycA \vee CycB))$
CycB	$\overline{Cdc20} \wedge \overline{Cdh1}$

partition; then we show how such a classification of states help the interpretation of the attractor.

A. Constructing a semi-partition of the attractor

In order to construct a suitable semi-partition, the first step consists in choosing the number k of subsets. A logical choice in our case would be $k = 4$, as the cell cycle is generally described as the sequence of four distinct phases (see e.g. [6]): G1 (growth and preparation of chromosomal duplication), S (chromosomal duplication), G2 (growth and preparation of division) and M (mitosis or cell division). The next step consists in determining the states of the attractor in each phase, thanks to biological information on the variables of the model. Thus, we know that during G1 two types of cyclins operate: CycD and CycE [18], [10]. CycD being constant in the attractor we focus on CycE, which is produced by the indirect activation of transcription factor E2F. Since it has been observed that CycE peaks at the end of G1 just before the transition to S [17], we decide to split G1 into two subsets: G1 and G1/S, the former characterizing the bulk of G1 phase and the latter characterizing the end of G1 and the transition to S. In summary in G1, E2F is “on” ($E2F = 1$), while CycE is growing (it can therefore be 0 or 1, denoted by $CycE = *$); in G1/S, E2F can be “on” or “off” ($E2F = *$) while CycE is “on” ($CycE = 1$). This information is recorded into Table II below. The next master regulator is CycA, which is known to be “on” in the end of G1 until G2 [6], [9], succeeded by CycB in the end of G2 and in M [15]. Therefore, we decide to group S and G2 into a single phase SG2, and to decompose the M phase into G2/M (transition from the end of G2 to mitosis) and M. Finally, [22] indicates that Cdh1 turns on in the end of mitosis and remains on in the G1 phase of the next cycle (phase M/G1 in Table II).

Table II defines 24 states (by replacing the $*$ with 0 and 1), which all belong to the attractor. Therefore, thanks to biological information we are able to classify 24 out of the 112 states into six biologically interpretable phases (or sub-phases) of the cell cycle. Unfortunately a lot of states remain unclassified (88 out of 112), implying a low-quality summary graph with a lot of spurious transitions (not shown). To circumvent this issue and refine the semi-partition, we restrict ourselves to four rows in Table II, corresponding to

TABLE II
BUILDING A FIRST SEMI-PARTITION.

Variables	G1	G1/S	SG2	G2/M	M	M/G1
CycD	1	1	1	1	1	1
Rb	0	0	0	0	0	0
E2F	1	*	0	0	0	0
CycE	*	1	0	0	0	0
CycA	0	1	1	1	0	0
P27	0	0	0	0	0	0
Cdc20	0	0	0	0	*	*
Cdh1	1	*	0	0	0	1
UbcH10	*	*	*	*	*	*
CycB	0	0	0	1	1	0

Note: 0/1 means that the variable is supposed to be off/on in the phase or that the variable is forced to be constant by the model, * means that both values are acceptable, or that information is lacking.

four master regulators of the cell cycle that are present in the model: E2F, CycE, CycA and CycB. We then interpret each 4-dimensional Boolean vector as one of the 6 phases defined above, leading to the truth table in Table III (left). By construction, it is ensured that the subsets are mutually exclusive. From this truth table, simple Boolean identification leads to the six rules in Table III (right). This new semi-partition, while still satisfying on a biological level, has the advantage of classifying 76 out of the 112 states in the attractor: $|G1| = 24$ states, $|G1/S| = 24$ states, $|SG2| = 8$ states, $|G2/M| = 6$ states, $|M| = 6$ states and $|M/G1| = 8$ states. It is denoted by $\mathcal{S}$ in the rest of the paper.

TABLE III
REFINEMENT OF THE SEMI-PARTITION.

0000	M/G1	$\left\{ \begin{array}{lcl} G1 & = & (E2F \vee CycE) \wedge \overline{CycA} \wedge \overline{CycB} \\ G1/S & = & (E2F \vee CycE) \wedge CycA \wedge \overline{CycB} \\ SG2 & = & CycA \wedge \overline{E2F} \wedge \overline{CycE} \wedge \overline{CycB} \\ G2/M & = & CycA \wedge CycB \wedge \overline{CycE} \wedge \overline{E2F} \\ M & = & CycB \wedge \overline{E2F} \wedge \overline{CycE} \wedge \overline{CycA} \\ M/G1 & = & \overline{CycB} \wedge \overline{E2F} \wedge \overline{CycE} \wedge \overline{CycA} \end{array} \right.$
0001	M	
0010	SG2	
0011	G2/M	
0100	G1	
0101	$\emptyset$	
0110	G1/S	
0111	$\emptyset$	
1000	G1	
1001	$\emptyset$	
1010	G1/S	
1011	$\emptyset$	
1100	G1	
1101	$\emptyset$	
1110	G1/S	
1111	$\emptyset$	

Left: truth table of the six phases with respect to the four master regulators (E2F, CycE, CycA, CycB). Right: corresponding Boolean rules.

B. Analysis of the attractor via a summary graph

The summary graph $G(\mathcal{S})$ constructed from semi-partition $\mathcal{S}$ is depicted in Fig. 4 (top left). Thanks to this summary graph, it becomes straightforward to check that the attractor of the model actually exhibits the expected succession of the phases of the cycle: $G1 \rightarrow G1/S \rightarrow SG2 \rightarrow G2/M \rightarrow M \rightarrow M/G1 \rightarrow G1$. Moreover, we checked that for each subset there exist paths joining entry points to exit points, ensuring the existence of closed trajectories (between states) that respect the correct succession of phases. Note

that some (but not all) of the elementary transitions mentioned above also exist in the reverse order, distinguishing “pausing”, reversible transitions and irreversible transitions. Interestingly, this property of the model may be viewed in relation with the concept of check points evoked for instance in [1], [14], [15]. Check points consist of observable states during the cycle where cells are known to verify a number of constraints before irreversibly transitioning to the next phase. For instance, a cell will not enter DNA replication phase unless some conditions are verified, for instance a sufficient nucleocytoplasmic ratio [1], [2]. In the summary graph, the transition $G1/S \rightarrow SG2$ is one-directional, reproducing this irreversibility. A second known check point takes place before the entry in mitosis [3], which again seems to correspond to irreversible transition $G2/M \rightarrow M$. However, a third check point is known to occur in the end of mitosis, which is not reproduced in the summary graph. Beyond the transitions respecting the cycle’s order, the summary graph also points to spurious transitions, indicating shortcuts or unrealistic paths between the phases. Such spurious paths are frequent in asynchronous networks, especially in these dimensions, and their existence was already detected in [8] and [21]. One interest of the summary graph is to rapidly identify these paths, providing directions to review the model. In this regard, an extension of the method is proposed in the next part, where the model is amended so as to suppress these paths.

Remark 2: It is interesting to note that the classification and the construction of the summary graph are based solely on the states of the attractor and not on the whole 2^n state space. It is particularly important for larger networks, where the method is still applicable provided the studied attractor remains of reasonable size.

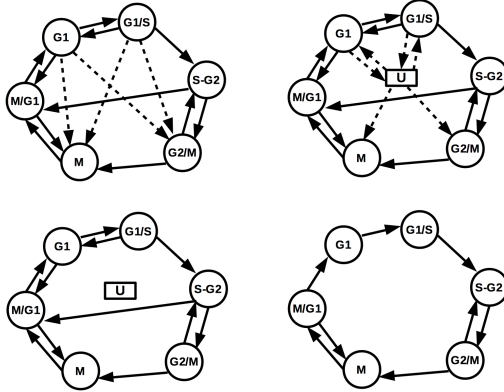


Fig. 4. *Top left:* Summary graph $G(\mathcal{S})$ of the attractor of [8] with semi-partition $\mathcal{S} = \{G1, G1/S, SG2, G2/M, M, M/G1\}$ defined in Table III. A dashed arrows indicates all paths pass through at least an unclassified state while a solid arrow indicates there exists a direct transition in the attractor. *Top right:* summary graph $G(\mathcal{P}(\mathcal{S}))$ obtained from the completed semi-partition $\mathcal{S}$. *Bottom left:* resulting graph when in/out-transitions of U have been removed. *Bottom right:* Final summary graph $H(\mathcal{S})$ obtained after the suppression of unwanted transitions.

V. MODEL AMENDMENT BASED ON SUMMARY GRAPH

We propose to modify the original model of [8] in order to suppress, in the attractor, undesired paths exhibited in the summary graph and thus to obtain a graph that is closer to the expected cell cycle. Beyond the modification of the model itself, the idea is to show how to use the summary graph to suggest minimal modifications in a network to adjust some asymptotic behavior in a prescribed way.

According to Fig. 4 (top left) there are five edges to suppress: $G1$ to M and $G2/M$, $G1/S$ to M and $G2/M$ and $SG2$ to $M/G1$. Among these edges four of them are dashed, which means that they represent sets of trajectories passing through the set of unclassified states, denoted by U . Therefore, as a first step we compute the summary graph $G(\mathcal{P}(\mathcal{S}))$ of the partition $\mathcal{P}(\mathcal{S})$ obtained by completion of $\mathcal{S}$ (Fig. 4, top right) in order to make U explicit. Since U is composed of states that are ambiguous (for instance because of the simultaneous presence of markers of different phases), we decide to simply remove all in- (respectively out-) transitions to (resp. from) U in the original attractor, leading to the new summary graph in Fig. 4, bottom left. We complete this step by removing the direct transitions from $SG2$ to $M/G1$.

Thanks to the summary graph, we were able to directly modify the original attractor G into a new graph G' , by suppressing unwanted transitions. This new graph G' can then be analyzed by the graph theoretical tools mentioned above: SCC decomposition and topological sort. We obtain a new terminal strongly connected component, denoted by H , containing only 53 states, which are all classified in the semi-partition $\mathcal{S}$. The final summary graph $H(\mathcal{S})$ of this new attractor is depicted in Fig. 4, bottom right. With respect to the original summary graph, it keeps the expected succession of the cell cycle’s phases and is cleared of spurious transitions.

Lastly, we propose an analysis of the above modifications in terms of the original model’s logical rules (Table I). To do so we use the method developed in [19] that uncovers the part of a Boolean network that is operational in a given attractor. It is based on Boolean identification, and it mainly consists in the reconstruction of *minimal* logical rules sufficient to reproduce an attractor (the reader is referred to [19] for more details). By applying this method successively on the original attractor G and on the modified version H , the same operational rules are inferred, except for two of them: the rules for CycA and CycB. For attractor G , these rules read:

$$\begin{cases} CycA' = (E2F \vee CycA) \wedge \overline{Cdc20} \wedge \overline{Cdh1} \wedge \overline{UbcH10}, \\ CycB' = \overline{Cdc20} \wedge \overline{Cdh1}, \end{cases}$$

whereas for attractor H they read:

$$\begin{cases} CycA' = E2F \wedge \overline{Cdc20} \wedge \overline{Cdh1} \wedge \overline{UbcH10} \vee \\ \quad CycA \wedge \overline{Cdc20} \wedge \overline{CycB}, \\ CycB' = \overline{E2F} \wedge \overline{CycE} \wedge \overline{Cdc20} \wedge \overline{Cdh1}. \end{cases}$$

Inspired by these rules, we propose the amended version of the model in Table IV. With respect to the original version, there is a new negative influence of CycB on CycA, and

two additional inhibitions, of E2F and CycE on CycB. The new version conserves the good properties of the initial model: a unique quiescent steady state when CycD=0 and a unique complex attractor when CycD=1. But this time, the latter attractor is exactly the graph H , respecting the exact succession of the phases of the cell cycle with no shortcut (Fig. 4, bottom right).

It is to be noted the proposed modifications are not unique; however, this amended version is strikingly close to the original version suggesting that the semi-partition $\mathcal{S}$ captures well the phases of the original attractor.

TABLE IV
MODIFIED VERSION OF THE MODEL IN [8].

Variable	Logical rule
CycD	CycD
Rb	$(\overline{\text{CycD}} \wedge \overline{\text{CycE}} \wedge \overline{\text{CycA}} \wedge \overline{\text{CycB}}) \vee (p27 \wedge \overline{\text{CycD}} \wedge \overline{\text{CycB}})$
E2F	$(\overline{\text{Rb}} \wedge \overline{\text{CycA}} \wedge \overline{\text{CycB}}) \vee (p27 \wedge \overline{\text{Rb}} \wedge \overline{\text{CycB}})$
CycE	$\text{E2F} \wedge \overline{\text{Rb}}$
CycA	$(\text{E2F} \wedge \overline{\text{Rb}} \wedge \overline{\text{Cdc20}} \wedge (\overline{\text{Cdh1}} \wedge \overline{\text{UbcH10}})) \vee (\text{CycA} \wedge \overline{\text{Rb}} \wedge \text{CycB} \wedge \text{Cdc20})$
p27	$(\overline{\text{CycD}} \wedge \overline{\text{CycE}} \wedge \overline{\text{CycA}} \wedge \overline{\text{CycB}}) \vee (p27 \wedge (\overline{\text{CycE}} \wedge \text{CycA}) \wedge \overline{\text{CycB}} \wedge \overline{\text{CycD}})$
Cdc20	CycB
Cdh1	$(\overline{\text{CycA}} \wedge \overline{\text{CycB}}) \vee \text{Cdc20} \vee (p27 \wedge \overline{\text{CycB}})$
UbcH10	$\overline{\text{Cdh1}} \vee (\text{Cdh1} \wedge \overline{\text{UbcH10}} \wedge (\text{Cdc20} \vee \text{CycA} \vee \text{CycB}))$
CycB	$\overline{\text{E2F}} \wedge \overline{\text{CycE}} \wedge \overline{\text{Cdc20}} \wedge \overline{\text{Cdh1}}$

Modifications with respect to the original rules in Table I appear in bold.

VI. CONCLUSION AND PROSPECTS

A method to represent and analyze a complex attractor of an asynchronous Boolean network has been proposed. The method is based on the construction of the summary graph of the attractor on a given partition or semi-partition of the state space. In terms of interpretation, the summary graph can be a powerful tool to visualize and understand the underlying characteristics of the global dynamics of the attractor. Different partitions may lead to different summary graphs and therefore bring different levels of details in the analysis of the attractor. Two general approaches may be used to find a suitable partition. First, one can search for a “simple” representation in a graph theoretical sense, thus reducing the problem to the optimization of some discrete criterion. Such optimization problem can become quite computationally heavy for large attractors. A second possibility, illustrated on the cell cycle example, is to use external knowledge about the phenomenon modeled by the network. Once such a partition has been found, we have shown how to use the summary graph to compare the attractor with the expected succession of the cell cycle’s phases, and even propose pointed modification of the network in order to adjust its behavior in a prescribed way.

The design of adapted partitions remains a central point of the proposed method, and algorithms to help design suitable partitions will be investigated in the future. For large attractors, the use of state-of-the-art clustering techniques, for instance, can provide interesting directions to explore. Secondly, in the example of the cell cycle model we have

found that four variables were sufficient to build a suitable partition. We plan to see the possibility to build a reduced version of the cell cycle around these variables, in order to provide further dynamical analyses.

REFERENCES

- [1] K.J. Barnum and M.J. O’Connell. Cell cycle regulation by checkpoints. *Methods in molecular biology (Clifton, NJ)*, 1170:29–40, 2014.
- [2] A. R. Barr, F. S. Heldt, T. Zhang, C. Bakal, and B. Novák. A dynamical framework for the all-or-none G1/S transition. *Cell Systems*, 2(1):27 – 37, 2016.
- [3] M.V. Blagosklonny and A. B. Pardee. The restriction point of the cell cycle. *Cell Cycle*, 1(2):102–109, 2002.
- [4] M. Chaves and L. Tournier. Analysis tools for interconnected boolean networks with biological applications. *Frontiers in physiology*, 9, 2018.
- [5] L. Comtet. *Advanced Combinatorics Adance*. Springer, Dordrecht, 1974.
- [6] G.M. Cooper. *The Cell: A Molecular Approach. 2nd edition*. Sunderland (MA): Sinauer Associates, 2000.
- [7] M.I. Davidich and S. Bornholdt. Boolean network model predicts cell cycle sequence of fission yeast. *PLOS ONE*, 3(2):1–8, 02 2008.
- [8] A. Fauré, A. Naldi, C. Chaouiya, and D. Thieffry. Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle. *Bioinformatics*, 22(14):e124–e131, 2006.
- [9] A. Fauré and D. Thieffry. Logical modelling of cell cycle control in eukaryotes: a comparative study. *Mol. BioSyst.*, 5:1569–1581, 2009.
- [10] D.A. Foster, Y. Paige, X. Limei, and S. Mahesh. Regulation of G1 cell cycle progression: Distinguishing the restriction point from a nutrient-sensing cell growth checkpoint(s). *Genes & Cancer*, 1(11):1124–1131, 2010.
- [11] L. Frécon. *Éléments de mathématiques discrètes*. Presses Polytechniques et Universitaires Romandes, 1st edition, 2002.
- [12] F. Li, T. Long, Y. Lu, O. Qi, and T. Chao. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences*, 101(14):4781–4786, 2004.
- [13] Haitao Li, Yuzhen Wang, and Peilian Guo. State feedback based output tracking control of probabilistic boolean networks. *Information Sciences*, 349:350:1 – 11, 2016.
- [14] A. Murray. Cell cycle checkpoints. *Current Opinion in Cell Biology*, 6(6):872 – 876, 1994.
- [15] E.A. Nigg. Cyclin-dependent protein kinases: Key regulators of the eukaryotic cell cycle. *BioEssays*, 17(6):471–480, 1995.
- [16] E. Ortiz-Gutiérrez, K. García-Cruz, E. Azpeitia, M. Sánchez A. Castillo, and E.R.A. Álvarez Buylla. A dynamic gene regulatory network model that recovers the cyclic behavior of arabidopsis thaliana cell cycle. *PLOS Computational Biology*, 11(9):1–28, 09 2015.
- [17] Z. Qu, J.N. Weiss, and W. R. MacLellan. Regulation of the mammalian cell cycle: a model of the G1-to-S transition. *American Journal of Physiology-Cell Physiology*, 284(2):C349–C364, 2003.
- [18] C.J. Sherr and J.M. Roberts. Cdk inhibitors: positive and negative regulators of G1-phase progression. *Genes & development*, 13(12):1501–1512, June 1999.
- [19] L. Tournier and M. Chaves. Uncovering operational interactions in genetic networks using asynchronous boolean dynamics. *J Theor Biol*, 260(2):196–209, 2009.
- [20] L. Tournier and M. Chaves. Interconnection of asynchronous boolean networks, asymptotic and transient dynamics. *Automatica*, 49(4):884–893, 2013.
- [21] P. Traynard, A. Fauré, F. Fages, and D. Thieffry. Logical model specification aided by model-checking techniques: application to the mammalian cell cycle regulation. *Bioinformatics*, 32(17):772–780, 2016.
- [22] J.J. Tyson and B. Novak. Regulation of the eukaryotic cell cycle: Molecular antagonism, hysteresis, and irreversible transitions. *Journal of Theoretical Biology*, 210(2):249 – 263, 2001.
- [23] N. Zhang, Y. Tian, and J. M. Patel. Discovery-driven graph summarization. In *2010 IEEE 26th International Conference on Data Engineering (ICDE 2010)*, pages 880–891, March 2010.



Qualitative Analysis of Mammalian Circadian Oscillations: Cycle Dynamics and Robustness

Ousmane Diop^{1(✉)}, Madalena Chaves², and Laurent Tournier¹

¹ MaIAGE, INRAE, Université Paris-Saclay, 78350 Jouy-en-Josas, France
ousmane.diop@inrae.fr

² Université Côte d’Azur, Inria, INRAE, CNRS, Sorbonne Université, Biocore Team,
Sophia Antipolis, France

Abstract. In asynchronous Boolean models, periodic solutions are represented by terminal strongly connected graphs, which are typically composed of hundreds of states and transitions. For biological systems, it becomes a challenging task to compare such mathematical objects with biological knowledge, or interpret the transitions inside an attractor in terms of the sequence of events in a biological cycle. A recent methodology generates summary graphs to help visualizing complex asynchronous attractors and order the dynamic progression based on known biological data. In this article we apply this method to a Boolean model of the mammalian circadian clock, for which the summary graph recovers the main phases of the cycle, in the expected order. It also provides a detailed view of the attractor, suggesting improvements in the design of the model’s logical rules and highlighting groups of transitions that are essential for the attractor’s robustness.

Keywords: Mammalian circadian clock · Asynchronous Boolean network · Complex attractor · Summary graph

1 Introduction

The analysis of periodic orbits and their properties remains a most challenging problem in dynamical systems theory. Many living systems exhibit periodical dynamics and the current literature covers a large diversity of mathematical models used to represent, explore, and study the mechanisms leading to physical or biological rhythms [7]. A thorough analysis of such cyclic attractors opens the door to a whole family of meaningful questions related to the robustness of the oscillatory behavior, the estimation and control of the period or amplitude of oscillations in terms of the parameters of the system, the location of the orbit in the state space, *etc.*

Supported by the ANR (French agency for research) through project ICycle ANR-16-CE33-0016-01.

Very little is known on how to express the properties of a periodic solution in terms of the system's parameters but qualitative models, such as piecewise linear or Boolean models, suggest some ideas. Piecewise linear systems partition the state space into regions where solutions of the system can be explicitly computed leading, in some examples, to the estimation of the period and other quantities in terms of the parameters [13,14]. Boolean models provide an ideal framework to analyze qualitative dynamical properties by enabling algorithmic approaches to characterize, for instance, the location of a periodic orbit in the state space, or the influence of the interaction graph on the dynamical behavior [17,21].

In an asynchronous Boolean network, a periodic orbit corresponds to a terminal strongly connected component of the state transition graph. If the computation of such an object is not an issue from a theoretical point of view, its size can grow very large, strongly limiting the biological interpretation of the states and transitions inside the attractor. To tackle this issue, general approaches can be used such as model checking techniques [23]. Recently, a dedicated method developed by Diop et al. [5] proposes to generate a *summary graph* of an asynchronous attractor, based on a classification of its states according to experimentally observed phases of the biological system. The summary graph provides a qualitative view of the general progression along the periodic orbit, capturing the underlying dynamics within the cyclic attractor. In [5] it was successfully applied to a Boolean model of the mammalian cell cycle [6].

In this paper, we propose to apply the summary graph method to the mammalian circadian clock, a biological rhythm which is based on the interactions among five main proteins CLOCK:BMAL1, REV-ERB α , ROR, PER, and CRY. The core of the clock mechanism is formed by three feedback loops. First, the BMAL1 complex promotes the transcription of *Per* and *Cry* genes. The corresponding proteins bind to form a complex PER:CRY, which then translocates to the nucleus where it will block the transcriptional activity of BMAL1, thus forming a first negative feedback loop. In addition, BMAL1 also promotes the transcription of the two genes *Rev-erb* and *Ror*. Eventually, *Bmal1* transcription will be inhibited by REV-ERB and activated by ROR, leading to a second negative loop and a new positive loop, respectively. The periodic behavior of the clock system is determined in large part by the phase opposition between CLOCK:BMAL1 and PER:CRY, which corresponds to the day/night succession (BMAL1 peaks during the day).

Examples of Boolean models for circadian clocks include one for the plant *Arabidopsis thaliana* [1] and a compact mammalian clock model [4] which reproduces the interplay between the negative feedback loops induced by BMAL1 activity. However, our present objective is to have a deeper understanding of the circadian cycle recently developed by Almeida et al. [2]. This is a continuous model of the core clock mechanism that faithfully reproduces the circadian rhythm, by including not only the five main proteins but also their transcription regulated by clock controlled elements (CCE). To use the graph method [5], we first construct a Boolean version of the continuous model in [2] that exhibits one cyclic attractor and, in addition, correctly reproduces the effect of some well

known gene knock-outs. Next, Sect. 3 analyzes the cyclic attractor with the summary graph method, by classifying groups of Boolean states according to their corresponding circadian time zones. Finally, in Sect. 4 the summary graph is further used to identify key groups of transitions within the attractor and relate them back to parts of the network's logical rules. This analysis has two outcomes, first suggesting an improved Boolean rule for one of the variables where the model lacked clear information. As a second outcome, our analysis predicts that some transitions, while allowing some short-circuits between phases, seem to contribute to the overall robustness of the attractor and globally ensure the good progression of the cell clock.

2 Proposing a New Boolean Model of the Circadian Clock

As evoked previously, the following Boolean model of the mammalian circadian clock is highly inspired by the continuous model developed in [2]. In addition to the three feedback loops already described, this model takes into account the transcription of the five clock proteins, each regulated through a particular combination of transcription factors. The latter bind to specific sites on the promoters, called clock controlled elements (CCEs): *Ebox* (enhancer box), *Dbox* (DBP/E4BP4 response element), and *Rbox* (REV-ERB/ROR response element). With the introduction of *Dbox* (which activates both REV-ERB and PER), two new proteins are added to the model, DBP and E4BP4, each also regulated by one of the CCE.

A complete and detailed justification of the continuous model assumptions and construction can be found in [2], but we provide a brief summary of the eight variables and corresponding differential equations in Table 1.

2.1 Construction of the Boolean Model

The design of a Boolean or of a continuous model for the same biological system differs in some fundamental aspects and, in general, there is no direct equivalence between terms in the two frameworks. For instance, activation and inhibition links typically have clear logical representations, but the effects of detailed mass-action kinetics or mass conservation laws are harder to represent in a Boolean model, and may require the definition of new variables. The purpose of our Boolean model is to transcribe as closely as possible the interactions in the continuous model in [2], as described below. The continuous equations and corresponding logical rules are shown in Table 1, for a clear comparison between the two models.

In [2], there are three CCE named *Dbox*, *Ebox*, and *Rbox*, each responding to the conjugation of two components, an activator and an inhibitor. More precisely, *Dbox* is activated by DBP and inhibited by E4, *Ebox* is activated by Bmal and inhibited by CRY and *Rbox* is activated by ROR and inhibited by REV. The continuous equations are formed by a synthesis term depending on one or two

Table 1. Differential equations of the circadian clock model in [2] and corresponding logical rules in the Boolean model. The term $M_{PC} = \gamma_{pc}PER \cdot CRY - \gamma_{cp}PC$ represents the kinetics of the complex binding: $PER + CRY \rightleftharpoons PC$ (mass action law). We use classical Boolean operators: $X \vee Y$ (X or Y), $X \wedge Y$ (X and Y) and $\overline{X}$ (not X).

Continuous equation	Logical rule
$dBMAL1/dt = Rbox - \gamma_{bp}BMAL1 \cdot PC$	$Bmal' = Rbox \wedge \overline{PC}$
$dROR/dt = Ebox + Rbox - \gamma_{ror}ROR$	$ROR' = Rbox_2$
$dREV/dt = 2Ebox + Dbox - \gamma_{rev}REV$	$ROR2' = Ebox \wedge Rbox \wedge ROR$
$dDBP/dt = Ebox - \gamma_{dbp}DBP$	$REV' = Ebox \vee REV2$
$dE4/dt = 2Rbox - \gamma_{e4}E4$	$REV2' = Ebox \wedge Dbox \wedge REV$
$dCRY/dt = Ebox + 2Rbox - M_{PC} - \gamma_cCRY$	$DBP' = Ebox$
$dPER/dt = Ebox + Dbox - M_{PC} - \gamma_pPER$	$E4' = Rbox$
$dPC/dt = M_{PC} - \gamma_{bp}BMAL1 \cdot PC$	$CRY' = Ebox \vee CRY2$
	$CRY2' = Ebox \wedge Rbox \wedge CRY$
	$PER' = Ebox \vee Dbox$
	$PC' = PER \wedge CRY$

CCE and a degradation term. The CRY , PER and PC equations also contain the binding and dissociation terms denoted M_{PC} in Table 1.

In general, Boolean variables are assumed to degrade when not updated in the next step and, therefore, linear degradation terms do not appear explicitly. Thus, a naive approach to construct a corresponding Boolean model is to combine the synthesis terms as logical conjugations or disjunctions of the given variables and directly obtain the rule, for instance: $DBP' = Ebox$ or $E4' = Rbox$. However, not all continuous equations follow this simple construction and other properties that strongly contribute to the dynamics must be taken into account.

Indeed, (i) some variables have nonlinear degradation (cf. $BMAL1$), (ii) others contain mass-action terms (cf. CRY, PER, PC) and (iii) three of the variables are regulated by two CCEs and are themselves regulators of CCEs (ROR , REV , CRY). This last property implies the existence of different thresholds for the different regulatory activities. For this reason, ROR , REV , and CRY are assumed to have an extra discrete level, here represented by extra Boolean variables $ROR2$, $REV2$, and $CRY2$ (following [24]), under the assumption

$$x = \begin{cases} 0 & \text{when } X = X2 = 0, \\ 1 & \text{when } X = 1, X2 = 0, \\ 2 & \text{when } X = X2 = 1. \end{cases}$$

The states corresponding to $X = 0, X2 = 1$ have no biological meaning (also called “forbidden”) and trajectories towards these states from the other “biological” states must be excluded. This can be achieved by using the method described in [3] which complements the rules for $X' = (\dots)$ and $X2' = (\dots)$ as follows: $X' = (\dots) \vee X2$ and $X2' = (\dots) \wedge X$. Since the CCEs appear in additive

form in the continuous equations, we assume that only one CCE is sufficient to trigger the activity of X , while the two CCEs are needed to trigger the activity of $X2$. The order in which the CCEs are activated and in turn activate each variable, was decided by comparing to the continuous solutions. The expressions for the CCEs are written as logical conjunctions:

$$\begin{aligned} Ebox &= Bmal \wedge \overline{CRY2}, & Dbox &= DBP \wedge \overline{E4}, \\ Rbox &= ROR \wedge \overline{REV}, & Rbox_2 &= ROR2 \wedge \overline{REV}. \end{aligned}$$

In the continuous model, there is only one nonlinear degradation term, corresponding to the inhibition of BMAL1 transcription by PC ($-\gamma_{bp} BMAL1 \cdot PC$). Accordingly, we assume $Bmal$ is explicitly repressed by PC but, conversely, PC is not strongly affected (indeed, our analysis showed that a rule of the form $PC = PER \wedge CRY \wedge \overline{Bmal}$ prevents oscillatory behavior).

The binding of PER and CRY to form the complex PC is described by mass-action kinetics in the continuous model (see term M_{PC} in Table 1). In the Boolean model, for simplicity, we assumed that PC is produced when both PER and CRY are available, leading to the rule $PC = PER \wedge CRY$, but no explicit effect from PC on PER or CRY . Although there is no systematic way to translate mass-action kinetics or other mass conservation laws into Boolean factors, and these are usually treated on a case-by-case basis, we will see in Sect. 4 that the analysis by the summary graph method suggests a refinement of the PER rule. This refinement can be interpreted as a more suitable way to include the mass-action terms into the circadian clock Boolean model.

In this way, we obtained a qualitative multi-valued model which closely translates the differential model of [2]. The multi-valued model is equivalent to the 11-dimensional Boolean network depicted in Table 2, to which the methodology in [5] can now be applied.

Table 2. Logical rules of the Boolean clock model, with 11 variables.

$Bmal' = ROR \wedge \overline{REV} \wedge \overline{PC}$
$ROR' = ROR2 \vee \overline{REV}$
$REV' = (Bmal \wedge \overline{CRY2}) \vee REV2$
$DBP' = Bmal \wedge \overline{CRY2}$
$E4' = ROR \wedge \overline{REV}$
$CRY' = Bmal \vee CRY2$
$PER' = (Bmal \wedge \overline{CRY2}) \vee (DBP \wedge \overline{E4})$
$PC' = PER \wedge CRY$
$REV2' = Bmal \wedge \overline{CRY2} \wedge DBP \wedge \overline{E4} \wedge REV$
$CRY2' = Bmal \wedge \overline{CRY2} \wedge ROR \wedge \overline{REV} \wedge CRY$
$ROR2' = Bmal \wedge \overline{CRY2} \wedge ROR \wedge \overline{REV}$

2.2 First Dynamical Analysis of the Boolean Model

To compute and analyze the dynamical behavior of a Boolean model $X'_i = F_i(X_1, \dots, X_n)$, $i = 1, \dots, n$, we need to specify an updating order for the variables. Applying the logical rules in the order defined by the updating schedule leads to a *state transition graph* (STG) with 2^n states, where a sequence of transitions represents a trajectory of the system. To analyze the STG, we first compute its strongly connected components (SCCs) which are defined as sets of states C , such that for every pair of states $x, y \in C$, there exist two paths (or sequences of transitions) in C leading from x to y and from y to x . SCCs may consist of single or multiple states and may have incoming and outgoing transitions, but two distinct SCCs can not be mutually connected, otherwise they would form a single SCC. The asymptotic behavior of the system is thus characterized by the SCCs without outgoing transitions, also called terminal SCCs or *attractors*. An attractor with multiple states represents a periodic orbit of the system.

As a preliminary analysis, we considered the basic synchronous updating schedule, where all variables are simultaneously updated: $X_i[t+1] = F_i(X[t])$. The synchronous STG of the model contains a single attractor, composed of only five states. This simple cyclic attractor captures the *Bmal*/*PER*:*CRY* phase opposition, a central feature in the circadian clock (see Fig. 1).

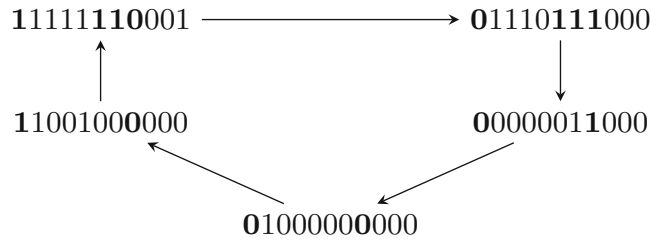


Fig. 1. Synchronous attractor of the Boolean model (variables are ordered as in Table 2). Bold digits indicate the succession of states corresponding to the *Bmal*/*PC* phase opposition: in the left column, *Bmal* is expressed while *PC* is turned off, the former leading to expression of *PER* and *CRY* (top left); next *PC* becomes expressed and *Bmal* is turned off (right column), and *CRY* and *PER* eventually turn off; at the bottom both *PC* and *Bmal* are off, before a new cycle begins.

In this paper, we will prefer an asynchronous updating schedule as it is much more realistic from a biological perspective. In this scheme, at most one variable is updated at each instant. To construct the STG for this asynchronous schedule, for each state $X = (X_1, \dots, X_n)$ define the subset of variables $I_X = \{i : X_i \neq F_i(X)\}$. Then, for each $i \in I_X$, add a transition $X \rightarrow Y$ where $Y_i = F_i(X)$ and $Y_j = X_j$ for all $j \neq i$. More details on the asynchronous strategy can be found for instance in [21]. From a biological point of view, asynchronous updating is preferable as it allows for variability and different timescales in the network interactions. From a graph theoretical point of view, synchronous and

asynchronous STGs are very different as the asynchronous STG is non deterministic, *i.e.* a state no longer has one and only one successor. Usually, this generates lots of transitions, making biological interpretation more difficult.

When applied to our circadian model, the asynchronous strategy still gives a single attractor, but the terminal SCC is now composed of 442 states and 1737 transitions. Interestingly, the five states of the synchronous attractor are all present in the asynchronous attractor, but the simple cyclic trajectory has now been replaced by a complex graph with a much larger amount of details. To further analyze this attractor and validate our model, we consider six different versions of the model representing six known mutations (see Table 3). For some of them, circadian oscillations may be completely lost, which in our model translates to the attractor shrinking to a single fixed state. For others, oscillations endure but are somewhat degraded (*eg.* with a shorter period); in those cases our model conserves a complex attractor, but with less states than the original. These results seem to confirm that the model in Table 2 reproduces the essential core of the circadian clock dynamics.

Table 3. Effects of some mutations on the dynamics of the Boolean model.

Mutation	Biological phenotype	Effect on the attractor
Bmal=0	Arrhythmic (complete loss of circadian oscillations) [9, 18, 19]	Single state attractor 0100100000
PER=0	Abnormal circadian oscillations [9, 18, 19]	Attractor of 114 states
CRY=0	Abnormal circadian oscillations [9, 18, 19]	Attractor of 106 states
REV=0	Shorter cycle period [15, 18]	Attractor of 80 states
ROR=0	Arrhythmic [9, 18]	Single state attractor 000000000000
REV=1	Arrhythmic [10]	Single state attractor 001000000000

Mutant analysis constitutes an interesting way to validate a discrete dynamical model. Indeed, the comparison between wild type and mutant phenotypes usually provides qualitative differences, such as the disappearance of oscillations for instance, that can be well captured by a Boolean model. Nevertheless, when dealing with such a complex attractor (hundreds of states, thousands of transitions), a more direct comparison of the attractor with biological data is rapidly limited, hindering model validation. For example, in Table 3 the degradation of circadian oscillations is paralleled with the number of states in the attractor, which is questionable. In the following we use the methodology proposed in [5], constructing a reduced version of the attractor based on biological knowledge. This summary graph leads to refine the analysis of the attractor, providing further validation of the model.

3 Comparing the Attractor with Circadian Oscillations

3.1 Dividing the Circadian Cycle into Qualitative Phases

In order to further analyze the model's attractor, a necessary first step is to classify its states into groups that will correspond to different stages of the circadian clock. This is essentially a modeling step, therefore there is not a unique way to make this classification, as it is based on a compromise between available biological data on the one hand, and the different variables and interactions included in the model on the other. In the following, we give a brief description of the main regulatory events during the circadian cycle, together with the modeling choices we made to deduce the corresponding partition of the attractor's states. For a comprehensive biological review we mainly referred to [20]. Note that this article is based on data at the transcriptional level; therefore, we sometimes used other sources to complete our classification (see [2, 16] and references therein).

As already mentioned, a hallmark of circadian rhythm progression is the phase opposition between the CLOCK:BMAL1 complex on the one hand and the PER:CRY complex on the other. This opposition divides the cycle into two major steps, approximately correlated with the day/night separation (see Fig. 2). More precisely, [20] introduces two biological phases, respectively called *Activation* and *Repression*. The first one takes place during the day and corresponds to the activation of Bmal, while PER:CRY is absent. The second one sees the repression of Bmal and takes place during the night. Projecting this on the variables of our model, this leads to consider two groups of states: one where $Bmal = 1$, $PC = 0$ and one where $Bmal = 0$, $PC = 1$.

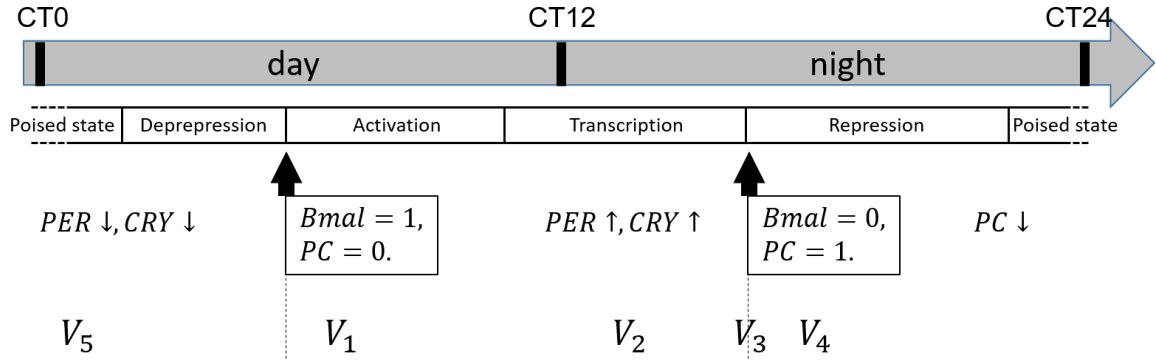


Fig. 2. Description of the main qualitative stages of the circadian clock. CT stands for Circadian Time and is an standard marker of time arbitrarily starting (CT0) at the beginning of activity for a diurnal organism. Top: temporal succession of biological phases as described in [20]; middle: projection of the main regulatory events on the model's variables; bottom: corresponding qualitative phases V_i defined in (1).

These two main phases are separated by intermediate phases called *Transcription* and *Poised state-Depression* in [20]. In the *Transcription* phase, *Per* and *Cry* genes are transcribed, followed by their complexation and the translocation of the complex into the nucleus. With the lack of precise timing of these

events, we decided to subdivide this into a first step where PER or CRY are present but not at the same time, followed by a second step where they are both present. The latter corresponds to the pre-formation step of the PER:CRY complex, *i.e.* PER and CRY have sufficiently accumulated and the complex is about to form in the cytosol [19]. Finally, after the repression phase, the PER:CRY complex disappears due to auto-repression of PER and CRY [8]. Again, without knowing the precise timing of the disappearance of PER, CRY and the complex, we simply consider intermediary states where $Bmal = PC = PER = CRY = 0$, just before the activation of Bmal and the subsequent beginning of a new cycle.

This description leads to the definition of five groups of states in the attractor, or qualitative phases, denoted by $(V_i)_{1 \leq i \leq 5}$ and defined as follows.

$$\begin{cases} V_1 = (Bmal \wedge \overline{PC}) \wedge (\overline{PER} \vee \overline{CRY}) & \text{(afternoon),} \\ V_2 = (Bmal \wedge \overline{PC}) \wedge PER \wedge CRY & \text{(late afternoon),} \\ V_3 = Bmal \wedge PC \wedge PER \wedge CRY & \text{(transition day-night),} \\ V_4 = \overline{Bmal} \wedge PC & \text{(night),} \\ V_5 = \overline{Bmal} \wedge \overline{PC} \wedge \overline{PER} \wedge \overline{CRY} & \text{(late night to next morning).} \end{cases} \quad (1)$$

These phases are defined by taking into account the main variables of our Boolean model. They are based on qualitative and not temporal considerations; however, thanks to the description in [20] we were able to approximately place them along the circadian time scale (see Fig. 2).

Remark 1. For the sake of convenience, we use the same symbol V_i to designate both the subset of states in the attractor and the Boolean formula describing those states. For instance, $V_4 = \overline{Bmal} \wedge PC$ denotes the set of states in the attractor such that $Bmal = 0$ and $PC = 1$.

3.2 Construction of the Summary Graph

Let $A = (V, E)$ denote the attractor, which is a directed graph over $|V| = 442$ states. The sets V_i defined by (1) are, by definition, subsets of V that are mutually exclusive. To complete them into a partition of V , introduce the set $U = V \setminus \left(\bigcup_{i=1}^5 V_i\right)$, containing “unclassified” states. The first thing to note with this partition is that every V_i is actually not empty, confirming the attractor accurately captures all important phases of the circadian clock. To be more precise, $|V_1| = 77$, $|V_2| = 31$, $|V_3| = 31$, $|V_4| = 136$ and $|V_5| = 30$ and there are $|U| = 137$ unclassified states. We now briefly recall the definition of a summary graph (interested reader may refer to [5] for more details).

Definition 1. *The summary graph of the graph $A = (V, E)$ on a partition $\mathcal{P} = \{V_1, V_2, \dots, V_k\}$ of V is the directed graph $\mathcal{G} = (\mathcal{P}, \mathcal{E})$, whose vertices are the V_i , $i = 1, \dots, k$ and where there is an edge from V_i to V_j iff $i \neq j$ and there exist $x \in V_i$ and $y \in V_j$ such that $(x, y) \in E$.*

From a graph theoretical point of view, $\mathcal{G}$ is simply the quotient graph of A on the partition $\mathcal{P}$. For the sake of simplicity, we use the same symbol V_i to design

the set of states V_i and the vertex V_i of the summary graph. This enables an easy description of sets of trajectories in the attractor A : actually, each edge of the summary graph is unequivocally associated with a subset of asynchronous transitions in A . The summary graph of the attractor is depicted in Fig. 3 (top left).

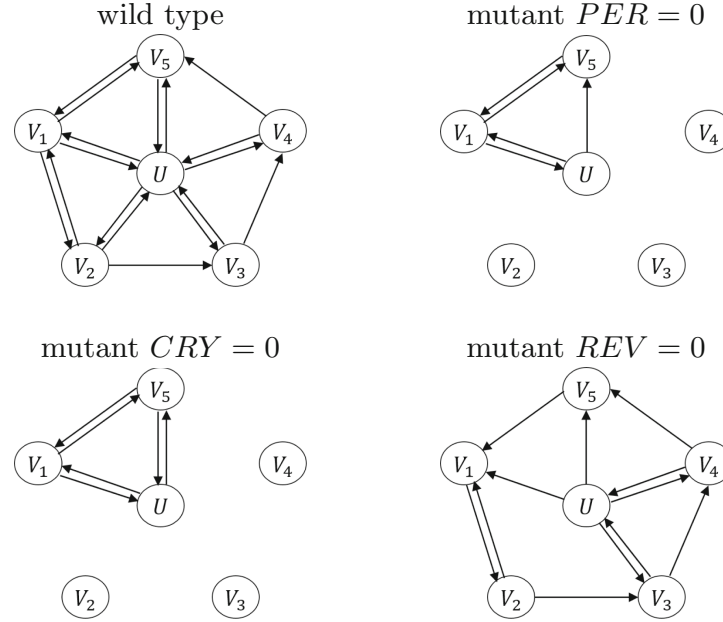


Fig. 3. Summary graph of the attractor on the partition $\{U, V_i, i = 1 \dots 5\}$ defined in (1), in wild type condition as well as for mutants $PER = 0$, $CRY = 0$ and $REV = 0$.

The summary graph in Fig. 3(top left) shows that the expected succession of phases $V_1 \rightarrow V_2 \rightarrow V_3 \rightarrow V_4 \rightarrow V_5 \rightarrow V_1$ is actually present, confirming the existence of accurate circadian oscillations in the attractor, with respect to regulators $Bmal$, PER , CRY and $PER:CRY$. Moreover, three one-directional transitions $V_2 \rightarrow V_3$, $V_3 \rightarrow V_4$, and $V_4 \rightarrow V_5$ even indicate irreversible progression through the cycle. The edge $V_2 \rightarrow V_3$ corresponds to the formation and translocation of the PC complex, following the accumulation of PER and CRY . This irreversibility is in adequacy with [19]: as the main role of PC is to inhibit $Bmal$, it will switch off only after $Bmal = 0$. Similarly, the edge from V_3 to V_4 corresponds to transitions where $Bmal$ is repressed, which is known to be irreversible. As for the edge $V_4 \rightarrow V_5$, it is associated to a set of transitions where PC switches off and can be viewed as a consequence of PER and CRY auto-repression [8]. From a graph point of view, the presence of these three one-directional edges imposes a general orientation for the cycle, consistent with what is known on the opposition between the *Activation* and *Repression* phases of [20].

In addition to the summary graph in wild type condition, Fig. 3 also shows the summary graphs for the three mutants that exhibited oscillatory behaviors in Table 3: $PER = 0$, $CRY = 0$ and $REV = 0$. Interestingly, the mutation of REV has a limited impact on the succession of the phases, showing this

perturbation does not impair the cycle progression, as it was observed in [15, 18]. For *PER* and *CRY* mutants, the perturbation is stronger and the cycle is impaired. However, we still observe a *Bmal* oscillation between V_1 and V_5 (notably through U), confirming circadian oscillations may persist, although in a degraded form [9, 18, 19].

Whether in wild type or in mutant conditions, the summary graph reveals to be a powerful tool to analyze asynchronous attractors of complex oscillatory biological systems, such as the circadian clock. In the next part, we propose an extension of this analysis, where we show how to use this tool to help redesign the network, and assess the general robustness of the model.

4 Advanced Analysis of the Attractor

The summary graphs in Fig. 3 allow a direct comparison of the model's attractor with biological knowledge on the circadian cycle, both in wild type and mutant conditions. In particular, we were able to track the expected succession of qualitative phases within the attractor, thus validating that our model captures the essential core of the regulatory network. Nevertheless, the summary graphs also point to spurious transitions to (and from) the set of unclassified states U , responsible for “short-circuits” between phases. Such spurious trajectories may reveal abnormal behaviors, with unwanted transitions in the model, or they may catch important links ensuring the overall robustness of the cycle. In both cases, the summary graph provides an ideal tool to detect those spurious transitions and analyze them in a biological context.

4.1 Adjustment of the Attractor and Refinement of the Model

We start with a more in-depth examination of the set U . In Sect. 3 it is defined as $U := V \setminus (\bigcup_i V_i) \subseteq \{0, 1\}^{11}$ and is a subset of 137 states in the attractor that we were not able to classify according to the general description of the circadian cycle. By assigning the value 1 to the states in U and the value 0 to the states in $V \setminus U$, we obtain a partial Boolean function (PBF) that we can identify using Boolean inference techniques (see [21, 22]). The inference consists in identifying minimal supports¹, *i.e.* minimal subsets of variables that are sufficient to reproduce the PBF. This function provides an easy way to characterize the unclassified states, showing they can be decomposed into two components:

$$U = \underbrace{\overline{Bmal} \wedge \overline{PC} \wedge (PER \vee CRY)}_{U_1} \vee \underbrace{Bmal \wedge PC \wedge \overline{PER}}_{U_2}. \quad (2)$$

While the inference of a PBF does not lead to a unique solution in general, here (2) is actually unique, as the set $\{Bmal, PC, PER, CRY\}$ is the only minimal

¹ In [21], this step is performed by the algorithm REVEAL [11], in [22] it is performed by the algorithm presented in [12]. Overall, all the inferences performed in the present paper are fast, due to the small dimension (11) of the network. A general discussion about the complexity of the inference problem can be found for instance in [22].

support of the PBF U (it will be the case for all PBF identified in this article). As before, we conveniently use the same symbol U to designate the set of unclassified states (within the attractor) and the Boolean function that characterizes them (see Remark 1 above).

The Boolean formula of U is decomposed as in (2) to highlight two separate subsets of unclassified states: the first one U_1 is composed of states where variables $Bmal$ and PC are both off whereas the second one U_2 is composed of states where variables $Bmal$ and PC are both on. Specifically, in this second subset the PER:CRY complex is forming (since it has not repressed Bmal yet) while PER is switched off. Clearly, such states should not exist in the attractor since PER is essential for complex formation. In order to observe the interplay between U_1, U_2 with the different phases V_i , we reconstruct the summary graph of the attractor, using Definition 1 on the new partition $\{V_1 \dots, V_5, U_1, U_2\}$ (Fig. 4, left).

By looking at this graph, one can see that the set U_2 is connected to the rest of the attractor mainly through phase V_3 . More precisely, only six transitions in the attractor are responsible for the entry into U_2 , all coming from V_3 . Since an asynchronous strategy is used, all these transitions can be traced back to situations where PER has disappeared before the complex PER:CRY has completed the repression of Bmal, leading to a contradiction. Therefore, we decide to suppress these transitions in the attractor, leading to the disconnection of the set U_2 (Fig. 4, center).

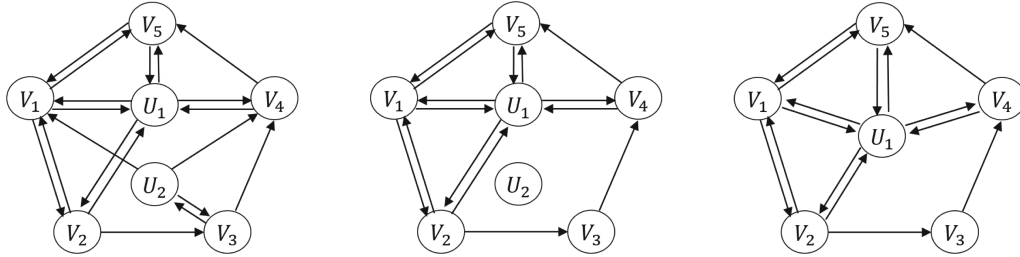


Fig. 4. Refined summary graphs. Left: initial attractor's summary graph with U decomposed in (U_1, U_2) ; center: summary graph once the six transitions $V_3 \rightarrow U_2$ have been removed; right: summary graph of the new amended model's attractor.

After the removal of the six transitions in the graph of the attractor, we obtain a new amended model by applying a few steps. First, we recompute the strongly connected component decomposition of the truncated attractor. We find a unique terminal SCC of 393 states, in which the states in U_2 have disappeared. From this SCC we reconstruct the partial truth table of the network on these 393 states, and then use an inference technique to identify the logical rules corresponding to this partial truth table. The last two steps (construction of partial truth table, followed by inference) are described in more details in [5] and, in a different context, in [21]. Finally, the only affected rule of the network is the rule of variable PER , with the addition of a new (unique) clause:

$$PER' = (Bmal \wedge \overline{CRY2}) \vee (DBP \wedge \overline{E4}) \vee (Bmal \wedge PC). \quad (3)$$

To verify that this modification did not alter other parts of the dynamics, we made the same analyses as in Sects. 2 and 3, confirming that the slight modification in (3) was sufficient to get rid of unwanted states U_2 while conserving good dynamical properties, in wild type and mutant conditions. The summary graph of this new amended model is depicted in Fig. 4, on the right.

The modified rule (3) adds a single new interaction in the network, which is a positive effect of PC onto PER (the positive effect of $Bmal$ to PER was already present in the original network). Interestingly, this effect was present in the ODE system of [2], as an unbinding term of the $PER:CRY$ complex. As already said, such mass action law kinetic terms are often made implicit by default in the Boolean framework. Here, our method points to the importance of this particular one to avoid unwanted transitions within the attractor. From a modeling point of view, the summary graph thus provides a valuable help in the design of Boolean models of complex, oscillating biological systems.

4.2 A Tool to Assess the General Robustness of the Attractor

The remaining set of unclassified states U_1 is composed of states where $Bmal$ and PC are off while PER or CRY can be on. Contrary to U_2 , this set is highly connected to almost every phases of the attractor (see Fig. 4). In total, the four edges entering in U_1 amount to 117 transitions. As before, we tried to remove those transitions in order to disconnect U_1 from the rest of the attractor, thus obtaining a final summary graph perfectly matching the order of the phases. However, although showing no more short-cut between phases, this version loses some essential dynamical property. Indeed, the PER and CRY mutants no longer exhibit oscillatory behaviors, suggesting that at least a subset of these transitions are needed for a proper behavior of the model.

Nevertheless, even though those transitions cannot be removed all together, we can still use the summary graph to investigate the model further and analyze each edge entering U_1 separately, in the context of the circadian cycle.

1. The edge $V_4 \rightarrow U_1$ is associated with 63 transitions in which the variable PC switches off while PER and CRY are not both deactivated. From a biological point of view, it means that the $PER:CRY$ complex disappears before PER and CRY have disappeared, indicating a short half-life of the complex. These transitions are to be compared with the $V_4 \rightarrow V_5$ transitions where PC switches off after PER and CRY , suggesting a longer half-life. In our model the two types of transitions coexist since the relative times of disappearance of the three components, PER , CRY and $PER:CRY$ are not taken explicitly into account. Interestingly, note that the removal of the 63 transitions does not affect the global dynamical properties of the attractor, as the mutants are not affected.
2. The edge $V_5 \rightarrow U_1$ corresponds to 9 transitions in the attractor, in which the variable PER switches on while $Bmal$ is not yet active. Since PER transcription is activated through the $CLOCK:BMAL1$ complex, these transitions

seem rather unrealistic. To investigate further, we removed the transitions and applied the same technique as before, to obtain the modified *PER* rule:

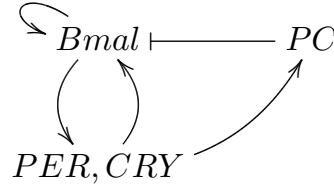
$$PER' = (Bmal \wedge \overline{CRY2}) \vee (Bmal \wedge PC) \\ \vee [(DBP \wedge \overline{E4}) \wedge (PER \vee CRY \vee PC)] .$$

Dynamically, the removal does not affect the main properties of the attractor (the summary graphs of wild type and mutant conditions are similar) however, the inferred logical rule exhibits new and undocumented interactions, namely an auto-activation of *PER* and a positive effect of *CRY* on *PER*. This points to a specific part of the model that will need a closer look in the future.

3. Finally, the edges $V_1 \rightarrow U_1$ and $V_2 \rightarrow U_1$ correspond to a total of 45 transitions in which the variable *Bmal* switches to 0 in the absence of *PC*. With respect to circadian clock events, these transitions describe an early deactivation of *Bmal*, leading the cycle to bypass the important steps of *PER:CRY* formation and translocation (phase V_3). However, it is the removal of these transitions that directly alter the behaviors of the mutants, suggesting they are necessary to ensure the robustness of the attractor. Their removal leads to a new rule for *Bmal*:

$$Bmal' = (ROR \wedge \overline{REV} \wedge \overline{PC}) \vee [(Bmal \wedge (PER \vee CRY) \wedge \overline{PC})],$$

that highlights, besides an auto-activation term, a direct positive effect of *PER* and *CRY* on *Bmal*. Those new interactions generate two positive feedback loops involving *Bmal*, *PER* and *CRY* that directly interfere with the negative feedback loop between *Bmal* and *PC*:



This alteration of the negative loop is directly responsible for the loss of robustness of the model.

The objective of this section was to analyse the “short-cut” transitions in the circadian attractor (see Fig. 4 left). As shown by the three examples above, the summary graph combined with Boolean inference indicates that some of these unwarranted transitions should not be removed and, in fact, appear to play a significant role during the progression of the cycle, namely in the system’s response to gene knock-outs.

5 Conclusions and Perspectives

In this article we used the summary graph, a novel tool introduced in [5], to analyze a Boolean model of the mammalian circadian cycle regulation. This tool is well adapted to study biological oscillations, as it provides a rational way to compare a complex Boolean attractor (with hundreds of states) with an oscillating phenomenon. Combined with Boolean inference techniques, it becomes particularly useful in model design as it allows to relate a local dynamical property of the attractor, such as the transition from one phase to the next, to the part of the network's topology directly responsible for it.

After verifying the model's attractor correctly captures the correct succession of phases, the summary graph was further analyzed to provide more insight on the model.

In a first step, our analysis indicates that the attractor contains many “short-cuts”, that is transitions between states (essentially towards sets U_1 and U_2) which may lead to distorted cycles with very short days or very short nights. In a second step, the analysis shows that some of these short-cuts (those passing through set U_2) can be removed by a refinement of the logical rules. Namely, the rule of PER should be modified to include the effect of PC dissociation which, for simplicity, was not taken into account in the initial Boolean model (compare Table 2 and Eq. (3)). Finally, in a third step, our analysis suggests that some short-cuts, specifically $V_1 \rightarrow U_1$ and $V_2 \rightarrow U_1$, are necessary for a correct performance. These transitions are characterized by an early BMAL1 deactivation, and they are responsible also for generating the short cycle mutants observed in the PER and CRY knock-outs. The transitions through U_1 may thus be necessary to generate circadian cycle robustness in response to perturbations in gene expression.

More generally, the summary graph provides an efficient way to tackle complex qualitative attractors, by testing the effect of specific perturbations on the dynamics. In future works we plan to further investigate the role of the states U_1 in maintaining circadian oscillations, by studying the links between state transitions and the topology of the circadian network.

Acknowledgements. The authors would like to thank Franck Delaunay for many useful discussions on the circadian clock events, as well as for pointers to relevant references.

References

1. Akman, O., Watterson, S., Parton, A., Binns, N., Millar, A., Ghazal, P.: Digital clocks: Simple boolean models can quantitatively describe circadian systems. *J. R. Soc. Interface* **9**, 2365–2382 (2012)
2. Almeida, S., Chaves, M., Delaunay, F.: Transcription-based circadian mechanism controls the duration of molecular clock states in response to signaling inputs. *J. Theor. Biol.* **484**, 110015 (2020)

3. Chaves, M., Tournier, L., Gouzé, J.L.: Comparing Boolean and piecewise affine differential models for genetic networks. *Acta Biotheor.* **58**(2–3), 217–232 (2010)
4. Comet, J.P., Bernot, G., Das, A., Diener, F., Massot, C., Cessieux, A.: Simplified models for the mammalian circadian clock. *Procedia Comput. Sci.* **11**, 127–138 (2012)
5. Diop, O., Tournier, L., Fromion, V.: Summarizing complex asynchronous Boolean attractors, application to the analysis of a mammalian cell cycle model. In: 18th European Control Conference (ECC), Naples, Italy, pp. 1677–1682 (2019)
6. Fauré, A., Naldi, A., Chaouiya, C., Thieffry, D.: Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle. *Bioinformatics* **22**(14), e124–e131 (2006)
7. Forger, D.B.: *Biological Clocks, Rhythms, and Oscillations: The Theory of Biological Timekeeping*. MIT Press (2017)
8. Gallego, M., Virshup, D.M.: Post-translational modifications regulate the ticking of the circadian clock. *Nat. Rev. Mol. Cell Biol.* **8**(2), 139–148 (2007)
9. Ko, C.H., Takahashi, J.S.: Molecular components of the mammalian circadian clock. *Hum. Mol. Genet.* **15**(suppl.2), R271–R277 (2006)
10. Kornmann, B., Schaad, O., Bujard, H., Takahashi, J.S., Schibler, U.: System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock. *PLoS Biol.* **5**(2), e34 (2007)
11. Liang, S., Fuhrman, S., Somogyi, R.: Reveal, a general reverse engineering algorithm for inference of genetic network architectures. In: *Pacific Symposium on Biocomputing*, vol. 3, pp. 18–29 (1998)
12. Murakami, K., Uno, T.: Efficient algorithms for dualizing large-scale hypergraphs. *Discrete Appl. Math.* **170**, 83–94 (2014)
13. Ndiaye, I., Chaves, M., Gouzé, J.L.: Oscillations induced by different timescales in signal transduction modules regulated by slowly evolving protein-protein interactions. *IET Syst. Biol.* **4**(4), 263–276 (2010)
14. Poignard, C., Chaves, M., Gouzé, J.L.: A stability result for periodic solutions of nonmonotonic smooth negative feedback systems. *SIAM J. Appl. Dyn. Syst.* **17**(2), 1091–1116 (2018)
15. Preitner, N., et al.: The orphan nuclear receptor rev-erb α controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* **110**(2), 251–260 (2002)
16. Relógio, A., Westermark, P.O., Wallach, T., Schellenberg, K., Kramer, A., Herzog, H.: Tuning the mammalian circadian clock: robust synergy of two loops. *PLoS Comput. Biol.* **7**(12), e1002309 (2011)
17. Remy, E., Mossé B., Thieffry D.: Boolean dynamics of compound regulatory circuits. In: Rogato, A., Zazzu, V., Guarracino, M. (eds.) *Dynamics of Mathematical Models in Biology*. Springer, Cham (2016)
18. Ripperger, J.A., Jud, C., Albrecht, U.: The daily rhythm of mice. *FEBS Lett.* **585**(10), 1384–1392 (2011)
19. Rosensweig, C., Green, C.B.: Periodicity, repression, and the molecular architecture of the mammalian circadian clock. *Eur. J. Neurosci.* **51**(1), 139–165 (2018)
20. Takahashi, J.S.: Transcriptional architecture of the mammalian circadian clock. *Nat. Rev. Genet.* **18**(3), 164 (2017)
21. Tournier, L., Chaves, M.: Uncovering operational interactions in genetic networks using asynchronous Boolean dynamics. *J. Theor. Biol.* **260**(2), 196–209 (2009)
22. Tournier, L., Goelzer, A., Fromion, V.: Optimal resource allocation enables mathematical exploration of microbial metabolic configurations. *J. Math. Biol.* **75**, 1349–1380 (2017). <https://doi.org/10.1007/s00285-017-1118-5>

23. Traynard, P., Feillet, C., Soliman, S., Delaunay, F., Fages, F.: Model-based investigation of the circadian clock and cell cycle coupling in mouse embryonic fibroblasts: Prediction of *reverb- α* up-regulation during mitosis. *BioSyst.* **149**, 59–69 (2016)
24. Van Ham, P.: How to deal with variables with more than two levels. In: *Kinetic Logic a Boolean Approach to the Analysis of Complex Regulatory Systems*, pp. 326–343. Springer, Berlin, Heidelberg (1979)

Bibliographie

- [1] Wassim Abou-Jaoudé, Pauline Traynard, Pedro T Monteiro, Julio Saez-Rodriguez, Tomáš Helikar, Denis Thieffry, and Claudine Chaouiya. Logical modeling and dynamical analysis of cellular networks. *Frontiers in genetics*, 7 :94, 2016.
- [2] Ozgur E Akman, Steven Watterson, Andrew Parton, Nigel Binns, Andrew J Millar, and Peter Ghazal. Digital clocks : simple boolean models can quantitatively describe circadian systems. *Journal of The Royal Society Interface*, 9(74) :2365–2382, 2012.
- [3] Urs Albrecht and Gregor Eichele. The mammalian circadian clock. *Current opinion in genetics & development*, 13(3) :271–277, 2003.
- [4] S Almeida, M Chaves, and F Delaunay. Transcription-based circadian mechanism controls the duration of molecular clock states in response to signaling inputs. *Journal of theoretical biology*, 484 :110015, 2020.
- [5] Sofia Almeida, Madalena Chaves, Franck Delaunay, and Celine Feillet. A comprehensive reduced model of the mammalian cell cycle. *IFAC-PapersOnLine*, 50(1) :12617–12622, 2017.
- [6] Thaddeus A Bargiello, F Rob Jackson, and Michael W Young. Restoration of circadian behavioural rhythms by gene transfer in drosophila. *Nature*, 312(5996) :752–754, 1984.
- [7] K.J. Barnum and M.J. O’Connell. Cell cycle regulation by checkpoints. *Methods in molecular biology (Clifton, NJ)*, 1170 :29–40, 2014.
- [8] Jiri Bartek, Jirina Bartkova, and Jiri Lukas. The retinoblastoma protein pathway and the restriction point. *Current Opinion in Cell Biology*, 8(6) :805 – 814, 1996.
- [9] Abraham Berman and Robert J Plemmons. *Nonnegative matrices in the mathematical sciences*. SIAM, 1994.
- [10] M.V. Blagosklonny and A. B. Pardee. The restriction point of the cell cycle. *Cell Cycle*, 1(2) :102–109, 2002.
- [11] Johannes Boonstra and Jan Andries Post. Molecular events associated with reactive oxygen species and cell cycle progression in mammalian cells. *Gene*, 337 :1 – 13, 2004.
- [12] Laurence Calzone, Laurent Tournier, Simon Fourquet, Denis Thieffry, Boris Zhivotovsky, Emmanuel Barillot, and Andrei Zinovyev. Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLoS Comput Biol*, 6(3) :e1000702, 2010.
- [13] Madalena Chaves, Reka Albert, and Eduardo D Sontag. Robustness and fragility of boolean models for genetic regulatory networks. *Journal of theoretical biology*, 235(3) :431–449, 2005.

- [14] Madalena Chaves and Miguel Preto. Hierarchy of models : From qualitative to quantitative analysis of circadian rhythms in cyanobacteria. *Chaos : An Interdisciplinary Journal of Nonlinear Science*, 23(2) :025113, 2013.
- [15] Madalena Chaves, Laurent Tournier, and Jean-Luc Gouzé. Comparing boolean and piecewise affine differential models for genetic networks. *Acta biotheoretica*, 58(2-3) :217–232, 2010.
- [16] Frédéric CHARBONNIER Christophe CHANOINE. *Biologie cellulaire et moléculaire de la cellule eucaryote*. Ellipses, 2018.
- [17] Krzysztof Ciesielski. The poincaré-bendixson theorem : from poincaré to the xxist century. *Open Mathematics*, 10(6) :2110–2128, 2012.
- [18] JEAN Colin, JEAN Timbal, CHARLES Boutelier, YVON Houdas, and MICHEL Siffre. Rhythm of the rectal temperature during a 6-month free-running experiment. *Journal of applied physiology*, 25(2) :170–176, 1968.
- [19] Jean-Paul Comet, Gilles Bernot, Aparna Das, Francine Diener, Camille Massot, and Amélie Cessieux. Simplified models for the mammalian circadian clock. *Procedia Computer Science*, 11 :127–138, 2012.
- [20] L. Comtet. *Advanced Combinatorics Adanced*. Springer, Dordrecht, 1974.
- [21] G.M. Cooper. *The Cell : A Molecular Approach. 2nd edition*. Sunderland (MA) : Sinauer Associates, 2000.
- [22] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2009.
- [23] D. Coudreuse and P. Nurse. Driving the cell cycle with a minimal cdk control network. *Nature*, 468(1074-1079), 2010.
- [24] M.I. Davidich and S. Bornholdt. Boolean network model predicts cell cycle sequence of fission yeast. *PLOS ONE*, 3(2) :1–8, 02 2008.
- [25] Etienne Farcot and Jean-Luc Gouzé. Limit cycles in piecewise-affine gene network models with multiple interaction loops. *International Journal of Control*, 41(1) :119–130, 2010.
- [26] A Fauré, A Naldi, C Chaouiya, and D Thieffry. Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle. *Bioinformatics*, 22(14) :e124–e131, 2006.
- [27] A. Fauré and D. Thieffry. Logical modelling of cell cycle control in eukaryotes : a comparative study. *Mol. BioSyst.*, 5 :1569–1581, 2009.
- [28] D. Fisher, L. Krasinska, D. Coudreuse, and B. Novák. Phosphorylation network dynamics in the control of cell cycle transitions. *Journal of Cell Science*, 125(20) :4703–4711, 2012.
- [29] D.A. Foster, Y. Paige, X. Limei, and S. Mahesh. Regulation of G1 cell cycle progression : Distinguishing the restriction point from a nutrient-sensing cell growth checkpoint(s). *Genes & Cancer*, 1(11) :1124–1131, 2010.
- [30] L. Frécon. *Eléments de mathématiques discrètes*. Presses Polytechniques et Universitaires Romandes, 1st edition, 2002.
- [31] M Gallego and D M Virshup. Post-translational modifications regulate the ticking of the circadian clock. *Nature reviews Molecular cell biology*, 8(2) :139–148, 2007.

- [32] C. Gérard and A. Goldbeter. Temporal self-organization of the cyclin/cdk network driving the mammalian cell cycle. *Proceedings of the National Academy of Sciences, National Acad Sciences*, 106(51) :21643–21648, 2009.
- [33] Claude Gérard and Albert Goldbeter. A skeleton model for the network of cyclin-dependent kinases driving the mammalian cell cycle. *Interface Focus*, 1(1) :24–35, 2011.
- [34] C Giacinti and Antonion Giordano. Rb and cell cycle progression. *Oncogene*, 25(38) :5220–5227, 2006.
- [35] Albert Goldbeter. *Au cœur des rythmes du vivant : La vie oscillatoire*. Odile Jacob, 2018.
- [36] Cecilia Gómez-Santos, Purificación Gómez-Abellán, Juan A Madrid, Juan J Hernández-Morante, Juan A Lujan, José M Ordovas, and Marta Garaulet. Circadian rhythm of clock genes in human adipose explants. *Obesity*, 17(8) :1481–1485, 2009.
- [37] A Gonzalez Gonzalez, Aurélien Naldi, Lucas Sanchez, Denis Thieffry, and Claudine Chaouiya. Ginsim : a software suite for the qualitative modelling, simulation and analysis of regulatory networks. *Biosystems*, 84(2) :91–100, 2006.
- [38] Brian C Goodwin. Oscillatory behavior in enzymatic control processes. *Advances in enzyme regulation*, 3 :425–437, 1965.
- [39] Charles M Grinstead and James Laurie Snell. *Grinstead and Snell’s Introduction to Probability : The CHANCE Project*. American Mathematical Society, 2006.
- [40] Paul E Hardin, Jeffrey C Hall, and Michael Rosbash. Feedback of the drosophila period gene product on circadian cycling of its messenger rna levels. *Nature*, 343(6258) :536–540, 1990.
- [41] J.W. Harper, Burton J.L., and Solomon M.J. The anaphase-promoting complex : it’s not just for mitosis anymore. *Genes Dev.*, (16) :2179–2206., 2002.
- [42] Stacy L Harvey, Alyson Charlet, Wilhelm Haas, Steven P Gygi, and Douglas R Kellogg. Cdk1-dependent regulation of the mitotic inhibitor wee1. *Cell*, 122(3) :407–420, 2005.
- [43] Isabel Heyde and Henrik Oster. Differentiating external zeitgeber impact on peripheral circadian clock resetting. *Scientific Reports (Nature Publisher Group)*, 9 :1–13, 2019.
- [44] S.A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22(3) :437 – 467, 1969.
- [45] John G Kemeny and J Laurie Snell. Finite markov chains springer-verlag. *New York Berlin Herdelberg Tokyo*, 1976.
- [46] Nobuya Koike, Seung-Hee Yoo, Hung-Chung Huang, Vivek Kumar, Choogon Lee, Tae-Kyung Kim, and Joseph S Takahashi. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science*, 338(6105) :349–354, 2012.
- [47] Takao Kondo and Masahiro Ishiura. The circadian clocks of plants and cyanobacteria. *Trends in plant science*, 4(5) :171–176, 1999.

- [48] Ronald J Konopka and Seymour Benzer. Clock mutants of drosophila melanogaster. *Proceedings of the National Academy of Sciences*, 68(9) :2112–2116, 1971.
- [49] B Kornmann, O Schaad, H Bujard, J S Takahashi, and U Schibler. System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock. *PLoS biology*, 5(2), 2007.
- [50] J-C Leloup and A Goldbeter. Modeling the mammalian circadian clock : sensitivity analysis and multiplicity of oscillatory mechanisms. *Journal of theoretical biology*, 230(4) :541–562, 2004.
- [51] Jean-Christophe Leloup and Albert Goldbeter. Toward a detailed computational model for the mammalian circadian clock. *Proceedings of the National Academy of Sciences*, 100(12) :7051–7056, 2003.
- [52] F. Li, T. Long, Y. Lu, O. Qi, and T. Chao. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences*, 101(14) :4781–4786, 2004.
- [53] Jiang Li, Xin-Ke Yang, Xin-Xin Yu, Meng-Liang Ge, Wen-Liang Wang, Jie Zhang, and Yun-De Hou. Overexpression of p27kip1 induced cell cycle arrest in g1 phase and subsequent apoptosis in hcc-9204 cell line. *World Journal of Gastroenterology*, 6(4) :513, 2000.
- [54] Shoudan Liang, Stefanie Fuhrman, Roland Somogyi, et al. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. In *Pacific symposium on biocomputing*, volume 3, pages 18–29, 1998.
- [55] Hanna-Stina Martinsson, Maria Starborg, Fredrik Erlandsson, and Anders Zetterberg. Single cell analysis of g1 check points—the relationship between the restriction point and phosphorylation of prb. *Experimental cell research*, 305(2) :383–391, 2005.
- [56] Nuno D Mendes, Rui Henriques, Elisabeth Remy, Jorge Carneiro, Pedro T Monteiro, and Claudine Chaouiya. Estimating attractor reachability in asynchronous logical models. *Frontiers in physiology*, 9 :1161, 2018.
- [57] L. Mendoza and X. Ioannis. A method for the generation of standardized qualitative dynamical systems of regulatory networks. *Theoretical Biology and Medical Modelling*, 3(1) :13, Mar 2006.
- [58] Jayashree Mitra and Greg H Enders. Cyclin a/cdk2 complexes regulate activation of cdk1 and cdc25 phosphatases in human cells. *Oncogene*, 23(19) :3361–3367, 2004.
- [59] Jennifer A Mohawk, Carla B Green, and Joseph S Takahashi. Central and peripheral circadian clocks in mammals. *Annual review of neuroscience*, 35 :445–462, 2012.
- [60] Keisuke Murakami and Takeaki Uno. Efficient algorithms for dualizing large-scale hypergraphs. *Discrete Applied Mathematics*, (170) :83–94, 2014.
- [61] A. Murray. Cell cycle checkpoints. *Current Opinion in Cell Biology*, 6(6) :872 – 876, 1994.
- [62] James D Murray. *Mathematical biology : I. An introduction*, volume 17. Springer Science & Business Media, 2007.
- [63] Emi Nagoshi, Camille Saini, Christoph Bauer, Thierry Laroche, Felix Naef, and Ueli Schibler. Circadian gene expression in individual fibroblasts : cell-autonomous and self-sustained oscillators pass time to daughter cells. *Cell*, 119(5) :693–705, 2004.

- [64] Aurélien Naldi, Céline Hernandez, Wassim Abou-Jaoudé, Pedro T Monteiro, Claudine Chaouiya, and Denis Thieffry. Logical modeling and analysis of cellular regulatory networks with ginsim 3.0. *Frontiers in physiology*, 9 :646, 2018.
- [65] E.A. Nigg. Cyclin-dependent protein kinases : Key regulators of the eukaryotic cell cycle. *BioEssays*, 17(6) :471–480, 1995.
- [66] B. Novák and J.J. Tyson. A model for restriction point control of the mammalian cell cycle. *Journal of Theoretical Biology*, 230(4) :563 – 579, 2004. Special Issue in honour of Arthur T. Winfree.
- [67] Motoaki Ohtsubo, Anne M Theodoras, Jill Schumacher, James M Roberts, and Michele Pagano. Human cyclin e, a nuclear protein essential for the g1-to-s phase transition. *Molecular and cellular biology*, 15(5) :2612–2624, 1995.
- [68] E. Ortiz-Gutiérrez, K. García-Cruz, E. Azpeitia, M. Sánchez A. Castillo, and E.R.A. Álvarez Buyla. A dynamic gene regulatory network model that recovers the cyclic behavior of arabidopsis thaliana cell cycle. *PLOS Computational Biology*, 11(9) :1–28, 09 2015.
- [69] Arthur B Pardee. A restriction point for control of normal animal cell proliferation. *Proceedings of the National Academy of Sciences*, 71(4) :1286–1290, 1974.
- [70] Michele S Park, Juan Rosai, Hai T Nguyen, Paola Capodiceci, Carlos Cordon-Cardo, and Andrew Koff. p27 and rb are on overlapping pathways suppressing tumorigenesis in mice. *Proceedings of the National Academy of Sciences*, 96(11) :6382–6387, 1999.
- [71] N Preitner, F Damiola, J Zakany, D Duboule, U Albrecht, U Schibler, et al. The orphan nuclear receptor rev-erb α controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell*, 110(2) :251–260, 2002.
- [72] Z. Qu, J.N. Weiss, and W. R. MacLellan. Regulation of the mammalian cell cycle : a model of the G1-to-S transition. *American Journal of Physiology-Cell Physiology*, 284(2) :C349–C364, 2003.
- [73] Michael Rape and Marc W Kirschner. Autonomous regulation of the anaphase-promoting complex couples mitosis to s-phase entry. *Nature*, 432(7017) :588–595, 2004.
- [74] A Relógio, P O Westermarck, T Wallach, K Schellenberg, A Kramer, and H Herzl. Tuning the mammalian circadian clock : robust synergy of two loops. *PLoS computational biology*, 7(12), 2011.
- [75] J A Ripperger, C Jud, and U Albrecht. The daily rhythm of mice. *FEBS letters*, 585(10) :1384–1392, 2011.
- [76] Clark Rosensweig and Carla B Green. Periodicity, repression, and the molecular architecture of the mammalian circadian clock. *European Journal of Neuroscience*, 51(1) :139–165, 2018.
- [77] Julian D. Schwab, Silke D. Kühlwein, Nensi Ikononi, Michael Kühl, and Hans A. Kestler. Concepts in boolean network modeling : What do they all mean ? *Computational and Structural Biotechnology Journal*, 18 :571 – 582, 2020.
- [78] Clayton Schwarz, Amy Johnson, Mardo Kõivomägi, Evgeny Zatulovskiy, Carolyn J Kravitz, Andreas Doncic, and Jan M Skotheim. A precise cdk activity threshold determines passage through the restriction point. *Molecular cell*, 69(2) :253–264, 2018.

- [79] C.J. Sherr and J.M. Roberts. Cdk inhibitors : positive and negative regulators of G1-phase progression. *Genes & development*, 13(12) :1501—1512, June 1999.
- [80] Jill C Sible and John J Tyson. Mathematical modeling as a tool for investigating cell cycle control networks. *Methods*, 41(2) :238–247, 2007.
- [81] Sabrina L Spencer, Steven D Cappell, Feng-Chiao Tsai, K Wesley Overton, Clifford L Wang, and Tobias Meyer. The proliferation-quiescence decision is controlled by a bifurcation in cdk2 activity at mitotic exit. *Cell*, 155(2) :369–383, 2013.
- [82] Claudia Stötzl, Susanna Röblitz, and Heike Siebert. Complementing ode-based system analysis using boolean networks derived from an euler-like transformation. *PloS one*, 10(10) :e0140954, 2015.
- [83] Beatrice M Sweeney. *Rhythmic phenomena in plants*. Academic Press, 2013.
- [84] J S Takahashi. Transcriptional architecture of the mammalian circadian clock. *Nature Reviews Genetics*, 18(3) :164, 2017.
- [85] René Thomas. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, 42(3) :563 – 585, 1973.
- [86] Yuanyuan Tian, Richard A Hankins, and Jignesh M Patel. Efficient aggregation for graph summarization. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 567–580, 2008.
- [87] L. Tournier and M. Chaves. Uncovering operational interactions in genetic networks using asynchronous boolean dynamics. *J Theor Biol*, 260(2) :196–209, 2009.
- [88] L Tournier, A Goelzer, and V Fromion. Optimal resource allocation enables mathematical exploration of microbial metabolic configurations. *Journal of mathematical biology*, 75(6-7) :1349–1380, 2017.
- [89] P. Traynard, A. Fauré, F. Fages, and D. Thieffry. Logical model specification aided by model-checking techniques : application to the mammalian cell cycle regulation. *Bioinformatics*, 32(17) :i772–i780, 2016.
- [90] Angela L Tyner and Andrei L Gartel. Roles of cyclin kinase inhibitors in g1 phase progression. *G1 phase progression*. Kluwer, New York, pages 58–76, 2003.
- [91] J.J. Tyson and B. Novak. Regulation of the eukaryotic cell cycle : Molecular antagonism, hysteresis, and irreversible transitions. *Journal of Theoretical Biology*, 210(2) :249 – 263, 2001.
- [92] Katja Vanselow, Jens T Vanselow, Pål O Westermarck, Silke Reischl, Bert Maier, Thomas Korte, Andreas Herrmann, Hanspeter Herzl, Andreas Schlosser, and Achim Kramer. Differential effects of per2 phosphorylation : molecular basis for the human familial advanced sleep phase syndrome (fasps). *Genes & development*, 20(19) :2660–2672, 2006.
- [93] Ying Xu, Quasar S Padiath, Robert E Shapiro, Christopher R Jones, Susan C Wu, Noriko Saigoh, Kazumasa Saigoh, Louis J Ptáček, and Ying-Hui Fu. Functional consequences of a cki δ mutation causing familial advanced sleep phase syndrome. *Nature*, 434(7033) :640–644, 2005.
- [94] Xiao-He Yang and Todd L Sladek. Overexpression of the e2f-1 transcription factor gene mediates cell transformation. *Gene Expression, The Journal of Liver Research*, 4(4-5) :195–204, 1995.

- [95] Guang Yao, Tae Jun Lee, Seiichi Mori, Joseph R Nevins, and Lingchong You. A bistable rb-e2f switch underlies the restriction point. *Nature cell biology*, 10(4) :476–482, 2008.
- [96] William A Zehring, David A Wheeler, Pranhitha Reddy, Ronald J Konopka, Charalambos P Kyriacou, Michael Rosbash, and Jeffrey C Hall. P-element transformation with period locus dna restores rhythmicity to mutant, arrhythmic drosophila melanogaster. *Cell*, 39(2) :369–376, 1984.
- [97] Hongkui Zeng, Zuwei Qian, Michael P Myers, and Michael Rosbash. A light-entrainment mechanism for the drosophila circadian clock. *Nature*, 380(6570) :129–135, 1996.
- [98] Anders Zetterberg, Olle Larsson, and Klas G Wiman. What is the restriction point ? *Current opinion in cell biology*, 7(6) :835–842, 1995.
- [99] Binhai Zheng, David W Larkin, Urs Albrecht, Zhong Sheng Sun, Marijke Sage, Gregor Eichele, Cheng Chi Lee, and Allan Bradley. The mper2 gene encodes a functional component of the mammalian circadian clock. *Nature*, 400(6740) :169–173, 1999.

Titre : Analyse mathématique de la dynamique de réseaux de régulation biologique

Mots clés : Système dynamique, Biologie des systèmes, Réseau Booléen asynchrone, Attracteur complexe, Cycle cellulaire, Horloge circadienne.

Résumé : Dans cette thèse, nous nous intéressons à l'analyse qualitative de la dynamique de deux cycles biologiques centraux dans les cellules eucaryotes, le cycle de division cellulaire et l'horloge circadienne. Nous utilisons pour cela des réseaux Booléens asynchrones, bien adaptés à une analyse qualitative. Dans ces réseaux, les cycles sont capturés par des attracteurs complexes, pouvant contenir des centaines d'états.

Nous proposons une nouvelle méthode d'analyse de ces attracteurs complexes, basée sur la construction d'un graphe résumé. Cette méthode permet de comparer les trajectoires contenues l'attracteur avec les propriétés qualitatives du cycle biologique. Nous illustrons notre méthode sur un modèle du cycle cellulaire

de la littérature et sur un modèle de l'horloge circadienne, que nous avons construit à partir d'un modèle continu existant. Dans ces deux modèles, notre méthode s'est montrée efficace pour visualiser la structure de l'attracteur complexe et le comparer avec un cycle biologique. En combinant le graphe résumé avec une chaîne de Markov, nous estimons les proportions de temps passé dans les phases décrites par les oscillations. En le combinant avec une méthode d'inférence Booléenne, nous montrons également comment ajuster localement la dynamique asymptotique du modèle, afin de forcer certaines propriétés dynamiques. Ces deux applications montrent l'intérêt de notre méthode pour la modélisation et l'analyse de réseaux de régulation cellulaire.

Title: Mathematical analysis of the dynamics of biological regulatory networks

Keywords: Dynamical system, Systems biology, Asynchronous Boolean Network, Complex attractor, Cell cycle, Circadian Clock.

Abstract: In this thesis, we are interested in the qualitative analysis of the dynamics of two biological cycles that are central in eukaryotic cells, the cell division cycle and the circadian clock. For that purpose, we use asynchronous Boolean networks that provide an adapted qualitative framework. In these networks, cycles are captured by complex attractors containing hundreds of states.

A new method for the analysis of such complex attractors is proposed. It is based on the construction of a summary graph of the attractor, enabling the comparison between the attractor's trajectories and qualitative properties

of the biological cycle. The method is illustrated on a cell cycle model from the literature and of a circadian clock model we built from an existing continuous model. In both models our method proves to be efficient to visualize the attractor's structure and to compare it with the biological cycle. By combining the summary graph with a Markov chain, proportions of time spent in each phase are estimated. By combining it with a Boolean inference technique, we show how to locally adjust the asymptotic dynamics of the model in order to force specific dynamical properties. These two applications show the interest of our method in the modeling and analysis of cellular regulatory networks.