

Reprogrammation comportementale : modèles, algorithmes et application aux maladies complexes

Célia Biane

▶ **To cite this version:**

Célia Biane. Reprogrammation comportementale : modèles, algorithmes et application aux maladies complexes. Logique en informatique [cs.LO]. Université Paris-Saclay; Université d'Evry-Val-d'Essonne, 2018. Français. NNT : 2018SACLE050 . tel-02107110

HAL Id: tel-02107110
https://hal.science/tel-02107110

Submitted on 23 Apr 2019

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.

université PARIS-SACLAY

NNT : 2018SACLE050

Thèse de doctorat

Reprogrammation comportementale : modèles, algorithmes et application aux maladies complexes

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université d'Évry Val d'Essonne

École doctorale n°580 Sciences et technologie de l'information et de la communication (STIC)
Spécialité de doctorat : Informatique

Soutenue à Évry, le 30 novembre 2018, par

CÉLIA BIANE FOURATI

Composition du Jury :

M. Philippe Dague Professeur, Université Paris-Sud (LRI)	Président
M. Sylvain Sené Professeur, Université d'Aix Marseille (LIS)	Rapporteur
Mme Anne Siegel Directrice de recherche, CNRS (IRISA)	Rapporteure
M. Jean Krivine Chargé de recherche, CNRS (IRIF)	Examinateur
M. Loïc Paulevé Chargé de recherche, CNRS (LRI)	Examinateur
Mme Élisabeth Remy Chargée de recherche, CNRS (I2M)	Examinatrice
M. Franck Delaplace Professeur, Université d'Évry Val d'Essonne (IBISC)	Directeur de thèse

Remerciements

Je ne remercierai jamais assez mon directeur de thèse, Franck Delaplace, de m'avoir permis, à la fin d'un master de Biologie, de me lancer dans un doctorat en informatique. Pendant mon stage de master, Franck m'a fait découvrir un champ de la modélisation dont je n'avais pas idée et m'a impressionnée par son imagination. Puis, pendant ma thèse, il m'a énormément appris, notamment à décomposer les problèmes, structurer des raisonnements, exprimer plus clairement mes idées, organiser efficacement mon travail. Et… à programmer récursif avec, je le cite, « des patterns, qui sont des filtres structurels ». Ce qui, je dois avouer, ne voulait absolument rien dire pour moi il y a trois ans. Au-delà de l'apprentissage de la méthode scientifique et de techniques, faire une thèse est aussi un travail sur soi : on met une part de soi-même dans ce que l'on produit et il faut se confronter au jugement des autres. Franck m'a rassurée et encouragée, à sa façon un peu brusque parfois mais efficace.

Je tiens à remercier Philippe Dague d'avoir présidé le jury de ma thèse et également pour les corrections attentives qu'il a apporté à mon manuscrit. Je souhaite remercier Sylvain Sené et Anne Siegel qui ont accepté de rapporter ma thèse. Sylvain Sené a relu avec énormément d'attention mon manuscrit et ses corrections et commentaires m'ont permis d'améliorer sa qualité. Enfin, je remercie Jean Krivine, Loïc Paulevé et Elisabeth Remy pour leurs conseils et d'avoir accepté de faire partie de ce jury.

Je remercie Tarek Melliti, qui a participé à nos réunions de travail au début de ma thèse et à qui je dois beaucoup. Je remercie également Hanna Klaudel, par qui tout est arrivé puisque c'est son cours qui m'a poussée à venir demander à faire un stage au laboratoire IBISC, pour le travail que nous avons fait ensemble, sa rigueur, son rôle de médiatrice et ses sages conseils. Je ne peux pas oublier Laurence Calzone, pour nos conversations sur la modélisation booléenne des cellules cancéreuses, ainsi que Vir-

ginie Mournetas et Christian Pinset pour l'intérêt qu'ils ont porté à mon travail. Merci également à Murielle Bourgeois pour son efficacité impressionnante mais surtout pour sa gentillesse et sa bonne humeur. Dans un autre contexte, je remercie également Serena Cerrito, Guillaume Hutzler, Franck Delaplace, Hanna Klaudel et Blaise Hanczar d'avoir fait confiance à la toute récente informaticienne que j'étais pour assurer les travaux dirigés et pratiques de leurs cours. Cela m'a permis de consolider mes bases en programmation, logique et théorie des langages. Et merci aussi à Serena pour son livre de logique qui a passé trois ans sur mon bureau !

Je remercie également les membres permanents, stagiaires, doctorants et ex-doctorants des équipes COSMO et AROBAS (Jérémy et Jérémie, Clément, Eli, Stan, Audrey, Ludo, Ahmed, Samy, Louis, Sara, Maria, Laurent, Jean-Marc, Jean-Christophe), pour les 3 ans de déjeuners, à base de surgelés Picard, à refaire le monde. Je remercie aussi la joyeuse bande de bio-informaticiens du Génoscope pour les bières post-travail. Parmi eux, il y a bien sûr Amin, son humour, son enthousiasme et ses conseils que je n'ai jamais suivis.

Je remercie très chaleureusement mes amis de longue date, Charlotte, Valentin, Clara, Théodore, Adrien ainsi que mes ex-colocs Antoine, Florent, Nadia de m'avoir rappelé qu'il y a autre chose, aussi, dehors.

Je souhaite remercier mes parents, mon petit frère et toute ma grande famille d'être là, avec leurs qualités et leurs défauts, et d'accepter les miens.

Enfin, je remercie Johan, pour son soutien constant et cette confiance que je peux avoir en lui.

Table des matières

Remerciements iii

Table des matieres v

Introduction **1**

1 La reprogrammation comportementale **5**

1 Modèles moléculaires des maladies 6

1.1 Modèle génétique 6

1.2 Modèle edgétique 8

1.3 Un modèle dynamique des maladies 10

2 Reprogrammer les réseaux booléens 11

2.1 Réseaux booléens 12

2.2 Réseaux booléens contrôlés 18

2.3 Contrôle par gel 19

3 Spécification de la reprogrammation 22

3.1 Modes de reprogrammation 22

3.2 Noyaux de reprogrammation 24

4 Bilan du chapitre 25

2 Calcul des noyaux de reprogrammation **27**

1 Formulation abductive du calcul des noyaux 27

1.1 Abduction et logique propositionnelle 28

1.2 Inférence abductive des noyaux 30

2 Méthodes de calcul d'impliquants premiers 34

2.1 Formes normales 34

2.2 Calcul des impliquants premiers 39

3 Méthodes d'inférence des noyaux 47
3.1 Méthodes PLNE-Noyau 47
3.2 Méthodes DDB-Noyau 50
3.3 Remarques sur les méthodes PLNE-Noyau et DDB-Noyau 50
4 Comparaison de méthodes 51
4.1 Approche par simulation 51
4.2 Approche par recherche de motifs stables 54
4.3 Approche par calcul algébrique 55
4.4 Synthèse 56

3 Application au Cancer **59**
1 Principes d'application 59
1.1 Dynamique booléenne des réseaux biologiques 60
1.2 Causes et modélisation du cancer 62
1.3 Épistasie et cancer 63
2 Cancer du sein 66
2.1 Conception du modèle 66
2.2 Requêtes et solutions 71
2.3 Analyse des solutions 72
3 Cancer de la vessie 76
3.1 Présentation du modèle 76
3.2 Requêtes et solutions 78
3.3 Analyse des solutions 79
4 Annexe du chapitre 84

4 Comparaison de performances **93**
1 PLNE-Noyau 93
2 DDB-Noyau 95
3 Comparaison de PLNE-Noyau et DDB-Noyau 98

Conclusion et perspectives **103**

Bibliographie **107**

Introduction

La médecine de précision est un champ de recherche en biologie moléculaire et médicale qui repose sur une meilleure compréhension des bases moléculaires des maladies humaines et le développement de thérapies ciblées permettant de les soigner. Le questionnement portant sur la définition des maladies à l'échelle moléculaire peut être appréhendé par deux approches complémentaires : l'empirisme et le rationnalisme [56, 103]. La méthode empirique consiste à établir des mesures statistiques entre les ensembles de symptômes des patients et leurs caractéristiques moléculaires. La méthode rationnelle repose sur l'élaboration de modèles mécanistiques théorisant le fonctionnement d'un organisme et expliquant l'émergence des symptômes de la maladie par des perturbations du modèle. De même, la découverte de thérapies ciblées peut reposer sur une approche par criblage, qui consiste à tester l'efficacité de nombreuses molécules, sans connaître leur mode d'action, ou sur une approche de conception rationnelle qui consiste à prédire les modes d'action de médicaments en se fondant sur des modèles biologiques [96]. Que ce soit en employant des méthodes empiriques ou rationnelles, les évolutions de la médecine de précision sont aujourd'hui indissociables des sciences informatiques [76, 43]. En effet, les méthodes empiriques supposent d'être capable d'analyser les masses de données considérables issues des progrès de la biologie moléculaire et les méthodes rationnelles supposent l'élaboration de modèles de maladies et de méthodes calculatoires permettant de raisonner sur ces modèles pour faire des prédictions.

Dans cette thèse, nous nous intéresserons à la définition d'un cadre de modélisation et de méthodes calculatoires pour l'étude des causes des maladies humaines et la découverte rationnelle de nouvelles cibles thérapeutiques. La définition des maladies par leurs causes moléculaires repose sur la question fondamentale suivante : *quelles*

sont les perturbations moléculaires responsables des dysfonctionnements des systèmes biologiques observés dans les maladies humaines ? et son corollaire pour la découverte de nouvelles thérapies ciblées : *comment perturber le système malade de façon à ce qu'il récupère un comportement sain ?* Par ailleurs, le déchiffrage de la relation causale entre l'information moléculaire (*le génotype*) contenue dans les organismes et leur comportement (*le phénotype*) requiert de s'appuyer sur des formalismes reflétant la complexité de cette relation.

Dans le **chapitre 1** nous présentons d'abord les *modèles génétique* et *edgétique* des maladies. Le modèle génétique établit une relation causale entre la perturbation d'un gène et l'émergence de maladies dites mendéliennes, c'est-à-dire causée par la perturbation d'un gène unique tandis que le modèle edgétique décrit la cause de ces maladies par des perturbations des interactions d'un gène avec d'autres gènes. Nous décrivons ensuite informellement un modèle dynamique des maladies complexes (multigéniques) qui assimile la maladie complexe à la *reprogrammation comportementale* d'un système dynamique discret sous l'action de diverses combinaisons de perturbations. Ce modèle de reprogrammation comportementale est ensuite formalisé sur les modèles dynamiques booléens par la définition des *réseaux booléens contrôlés* et de *modes de reprogrammation*. Les réseaux booléens contrôlés permettent la formalisation de perturbations dans les réseaux booléens grâce à des paramètres de contrôle tandis que la spécification de modes de reprogrammation décrivent les effets attendus de ces perturbations sur les états stables des réseaux booléens.

Dans le **chapitre 2**, nous nous intéressons aux méthodes algorithmiques permettant de calculer les contrôles à effectuer sur un réseau booléen contrôlé afin de le reprogrammer selon un mode donné. Nous montrons notamment que ce problème peut s'exprimer comme un problème abductif en logique, c'est-à-dire comme le calcul des causes d'effets observés. Nous présentons ensuite la *procédure du consensus* et la méthode de *minimisation logique* qui permettent le calcul des causes minimales d'observations exprimés sous forme d'une fonction booléenne et nous montrons comment les adapter à notre problématique. Enfin, nous comparons notre approche à des méthodes traitant de problèmes similaires et publiées récemment.

Le **chapitre 3** est consacré à l'application de la modélisation et des méthodes à

des modèles booléens de cellule pour l'identification des gènes causaux du cancer et la découverte de cibles thérapeutiques. Dans un premier temps, nous présentons les principes généraux de modélisation booléenne des réseaux Biologiques, et proposons un cadre interprétatif de notre cadre théorique. Puis, nous appliquons nos méthodes à un modèle conçu pour le cancer du sein et à un modèle publié précedemment portant sur le cancer de la vessie. Nous montrons que les résultats obtenus sont pertinents avec les connaissances biologiques.

Enfin, dans le **chapitre 4** les performances des méthodes PLNE-Noyau et DDB-Noyau sont comparées.

Chapitre 1

La reprogrammation comportementale

Les *modèles moléculaires de maladie* sont des cadres d'étude qui définissent une relation causale entre des perturbations à l'échelle moléculaire et les comportements malades ou sains de l'organisme. Nous présentons dans ce chapitre un nouveau modèle appelé *reprogrammation comportementale* qui assimile la maladie à une reprogrammation du comportement d'un système dynamique discret par des perturbations des composants et des interactions de ce système. Ce modèle s'insère dans la continuité des modèles génétiques et edgétiques, qui assimilent respectivement la maladie à la perturbation d'un gène et à la perturbation des gènes et/ou de leurs interactions. La reprogrammation y ajoute l'aspect comportemental émergeant de la dynamique du système génétique. Les modèles génétique et edgétique des maladies ainsi que le nouveau modèle dynamique de reprogrammation comportementale que nous proposons sont définis informellement et leur motivation justifiée par des observations biologiques dans la section 1 de ce chapitre. Puis nous introduisons un cadre théorique permettant l'analyse de la reprogrammation comportementale des systèmes dynamiques booléens : dans la section 2, nous définissons les *réseaux booléens contrôlés* qui constituent un formalisme intégrant la notion de perturbation dans les systèmes dynamiques booléens et permettent leur reprogrammation. Dans la section 3, nous proposons deux *modes de reprogrammation* décrivant les effets attendus des perturbations sur la dynamique booléenne et permettant de réaliser des requêtes d'inférence des *noyaux* de cette reprogrammation.

1 Modèles moléculaires des maladies

Les modèles moléculaires de maladies que nous présentons ont pour but d'établir une relation causale entre génotype et phénotype. Ils reposent sur le principe suivant : les variations de fonctionnement des systèmes biologiques sont déterminées par des variations des molécules constituant ces systèmes.

1.1 Modèle génétique

Considérons un gène α existant dans la population sous deux formes alléliques : A, appelé allèle *sauvage* codant pour une protéine normale et *a* appelé allèle *muté* codant pour une protéine non fonctionnelle. Selon le modèle génétique (schématisé en figure 1.1), la combinaison des deux allèles portés par un individu détermine si l'individu a un phénotype malade ou non. Ainsi, si l'allèle muté est *récessif*, le phénotype sera malade si l'individu est porteur de deux allèles mutés, et si l'allèle muté est *dominant*, le phénotype sera malade si l'individu est porteur d'au moins un allèle muté. Le modèle génétique des maladies établit donc une relation causale déterministe entre un génotype et un phénotype [69] qui est le suivant :

Mutation du gène x ⟶ Perte du produit du gène ⟶ maladie X.

Un exemple de maladie génétique récessive est la myopathie de Duchenne qui est caractérisée par la dégénérescence progressive des muscles causée par la mutation du gène codant pour la dystrophine et qui a pour conséquence l'absence d'expression de cette protéine [32].

Ce modèle a permis l'identification de plus de 1000 gènes associés aux maladies monogéniques par comparaison des génomes de patients malades et sains et qui sont répertoriées sur la base de données Online Mendelian Inheritance in Man (OMIM, [67]), une meilleure compréhension des bases moléculaires des maladies humaines et l'élaboration d'outils de diagnostic basés sur des analyses des données génétiques.

Cependant, ce modèle ne rend pas compte de la complexité des relations entre génotype et phénotype, même dans le cas des maladies monogéniques. Une des raisons de cette complexité est que la plupart des gènes sont en réalité *pléiotropes*, c'est-à-dire qu'ils sont associés à plusieurs traits phénotypiques. Une explication biologique

Gène α	A	*a*
A	**Sain**	**Sain**
a	**Sain**	**Malade**

Maladie génétique récessive

Gène β	B	*b*
B	**Sain**	**Malade**
b	**Malade**	**Malade**

Maladie génétique dominante

Figure 1.1 – Modèle génétique établissant une relation causale entre gène et maladie. Une maladie génétique est dite dominante si le phénotype malade s'exprime lorsque le génome contient au moins un allèle muté ou récessive si le phénotype malade s'exprime lorsque le génome contient deux allèles mutés.

Gène α	A	*a*	a'
A	**Sain**	**Maladie 1**	**Sain**
a	**Maladie 1**	**Maladie 1**	**Maladie 2**
a'	**Sain**	**Maladie 2**	**Maladie 2**

Figure 1.2 – Illustration schématique d'un phénomène de pléiotropie dû à une hétérogénéité allélique du gène α.

de cette pléiotropie est *l'hétérogénéité allélique* : un gène pléiotrope peut exister sous différents variants alléliques, eux-mêmes associés à différentes maladies. Considérons un gène α existant sous trois formes alléliques : l'allèle sauvage A et les allèles mutés *a* et a', si *a* est dominant sur A, A est dominant sur a' et a' dominant sur *a* (on pourra remarquer ici que les relations de dominance entre allèles ne sont pas nécessairement transitives), on obtient le tableau présenté figure 1.2. Le gène α est donc pléiotrope : il est associé à deux maladies, la maladie 1 et la maladie 2. La pléiotropie a pour conséquence de modifier la relation causale entre génotype et phénotype qui n'est plus déterministe, comme dans le modèle génétique.

Le modèle *edgétique* des maladies, proposé par [113], complexifie le modèle génétique en intégrant l'effet de la mutation d'un gène sur la protéine codée afin de représenter la pléiotropie.

1.2 Modèle edgétique

Le modèle *edgétique* propose un schéma biologique explicatif de la pléiotropie qui transfère l'identification des causes sur les effets des mutations sur les interactions du produit des gènes avec leurs partenaires. En effet, la mutation d'un gène n'a pas nécessairement pour conséquence la perte totale de l'expression mais peut affecter plus finement son expression. Les mutations ponctuelles, c'est-à-dire résultant du changement d'un seul nucléotide dans la séquence d'ADN, peuvent être classées en trois classes selon leur effet sur la séquence en acides aminés de la protéine :

une mutation même-sens induit un changement de nucléotide qui, du fait de la redondance du code génétique, n'induit aucun changement de séquence de la protéine ;

une mutation faux-sens induit le changement de l'acide aminé codé par le codon d'ADN et donc un changement de la séquence protéique ;

une mutation non-sens transforme le codon en un codon STOP et induit donc un arrêt de la transcription : la protéine résultante est tronquée.

Le changement de séquence protéique par les mutations faux-sens serait donc plus probablement responsable d'anomalies de la structure tridimensionelle et par conséquent de la perturbation de certaines interactions de cette protéine tandis qu'une mutation non-sens aurait plus de chance d'être responsable d'une absence totale de la protéine. Le modèle edgétique suggère donc que l'identification de la cause des maladies et l'élucidation de la relation entre génotype et phénotype repose sur l'analyse de l'*interactome*. Le terme d'interactome a été introduit en 1999 par Bernard Jacq [90] pour désigner la représentation des réseaux biologiques moléculaires sous forme de graphes dont les sommets correspondent à des entités moléculaires, possiblement de différentes natures (ADN, ARN, protéines, métabolites...), et les arêtes des interactions physiques ou fonctionnelles. La figure 1.3 illustre schématiquement le modèle edgétique des maladies défini sur l'interactome.

Les maladies humaines les plus courantes, parmi lesquelles on peut citer le cancer, la maladie d'Alzheimer ou la maladie de Parkinson, sont des maladies dites complexes dont l'identification des causes moléculaires ainsi que des mécanismes moléculaires des médicaments pouvant inverser leur effet est un enjeu majeur de santé publique.

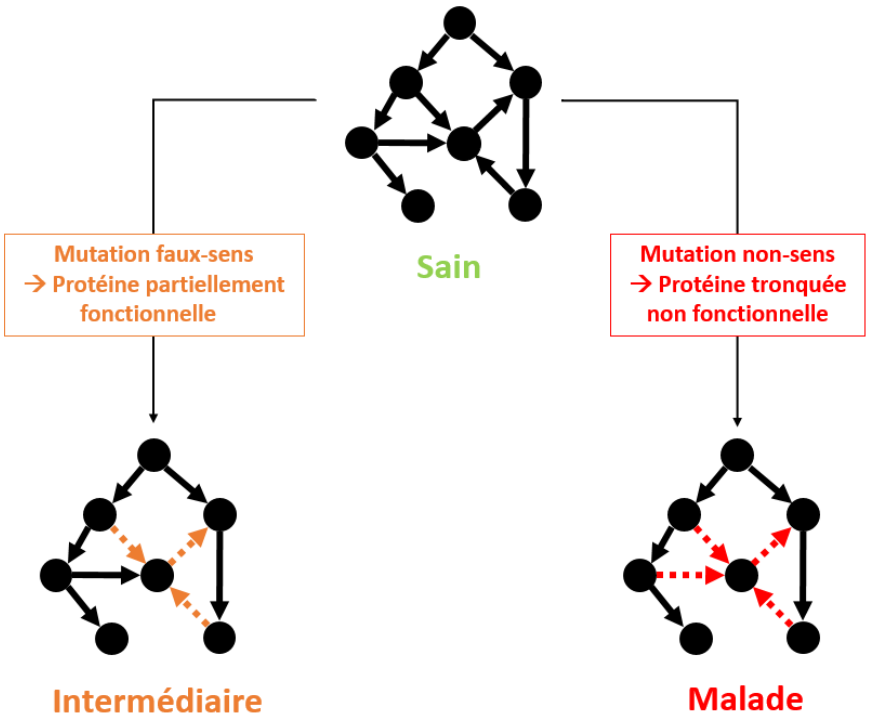


Figure 1.3 – Illustration schématique du modèle edgétique des maladies associant les maladies à des perturbations de l'interactome. Les perturbations des interactions sont illustrées par des arcs en pointillés, la perturbation de toutes les interactions d'un sommet est assimilable à la perte du sommet.

Ces maladies sont à la fois multigéniques, c'est-à-dire que la perturbation simultanée de plusieurs gènes est nécessaire à leur déclenchement, et diverses, c'est-à-dire que plusieurs combinaisons de perturbations peuvent aboutir au déclenchement de la maladie. Le terme *épistasie* regroupe les phénomènes de collaboration ou d'antagonisme entre gènes qui sont à l'origine de la diversité et de la multiplicité causale des maladies complexes. Considérons deux gènes α et β existant chacun sous une forme sauvage A et B et sous une forme mutée *a* et *b*. On parle d'épistasie négative quand la combinaison des deux allèles mutés a un effet délétère sur le phénotype par rapport à l'effet individuel de chaque mutation et on parle d'épistasie positive quand la combinaison des allèles mutés de ces deux gènes a un effet délétère sur le phénotype par rapport à l'effet individuel de chaque mutation [82] (voir l'illustration 3.2).

Afin de prendre en compte ces phénomènes d'épistasie et de multiplicité et diversité causales des maladies humaines, nous définissons un nouveau modèle des maladies s'appuyant sur la dynamique des réseaux biologiques.

β \ α	A	*a*
B	**Sain**	**Sain**
b	**Sain**	**Malade**

Epistasie négative

β \ α	A	*a*
B	**Sain**	**Malade**
b	**Malade**	**Sain**

Epistasie positive

Figure 1.4 – Illustration schématique des phénomènes d'épistasies positive et négative entre 2 gènes α et β

1.3 Un modèle dynamique des maladies

Nous définissons un modèle où les maladies sont assimilées à une *reprogrammation comportementale* d'un système dynamique discret et fini représentant un système biologique.

Un système dynamique discret F décrit l'évolution d'un état global x par un ensemble de n fonctions locales de transition f_i, $1 \leq i \leq n$ telles que chaque fonction locale f_i décrit l'évolution de l'état de la variable x_i, soit :

$$\mathrm{F}(x) = \begin{cases} x_1 \mapsto f_1(x_1, ..., x_n) \\ ... \\ x_n \mapsto f_n(x_1, ..., x_n) \end{cases}$$

Dans les modèles biologiques, les composants i correspondent aux gènes ou autres molécules contenus dans une cellule, et leurs états, qui peuvent représenter leur activité, expression, quantité, concentration, etc, évoluent localement en fonction des autres composants du système. Le comportement global du système est assimilé au comportement («phénotype») d'une entité biologique d'échelle supérieure (cellule, tissu, organe, etc.). La reprogrammation d'un système dynamique F consiste en la réécriture de ses fonctions locales de transition sous l'action de perturbations et engendre un nouveau système G dont le comportement global diffère du comportement global de F. La figure 1.5 illustre la reprogrammation comportementale d'un système dont l'état sain est représenté par une coloration particulière des sommets du graphe d'interaction, en un système malade, dont les sommets sont colorés différemment, sous l'action

de différents ensembles de perturbations.

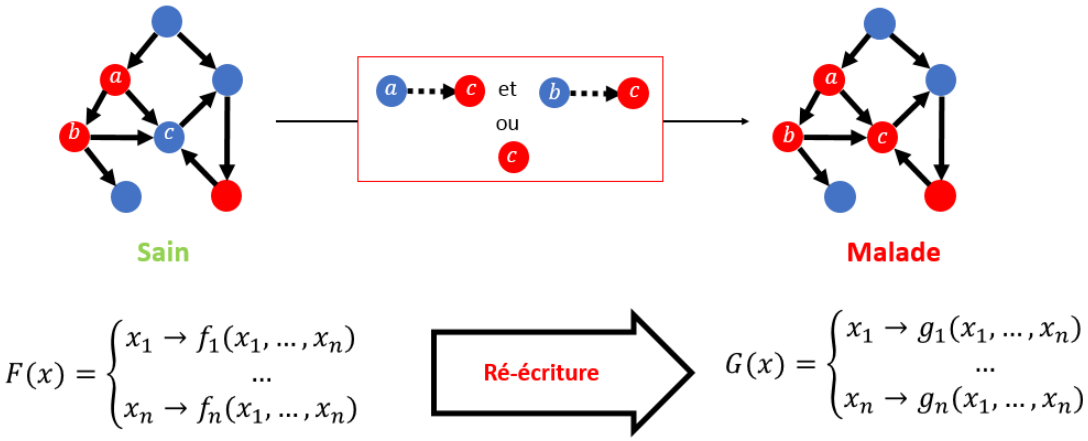


Figure 1.5 – Illustration schématique de la reprogrammation comportementale où les phénotypes sont assimilés à des états dynamiques (représentés par la couleur des sommets du graphe) et les différents types de mutations à la réécriture des fonctions locales de transition.

Dans la prochaine section, nous définissons d'une part un cadre théorique formalisant la réécriture des fonctions locales de transition par le *contrôle* des systèmes dynamiques booléens et d'autre part, un langage de spécification questionnant les contrôles à effectuer sur un système de façon à changer sa *propriété* «phénotypique» (être sain ou malade). Cette variation de propriété sera évaluée par la variation d'état de *biomarqueurs*, qui sont définis comme des espèces biologiques dont la mesure permet la discrimination de différents phénotypes. Le problème d'identification des causes consiste alors à trouver les ensembles de perturbations du système F responsable du changement d'état de ses biomarqueurs.

2 Reprogrammer les réseaux booléens

Les réseaux booléens sont des systèmes dynamiques discrets introduits en biologie par McCulloch et Pitts [66] en 1943 comme modèle de la transmission d'information entre neurones puis par Stuart Kauffman [50] et René Thomas [97] pour modéliser les réseaux de régulation génique. Cette modélisation peut s'étendre à celle de la transmission de l'information à travers les voies de signalisation. Les variables du système

sont alors des molécules prenant les valeurs booléennes, éléments de $\mathbb{B} = \{0,1\}$. Par convention, la valeur 1 correspond à un état actif ou présent et 0 à un état inactif ou absent. La reprogrammation des réseaux booléens consiste à modifier leur dynamique par réécriture des fonctions locales de transition. Dans cette partie, nous formaliserons cette réécriture en nous appuyant sur les *réseaux booléens contrôlés*, et, par analogie au modèle edgétique décrivant des perturbations des interactions et des sommets des graphes biologiques, nous introduirons la catégorie de *contrôle par gel*. Enfin, nous spécifierons deux modes de reprogrammation, la *nécessité* et la *possibilité* questionnant les contrôles à effectuer sur le réseau booléen de façon à modifier sa dynamique.

2.1 Réseaux booléens

Dans cette section, nous donnons d'abord quelques définitions générales puis nous définissons formellement les réseaux booléens et les notions qui leur sont associées : graphe de transition, cycle limites et graphes d'interaction.

2.1.1 Notations

Définition 2.1. Une *relation binaire* sur un ensemble E est un sous-ensemble du produit cartésien E × E. La *fermeture réflexive transitive* d'une relation binaire $\mathscr{R}$, notée $\mathscr{R}^*$, est la relation binaire $\mathscr{R}^* = \bigcup_{m\geq 0} \mathscr{R}^m$.

Exemple 2.1. Une relation binaire sur un ensemble E peut être représentée par un graphe orienté G dont les sommets correspondent aux éléments de E et les arcs aux éléments de R, soit $G = \langle E, R\rangle$. La fermeture réflexive transitive de R sur E est alors représentée par le graphe $C(G) = \langle E, R^*\rangle$. La figure 1.6 représente une relation binaire $R = \{(1,2),(2,3),(3,4)\}$ sur l'ensemble $E = \{1,2,3,4\}$ et sa fermeture réflexive transitive R^*.

On notera $\mathbf{2}^E$ l'ensemble des parties de l'ensemble E et $E \setminus \{e_i\}$ l'ensemble E privé de son élément e_i.

Soit $X = \{x_1, ..., x_n\}$ un ensemble de variables booléennes. Une *interprétation* correspond à l'assignation d'une valeur booléenne à chaque élément x_i de X, c'est-à-dire, $I : X \mapsto \mathbb{B}$.

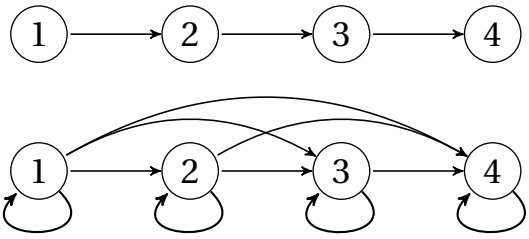


Figure 1.6 – Relation binaire R sur l'ensemble E = {1,2,3,4} (en haut) et sa fermeture réflexive transitive (en bas).

Une *formule propositionnelle* ϕ est construite à partir de *constantes booléennes* (0/ Faux, 1/ Vrai), d'un ensemble de variables booléennes X, de l'opérateur unaire $\neg$ (négation) et d'opérateurs logiques binaires ($\wedge$/ ET/ conjonction), ($\vee$/ OU/ disjonction), ($\Longrightarrow$ / DONC/ implication), ($\veebar$/ XOR/ ou exclusif), etc.

On notera ϕ(I) l'évaluation de ϕ par rapport à une interprétation I sur ses variables.

2.1.2 Réseaux booléens

Définition 2.2. Un *réseau booléen* est un système dynamique discret défini sur un ensemble fini de variables booléennes X = $\{x_1, ..., x_n\}$ et décrit par un système de n équations booléennes de la forme suivante :

$$\mathrm{F}(x) = \begin{cases} x_1 \mapsto f_1(x_1, ..., x_n) \\ \dots \\ x_n \mapsto f_n(x_1, ..., x_n) \end{cases},$$

où chaque $f_i : \{0,1\}^n \to \{0,1\}, 1 \leq i \leq n$ est une fonction locale de transition déterminant l'évolution de la variable x_i.

Exemple 2.2. Un réseau booléen possible est :

$$\mathrm{F}(x) = \begin{cases} x_1 \mapsto x_2 \vee x_3 \\ x_2 \mapsto \neg x_3 \\ x_3 \mapsto x_1 \wedge \neg x_2 \end{cases}. \tag{1.1}$$

L'évolution de sa dynamique correspond à des *transitions* sur *l'espace d'états* boo-

léens des variables X .

Définition 2.3. L'*espace d'états* d'un réseau est l'ensemble de toutes les interprétations possibles sur l'ensemble de variables X, et sera noté $S_X = (X \mapsto \mathbb{B})$.

Exemple 2.3. $s = (0,0,0)$ est un état dans lequel $s(x_1) = 0$, $s(x_2) = 0$ et $s(x_3) = 0$. L'espace d'états du réseau F (1.1) donné en exemple est :

$$S_{\{x_1,x_2,x_3\}} = \{(0,0,0),(0,0,1),(0,1,0),(0,1,1),(1,0,0),(1,0,1),(1,1,0),(1,1,1)\}.$$

L'évolution de chaque variable x_i dépend de l'application de sa fonction locale de transition f_i sur un état s, le tableau 1.1 montre le résultat de l'application des fonctions locales de transition du réseau booléen F (1.1) sur chaque élément de l'espace d'états S_X.

$s = (x_1, x_2, x_3)$	$f_1(s)$	$f_2(s)$	$f_3(s)$
(0,0,0)	0	1	0
(0,0,1)	1	0	0
(0,1,0)	1	1	0
(0,1,1)	1	0	0
(1,0,0)	0	1	1
(1,0,1)	1	0	1
(1,1,0)	1	1	0
(1,1,1)	1	0	0

Tableau 1.1 – Application des fonctions locales de transition f_i du réseau F sur les éléments de l'espace d'état S_X.

Le comportement dynamique global du réseau correspond à une *relation de transition* entre les états et dépend d'une loi décrivant la mise à jour des variables, appelée *mode*. Deux catégories de mode de mise à jour existent : les déterministes, où la relation de transition est une fonction et les non-déterministes, où la relation de transition est une relation quelconque (non-fonctionnelle). Nous nous intéresserons dans cette thèse à une catégorie de modes de mise à jour classiquement utilisé en modélisation des réseaux biologiques. Ces modes sont définis comme des couvertures des variables, $M \subseteq \mathbf{2}^X$ et $\bigcup_{m_i \in M} m_i = X$. Ils s'appliquent en chaque état de manière identique pour le

calcul de la transition. Le lecteur pourra se referer au livre [85] et à l'article [78] pour une description plus générale des modes de mise à jour.

La dynamique globale se représente par un système de transition étiqueté, comme suit :

Définition 2.4. Le système de transition étiqueté suivant $\langle \mathrm{S_X}, \longrightarrow, \mathrm{M} \rangle$ décrit les trajectoires possibles entre les états du système, tel que :

— $\mathrm{S_X} = (\mathrm{X} \mapsto \mathbb{B})$ est l'espace d'états ;

— $\longrightarrow \subseteq \mathrm{S_X} \times \mathrm{M} \times \mathrm{S_X}$ est la *relation de transition* dont chaque élément est étiqueté par une ensemble de variables $m \in \mathrm{M}$ correspondant aux variables mises à jour lors d'une transition et définie par :

$$s \xrightarrow{m} s' \stackrel{\text{def}}{=} \forall x_i \in \mathrm{X} : s'(x_i) = \begin{cases} f_i(x) \text{ si } x_i \in m \\ s(x_i) \text{ sinon} \end{cases}.$$

Concernant les modes, deux modes sont privilégiés pour la modélisation des réseaux biologiques : le mode *asynchrone* $\mathrm{M} = \{\{x_i\}\}_{x_i \in \mathrm{X}}$ où une seule variable est mise à jour par transition, et le mode *parallèle*, $\mathrm{M} = \{\mathrm{X}\}$, où toutes les variables sont mises jour lors d'une transition. Remarquons que le mode asynchrone est un mode non-déterministe et le mode parallèle est déterministe.

Le système de transition étiqueté se représente graphiquement par un *graphe de transition*.

Exemple 2.4. Le graphe de transition asynchrone (dont on a omis les transitions d'un état sur lui-même) du réseau booléen F (1.1) est représenté dans la figure 1.7 et le graphe de transition parallèle de ce même réseau est donné dans la figure 1.8.

Définition 2.5 (Équilibre, cycle limite et état stable)**.** Un équilibre est défini à partir de la fermeture transitive $\longrightarrow^*$ de la relation de transition. Un «état s est un équilibre», noté $\mathrm{E}q(s)$, si et seulement si :

$$\mathrm{E}q(s) \stackrel{\text{def}}{=} \forall s' \in \mathrm{S_X} : (s \longrightarrow^* s') \implies (s' \longrightarrow^* s).$$

Un *cycle limite* est un ensemble d'états A de $\mathrm{S_X}$ tel que chaque état de A est un équilibre et peut être atteint à partir de tout élément de A. Formellement, un cycle

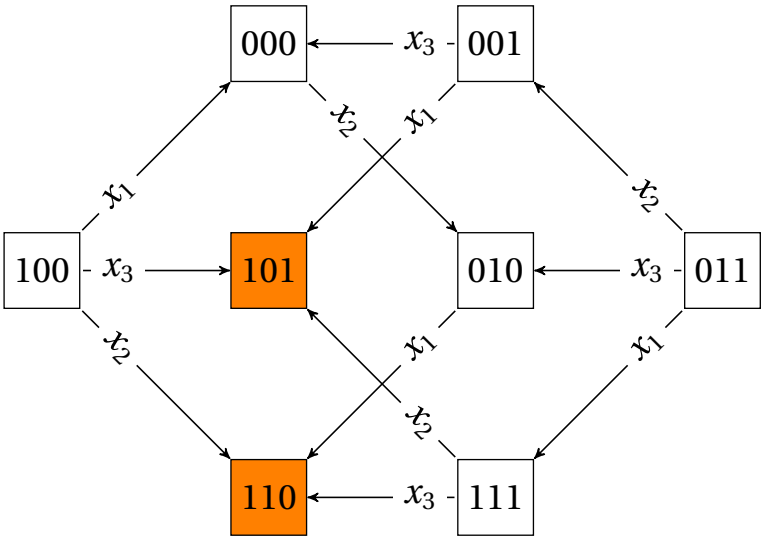


Figure 1.7 – Graphe de transition asynchrone de F. Les transitions d'un état sur lui-même ont été omises.

limite $A \subseteq S_X$ se définit par :

$$\forall s \in A, \forall s' \in A : Eq(s) \wedge Eq(s') \wedge (s \longrightarrow^* s').$$

Un cycle limite correspond à une composante fortement connexe (CFC) terminale du graphe de transition.

Un *état stable* est un cycle limite de cardinalité 1, qui peut être défini directement à partir d'un réseau booléen F de la façon suivante :

$$\text{STBL}_F(s) \stackrel{\text{def}}{=\!=} \forall 1 \leq i \leq n : f_i(s) = s(x_i) \tag{1.2}$$

Par conséquent, un état stable est une CFC terminale triviale du graphe de transition.

Exemple 2.5. Sur les figures 1.7 et 1.8, les états stables sont colorés en orange et les cycle limites en jaune.

Le *graphe d'interaction* d'un réseau booléen est un graphe dirigé qui est une abstraction du réseau dans laquelle sont représentées les dépendances entre les variables.

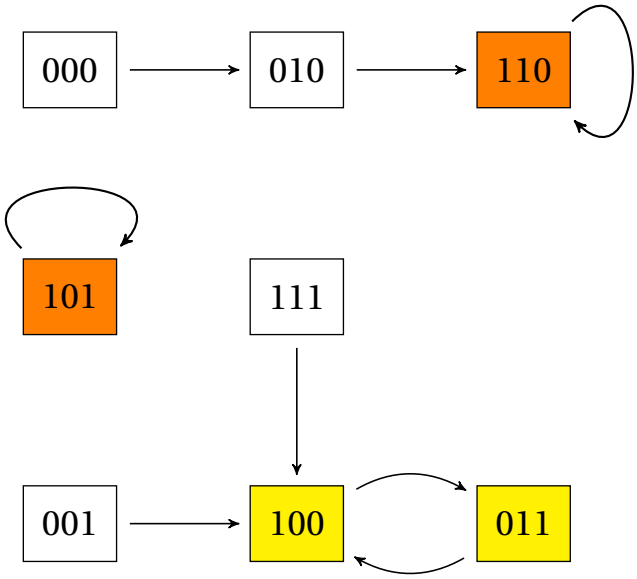


Figure 1.8 – Graphe de transition parallèle de F. Les transitions ne sont pas étiquetées car tous les agents sont mis à jour à chaque transition. Les états stables sont colorés en orange et le cycle limite en jaune.

Définition 2.6. Une *interaction* entre deux agents x_i et x_j est une relation définie par :

$$x_i \longrightarrow x_j \overset{\text{def}}{=} \exists s, s' \in \mathrm{S_X} : s(x_i) \neq s'(x_i) \wedge f_j(s) \neq f_j(s') \wedge$$
$$\forall x_k \in \mathrm{X} \setminus x_i : s(x_k) = s'(x_k)$$

Les *interactions signées* précisent la nature de l'influence et sont définies comme suit :

positive monotone :

$$x_i \xrightarrow{1} x_j \overset{\text{def}}{=} x_i \longrightarrow x_j \wedge (\forall s, s' \in \mathrm{S_X} : s(x_i) \leq s'(x_i) \wedge$$
$$\forall x_k \in \mathrm{X} \setminus x_i : s(x_k) = s'(x_k) \implies f_j(s) \leq f_j(s'))$$

négative monotone :

$$x_i \xrightarrow{-1} x_j \overset{\text{def}}{=} x_i \longrightarrow x_j \wedge (\forall s, s' \in \mathrm{S_X} : s(x_i) \leq s'(x_i) \wedge$$
$$\forall x_k \in \mathrm{X} \setminus x_i : s(x_k) = s'(x_k) \implies f_j(s) \geq f_j(s'))$$

non-monotone :

$$x_i \xrightarrow{0} x_j \overset{\text{def}}{=} x_i \longrightarrow x_j \wedge \neg(x_i \xrightarrow{+} x_j) \wedge \neg(x_i \xrightarrow{-} x_j)$$

Le graphe d'interaction *signé* représente les interactions décorées d'un signe préci-

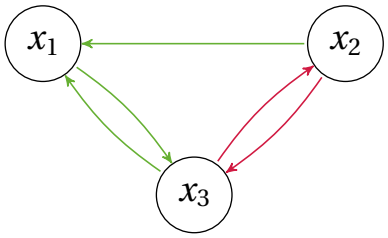


Figure 1.9 – Graphe d'interaction signé de F où les interactions positives monotones sont en vert et les interactions négatives monotones en rouge.

sant leur nature : positive +, négative − et non-monotone ±. Il peut également être déduit de la définition symbolique du réseau booléen si les formules sont bien formées [1] de telle façon que :

— une interaction est positive $x_i \xrightarrow{+} x_j$ si et seulement si le littéral x_i apparaît dans la formule définissant f_j et le littéral $\neg x_i$ n'apparaît pas dans la formule définissant f_j

— une interaction est négative $x_i \xrightarrow{-} x_j$ si et seulement si le littéral $\neg x_i$ apparaît dans la formule définissant f_j et le littéral x_i n'apparaît pas dans la formule définissant f_j

— une interaction est non monotone $x_i \xrightarrow{\pm} x_j$ si et seulement si les littéraux x_i et $\neg x_i$ apparaissent dans la formule définissant f_j.

Exemple 2.6. Le graphe d'interaction G(F) signé du réseau F (1.1) est donné dans la figure 1.9. Les interactions sont toutes monotones, les positives sont colorées en vert et les négatives en rouge.

2.2 Réseaux booléens contrôlés

Dans cette section, nous définissons formellement les réseaux booléens contrôlés (RBC), puis nous montrons comment nous modélisons les perturbations structurelles du graphe d'interaction par des paramètres de contrôle d'usage et de définition.

Définition 2.7. Un *réseau booléen contrôlé* étend le réseau booléen par l'ajout de *paramètres de contrôle*, qui sont des variables booléennes $u_j \in \mathrm{U}, 1 \leq j \leq m$, dans les fonctions locales de transition, c'est-à-dire :

1. En forme normale disjonctive minimale, définie au chapitre 2 dans la section 1

$$\mathrm{F}(x,u) = \begin{cases} x_1 \mapsto f_1(x_1,...x_n,u_1,...,u_m) \\ \dots \\ x_n \mapsto f_n(x_1,...x_n,u_1,...,u_m) \end{cases}.$$

Ainsi, un RBC à n variables et à m paramètres de contrôle est défini comme une fonction générant un réseau booléen à partir d'une valuation des paramètres de contrôle : $\mathrm{F_U} : \mathbb{B}^m \to (\mathbb{B}^n \to \mathbb{B}^n)$.

Exemple 2.7. Une extension du réseau booléen F donné en exemple dans la section précédente, à un RBC en ajoutant les paramètres de contrôle $\mathrm{U} = \{u_1, u_2, u_3, u_4\}$ est :

$$\mathrm{F}(x,u) = \begin{cases} x_1 \mapsto (x_2 \wedge u_1) \vee x_3 \\ x_2 \mapsto \neg(x_3 \vee \neg u_2) \\ x_3 \mapsto ((\neg x_2 \wedge x_1) \vee \neg u_3) \wedge u_4 \end{cases}. \tag{1.3}$$

Définition 2.8. Une *entrée de contrôle* est une interprétation sur l'ensemble des paramètres de contrôle : $\mu : \mathrm{U} \mapsto \mathbb{B}$.

L'application d'une entrée de contrôle $\mu \in \mathbb{B}^m$ modifie les fonctions locales de transition et par conséquent la dynamique globale du réseau.

Exemple 2.8. L'application de l'entrée de contrôle $\mu_1 = (0,1,1,1)$ au RBC 1.3 engendre le réseau présenté à la figure 1.10 et a pour conséquence d'inverser le sens de la transition $(0,1,0) \longrightarrow (1,1,0)$ dans le graphe de transition et un changement d'état stable avec $(0,1,0)$ qui devient stable et $(1,1,0)$ qui ne l'est plus.

2.3 Contrôle par gel

Les perturbations liées aux mutations ou à l'action des médicaments sont représentées par une catégorie particulière de contrôle appelée *contrôle par gel* de telle façon que l'activation du contrôle fige les états des variables à une valeur spécifique. On distingue le *gel de la définition* (gel-D) qui contrôle une fonction locale de transition du *le gel d'usage* (gel-U) qui contrôle une variable dans une fonction locale de transition

$$\mu_1 = (0,1,1,1)$$

$$F(x,\mu_1) = \begin{cases} x_1 \mapsto x_3 \\ x_2 \mapsto \neg x_3 \\ x_3 \mapsto (\neg x_2 \wedge x_1) \end{cases}$$

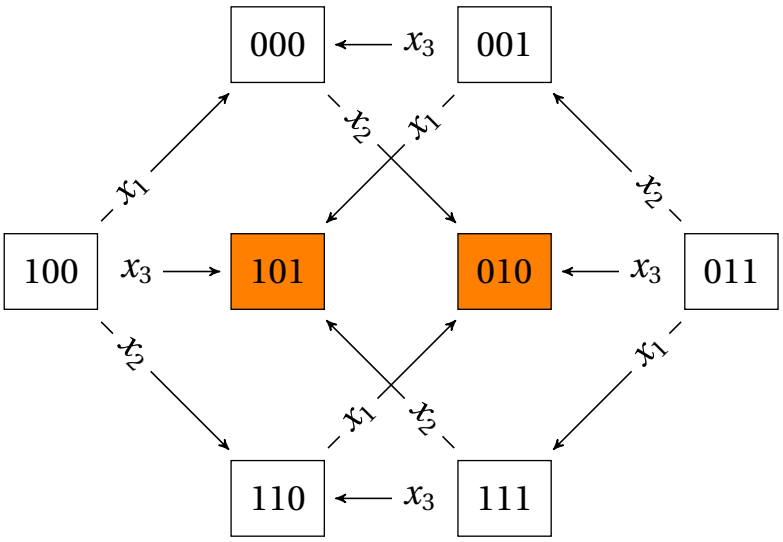


Figure 1.10 – Réseau $F(x,\mu_1)$ résultant de l'application de l'entrée de contrôle μ_1 et son graphe de transition asynchrone.

définissant la mise à jour d'une autre variable. Le paramètre gel-D régissant le gel de f_i est désigné par d_i et le paramètre gel-U est désigné par $u_{i,j}$ pour le contrôle par gel de la variable x_i dans son utilisation dans f_j. De plus, chaque paramètre de contrôle a deux régimes distincts : soit il fige la variable à une valeur spécifique, soit il reste inactif. La convention, inspirée par la température de gel de l'eau 0°C, est la suivante : l'action de gel est déclenchée lorsque le paramètre de contrôle est figé à 0 alors que la situation d'inactivité correspond à 1. Comme la valeur d'un paramètre indique l'activité de gel (active ou inactive), les deux résultats de gel possibles 0 ou 1 sont supportés par deux paramètres distincts notés respectivement $d_i^0, u_{i,j}^0$ et $d_i^1, u_{i,j}^1$.

Définition 2.9. Le *contrôle Gel-D* de i consiste à ajouter un paramètre $d_i^k, k \in \{0,1\}$ à la formule f_i de telle sorte que le réglage $\mu(d_i^k) = 0$ gèle la variable x_i à k et reste inactif sinon (c'est-à-dire $\mu(d_i^k) = 1$). La formule f_i est complétée en fonction de ce

comportement de contrôle comme suit :

$$\begin{aligned} f_i(x_1,\dots,x_n) \wedge d_i^0 \quad & \text{pour geler } i \text{ à } 0, \\ f_i(x_1,\dots,x_n) \vee \neg d_i^1 \quad & \text{pour geler } i \text{ à } 1. \end{aligned}$$

Le *contrôle Gel-U* suit les mêmes principes que le contrôle Gel-D mais s'applique à l'occurrence de variables dans les fonctions d'autres variables c'est-à-dire :

$$\begin{aligned} f_j(x_1,\dots,x_i \wedge u_{i,j}^0,\dots,x_n) \quad & \text{pour geler } x_i \text{ à } 0 \text{ dans } f_j, \\ f_j(x_1,\dots,x_i \vee \neg u_{i,j}^1,\dots,x_n) \quad & \text{pour geler } x_i \text{ à } 1 \text{ dans } f_j. \end{aligned}$$

Exemple 2.9. Dans l'exemple 1.3, u_1 est assimilé au paramètre de gel-U de x_2 à 0 ($u_1 = u_{2,1}^0$) utilisé dans la définition de x_1, u_2 peut être interprété comme le paramètre de gel-U de x_3 ($u_2 = u_{3,2}^1$), et u_3, u_4 sont les paramètres gel-D de x_3 gelant la variable à 1 et 0 respectivement ($u_3 = d_3^1, u_4 = d_3^0$). Par conséquent, le RBC peut être réécrit en utilisant la convention d'appellation appropriée comme suit :

$$\mathrm{F}_{(u_{2,1}^0,d_2^0,d_3^2,d_3^1)} = \begin{cases} x_1 \mapsto \left(x_2 \wedge u_{2,1}^0\right) \vee x, \\ x_2 \mapsto \neg(x_3 \vee \neg u_{3,2}^1) \\ x_3 \mapsto \left((\neg x_2 \wedge x_1) \vee \neg d_3^1\right) \wedge d_3^0 \end{cases} . \tag{1.4}$$

Effet du contrôle par gel sur le graphe d'interaction. Le contrôle par gel-D et gel-U défini ainsi modélise les effets dynamiques des perturbations des sommetset des arcs du graphe d'interaction. Ainsi, lorsqu'un paramètre de gel-D d_i est activé, cela a pour conséquence de geler la valeur de i et donc de supprimer tous les arcs entrants du sommet i du reste du graphe d'interaction, tandis que l'activation d'un paramètre de gel-U $u_{i,j}$ a pour conséquence de supprimer l'arc $i \to j$ du graphe d'interaction. La figure 1.11 montre les effets de l'activation individuelle du paramètre de gel-U $u_{2,1}^0$ et celle du paramètre de gel-D d_3^0 sur le réseau booléen et son graphe d'interaction.

Les RBC constituent un cadre formel discret pour la reprogrammation comportementale des réseaux booléens et la catégorie de contrôle par gel permet de modéliser des perturbations des sommets et des interactions du graphe d'interaction du réseau

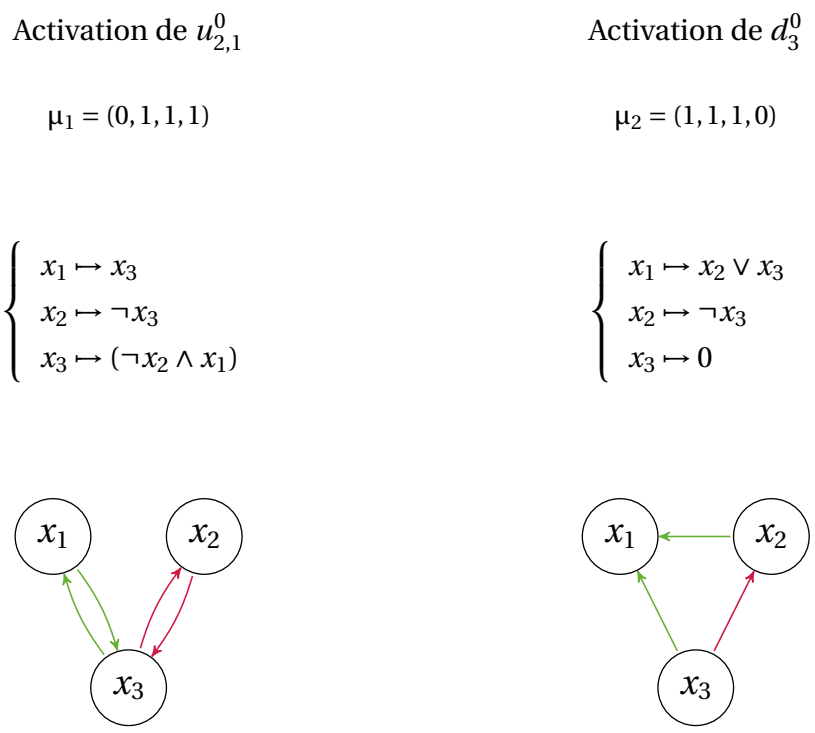


Figure 1.11 – Suppression d'arcs dans le graphe d'interaction résultant de l'activation des paramètres de contrôle de gel-U et gel-D.

booléen. Dans la suite, nous définissons *des modes de reprogrammation* décrivant les effets de ces perturbations des fonctions locales sur la dynamique globale du réseau booléen.

3 Spécification de la reprogrammation

La détermination des *paramètres de contrôle causaux* des changements comportementaux requiert de définir des *modes de reprogrammation* définissant les effets attendus ou les objectifs de cette reprogrammation. Nous définissons deux modes de reprogrammation : la *possibilité* et la *nécessité* d'une propriété rencontrée aux états stables du réseau.

3.1 Modes de reprogrammation

Soit F un RBC à n variables d'état, contrôlé par un ensemble U de m paramètres de contrôle. Soit S_X l'ensemble des états du réseau F et p, une *propriété*, correspondant à

une fonction booléenne associant une valeur de vérité à l'état s.

La possibilité est un mode de reprogrammation défini par un prédicat spécifiant que la propriété p doit être vraie dans **au moins un état stable** du réseau booléen F reprogrammé par une entrée de contrôle μ.

Définition 3.1. Le mode de reprogrammation par *possibilité* de p (notée $\Diamond p$) se définit par :

$$\exists s \in \mathrm{S_X} : \mathrm{STBL}_{\mathrm{F}_\mu}(s) \wedge p(s) \qquad (\Diamond p)$$

Exemple 3.1. Considérons le RBC 1.3 donné en exemple plus haut. Le réseau booléen $\mathrm{F}_{(1,1,1,1)}$ sans contrôle actif a deux états stables $(1,0,1)$ et $(1,1,0)$ et le réseau booléen $\mathrm{F}_{(0,1,1,1)}$ dont le paramètre de contrôle $u^0_{2,1}$ a été activé a deux états stables $(1,0,1)$ et $(0,1,0)$. Le réseau $\mathrm{F}_{(0,1,1,1)}$ satisfait $\Diamond p$ avec $p = \neg x_1 \wedge x_2 \wedge \neg x_3$ tandis que le réseau $\mathrm{F}_{(1,1,1,1)}$ ne le satisfait pas.

La nécessité est un mode de reprogrammation défini par un prédicat spécifiant que la propriété p doit être vraie **dans tous les états stables** du réseau booléen reprogrammé par une entrée de contrôle μ.

Définition 3.2. Le mode de reprogrammation par *nécessité* de p (notée $\Box p$) se définit par :

$$\forall s \in \mathrm{S_X} : \mathrm{STBL}_{\mathrm{F}_\mu}(s) \implies p(s) \qquad (\Box p)$$

Exemple 3.2. Le réseau $\mathrm{F}_{(0,1,1,1)}$ satisfait $\Box \neg q$ avec $q = x_1 \wedge x_2 \wedge \neg x_3$ tandis que le réseau $\mathrm{F}_{(1,1,1,1)}$ ne le satisfait pas.

Ces modes de reprogrammation permettent de spécifier des *requêtes* définissant l'effet attendu par le contrôle de variables. Trois types de requêtes seront utilisés par la suite : la possibilité d'obtenir une propriété p à l'état stable $\Diamond p$, la nécessité de l'obtenir

$\Box p$ et la nécessité de l'obtenir en imposant qu'un état stable existe $\Diamond p \wedge \Box p$. En effet, par définition, la nécessité seule n'impose pas l'existence d'un état stable.

3.2 Noyaux de reprogrammation

Une solution d'une reprogrammation vérifiant une requête spécifiée en utilisant l'un des modes se nomme le *noyau de reprogrammation*. Ces solutions consituent des solutions *causales minimales* pour la reprogrammation. En effet, de nombreuses entrées de contrôle permettent de vérifier une requête. Le tableau 1.2 présente les ensembles d'états stables des réseaux booléens résultant de l'application exhaustive de toutes les entrées de contrôle possibles. Cependant, on remarquera que certaines étendent un profil de contrôle actif (variables de contrôle à 0) comme par exemple la première ligne de contrôle pour la nécessité. Dans ce cadre, le noyau est le contrôle minimal pour vérifier la requête, incarnant en modélisation les causes essentielles.

$\mu = (u^0_{2,1}, u^1_{3,2}, d^1_3, d^0_3)$	Etats stables	$\Diamond p$	$\Box q$
(0,0,0,0)	{(000)}	Faux	Vrai
(0,0,0,1)	{(101)}	Faux	Vrai
(0,0,1,0)	{(000)}	Faux	Vrai
(0,0,1,1)	{(000),(101)}	Faux	Vrai
(0,1,0,0)	{(010)}	Vrai	Vrai
(0,1,0,1)	{(101)}	Faux	Vrai
(0,1,1,0)	{(010)}	Vrai	Vrai
(0,1,1,1)	{(101),(010)}	Vrai	Vrai
(1,0,0,0)	{(000)}	Faux	Vrai
(1,0,0,1)	{(101)}	Faux	Vrai
(1,0,1,0)	{(000)}	Faux	Vrai
(1,0,1,1)	{(000),(101)}	Faux	Vrai
(1,1,0,0)	{(110)}	Faux	Faux
(1,1,0,1)	{(101)}	Faux	Vrai
(1,1,1,0)	{(110)}	Faux	Faux
(1,1,1,1)	{(101),(110)}	Faux	Faux

Tableau 1.2 – Evaluation exhaustive de l'effet des entrées de contrôle sur les états stables et la valeur de vérité de $\Diamond p$ et $\Box q$.

Les entrées de contrôle vérifiant un mode de reprogrammation sont composées de

paramètres de contrôle actifs (dont la valeur est 0) et de paramètres de contrôle inactifs (dont la valeur est 1). Afin de restreindre les solutions aux *paramètres de contrôle causaux*, on définit que les solutions de la reprogrammation doivent être *minimales* en paramètres de contrôle actifs.

Définition 3.3. Un *noyau de reprogrammation* est un ensemble minimal pour l'inclusion de paramètres de contrôle actifs et vérifiant une requête donnée.

Exemple 3.3. Trois différentes entrées de contrôle peuvent convenir comme solution pour le mode de reprogrammation $\Diamond p$, à savoir :

$$(0,1,1,1),(0,1,1,0),(0,1,0,0).$$

Dans ces trois entrées de contrôle, on peut distinguer les paramètres de contrôle qui doivent nécessairement être activés, paramètres causaux, c'est le cas par exemple de $u_{2,1}$ pour le mode $\Diamond p$, des paramètres de contrôle non-causaux. Les paramètres de contrôle non causaux peuvent se diviser en deux classes : les paramètres de contrôle qui peuvent être indifféremment valués à 0 ou à 1, c'est le cas de d_3^1 et de d_3^0, et ceux qui doivent être nécessairement inactifs, c'est le cas de $u_{3,2}^1$. Plusieurs noyaux peuvent être trouvés pour un problème donné. Par exemple, trois noyaux différents $\{d_3^1\},\{u_{2,1}^0\},\{u_{3,2}^1\}$ permettent la perte de l'état stable 110 vérifiant $\Box q$.

4 Bilan du chapitre

La reprogrammation comportementale des systèmes dynamiques booléens s'appuie sur les RBC qui modélisent l'action de perturbations des réseaux booléens. Ces perturbations ont pour objectif de dévier le comportement à l'équilibre du réseau pour vérifier des requêtes spécifiées selon deux modes de reprogrammation : la possibilité ou la nécessité de rencontrer une propriété à l'état stable. Les paramètres de contrôle sont de deux types : les paramètres en Gel-D qui gèlent les valeurs des variables du réseaux et les paramètres en Gel-U qui gèlent les valeurs des interactions entre les variables du réseau. Ces contrôles représentent l'action de la maladie sur la cellule provoquées par des mutations ou celle des médicaments altérant l'expression ou l'activité

d'une molécule. Le noyau détermine les causes essentielles conduisant à cette reprogrammation. L'identification complète et correcte des noyaux de la reprogrammation vise à trouver tous les noyaux de reprogrammation pour une requête donnée et seulement ceux-là. Cette identification requiert des méthodes algorithmiques que nous développerons dans le chapitre suivant.

Chapitre 2

Calcul des noyaux de reprogrammation

Ce chapitre est consacré au calcul des noyaux de reprogrammation. Une méthode «force brute» consiste à tester par simulation toutes les entrées de contrôle possibles pour caractériser celles répondant à une requête. Cette méthode implique cependant de tester 2^m interprétations possibles pour m paramètres de contrôle. Le temps exponentiel de ce calcul limite la taille des réseaux booléens pouvant être traités (empiriquement $\lesssim 10$ variables d'état avec des paramètres de contrôle en Gel-D). L'inférence directe des noyaux apparait donc nécessaire pour permettre son application à de grands réseaux. Cet objectif pose la question fondamentale de définir un cadre méthodologique pour cette inférence afin d'en déduire des méthodes efficaces la calculant.

Dans la section 1, nous montrons que le calcul de tous les noyaux de reprogrammation et uniquement ceux-ci peut se définir comme un problème d'*abduction* en logique propositionnelle. Dans la section 2, nous présentons des algorithmes permettant le calcul d'impliquants premiers, qui sont un élément fondamental de l'abduction et à partir desquels sont développées les méthodes d'inférence des noyaux de reprogrammation présentées dans la section 3. Enfin, dans la section 4, nous comparons ces méthodes à des méthodes récentes publiées dans la littérature.

1 Formulation abductive du calcul des noyaux

L'*abduction* est un mode de raisonnement logique correspondant à *l'inférence d'une hypothèse explicative de faits à partir de ce qui est déjà connu (la théorie).*

Nous rappellerons des définitions de logique propositionnelle en concluant par celles relatives à l'abduction en logique propositionnelle, puis nous démontrerons comment le calcul d'un noyau de reprogrammation s'exprime par l'inférence d'un *impliquant premier*. Pour illustrer ces définitions, nous utiliserons la fonction booléenne définie par la table de vérité donnée dans le tableau 2.1.

x_1	x_2	x_3	$\phi(x_1, x_2, x_3)$
0	0	0	1
0	0	1	0
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	0
1	1	0	1
1	1	1	1

Tableau 2.1 – Table de vérité de la fonction booléenne $\phi : \{0,1\}^3 \to \{0,1\}$ telle que $\phi(x) = x_2 \veebar (\neg x_1 \wedge \neg x_2 \wedge \neg x_3)$.

1.1 Abduction et logique propositionnelle

Les définitions que nous donnons dans ce chapitre sont adaptées de [26] pour la logique propositionnelle et de [64] pour l'abduction.

Soit $X = \{x_1, \ldots, x_n\}$ un ensemble de variables booléennes. Rappelons qu'une *interprétation* correspond à l'assignation d'une valeur booléenne à chaque élément x_i de X, c'est-à-dire, $I : X \mapsto \mathbb{B}$. On notera $\phi(I)$ l'évaluation de ϕ par rapport à une interprétation I sur ses variables et $V(I)$ les variables sur lesquelles I est définie.

Définition 1.1. Une interprétation I est un *modèle* d'une formule ϕ si et seulement si $\phi(I)$ est vraie, qui est noté $I \models \phi$.

Une *tautologie* est une formule dont toutes les interprétations sont des modèles et une *contradiction* est une formule dont aucune interprétation n'est un modèle. Une formule est dite *satisfaisable* si elle possède au moins un modèle.

Exemple 1.1. $I = (x_1 \mapsto 0, x_2 \mapsto 1, x_3 \mapsto 0)$ est un modèle de la formule ϕ, qui est donc satisfaisable. La formule $x_1 \vee \neg x_1$ est une tautologie et la formule $x_1 \wedge \neg x_1$ est une contra-

diction.

Définition 1.2. Soient ϕ_1 et ϕ_2 deux formules booléennes. On dit que ϕ_2 est une *conséquence logique* de ϕ_1, que l'on note $\phi_1 \models \phi_2$ [1], si tout modèle de ϕ_1 est aussi modèle de ϕ_2, c'est à dire :

$$\phi_1 \models \phi_2 \stackrel{\text{def}}{=} \forall I \in (X \to \mathbb{B}) : I \models \phi_1 \implies I \models \phi_2.$$

Exemple 1.2. ϕ est une conséquence logique de $x_2 \wedge (x_1 \vee \neg x_1)$. En effet, cette formule a quatre modèles sur $X = \{x_1, x_2, x_3\} : (0,1,0), (0,1,1), (1,1,0)$ et $(1,1,1)$ qui sont tous des modèles de ϕ.

En logique propositionnelle, l'abduction consiste donc à proposer les *causes* «les plus explicatives» dont les *faits observables* sont la conséquence logique et en prenant en considération une connaissance constituant la *théorie*. Plus formellement, si on pose T une formule spécifiant la *théorie*, f une formule spécifiant les *observations* et la formule H désignant une conjonction de faits constituant la *cause* de f, on note :

$$H, T \models f.$$

Afin d'éviter les conjonctions de faits ne vérifiant pas la théorie, la cause abductive H doit être *cohérente* avec la théorie.

Définition 1.3. Une cause abductive H de f est *cohérente* avec la théorie T si et seulement si $T \wedge H$ est satisfaisable.

Enfin, afin d'éliminer les causes trop générales, on contraint une cause abductive H à être *minimale*.

Définition 1.4. Une cause abductive H de f est *minimale* si, pour toute cause abductive H' de f cohérente avec T et telle que $H \models H'$ on a également $H' \models H$.

En logique propositionnelle, les causes abductives minimales et cohérentes sont assimilées à des *cubes* qui sont des *impliquants premiers* de la formule booléenne f représentant les faits.

1. On peut noter ici que le symbole $\models$ est surchargé sur les formules.

Définition 1.5. Un *littéral positif* (respectivement *littéral négatif*) est une formule de la forme x_i (respectivement $\neg x_i$). Un *cube* est une conjonction élémentaire de littéraux, c'est-à-dire, une formule de la forme $C = l_1 \wedge l_2 \wedge ... \wedge l_n$, notée $\bigwedge_{i=1}^{n} l_i$ et une *clause* est une disjonction élémentaire de littéraux, c'est-à-dire, une formule de la forme $D = l_1 \vee l_2 \vee ... \vee l_n$, notée $\bigvee_{i=1}^{n} l_i$. Une cube ou une clause peuvent être représentés par un ensemble de littéraux.

Définition 1.6. Soit ϕ une formule propositionnelle et C un cube, on dit que C est un *impliquant* de ϕ si et seulement si $C \models \phi$ et C n'est pas une contradiction.

Exemple 1.3. Les formules $(\neg x_1 \wedge \neg x_2 \wedge \neg x_3)$ et x_2 sont deux impliquants de ϕ.

Définition 1.7. Soit C un impliquant d'une formule ϕ, on dit que C est un *impliquant premier* de ϕ si tout cube obtenu en éliminant un littéral arbitraire de C n'est pas un impliquant, c'est-à-dire, $\forall l \in C : C \setminus l \not\models \phi$. On notera $PI(\phi)$ l'ensemble des impliquants premiers de la formule ϕ.

Exemple 1.4. x_2 est un impliquant premier de ϕ. La formule $(\neg x_1 \wedge \neg x_2 \wedge \neg x_3)$ est un impliquant de ϕ mais n'est pas premier car le cube $(\neg x_1 \wedge \neg x_3)$ obtenu en éliminant $\neg x_2$ de la formule est un impliquant de ϕ.

1.2 Inférence abductive des noyaux

Un noyau de reprogrammation, tel que nous l'avons défini au chapitre 1, est un ensemble de paramètres actifs de contrôle minimal par inclusion et vérifiant un mode de reprogrammation. Nous montrons ici que le calcul de l'ensemble des noyaux de reprogrammation selon les modes de reprogrammation $\Diamond$ et $\Box$ s'exprime par un problème d'inférence abductive en logique propositionnelle (Lemme 1.1)

Définition 1.8. Le *minterm* C_I d'une interprétation I est l'unique cube tel que $V(I) = V(C_I)$ et $I \models C_I$.

Exemple 1.5. Le *minterm* de $I = (x_1 \mapsto 0, x_2 \mapsto 1, x_3 \mapsto 0)$ est $C_I = \neg x_1 \wedge x_2 \wedge \neg x_3$.

Lemme 1.1. Trouver une entrée de contrôle μ telle que :

$$\exists s \in S : \mathrm{STBL}_{F_\mu}(s) \wedge p(s) \qquad (\Diamond p)$$

est équivalent à trouver un cube C_μ tel que :

$$C_\mu \wedge C_s \models (\text{STBL}_{F_U} \wedge p),$$

où $(C_s \wedge C_\mu)$ *est un impliquant des observations* $(\text{STBL}_{F_U} \wedge p)$ *et en considérant la théorie* T *comme étant toujours vraie.*

Trouver une entrée de contrôle μ telle que :

$$\forall s \in S : \text{STBL}_{F_\mu}(s) \implies p(s) \qquad (\Box p),$$

est équivalent à trouver un impliquant C_μ tel que :

$$C_\mu \models (\text{STBL}_{F_U} \implies p),$$

où la formule spécifiant les observations $(\text{STBL}_{F_U} \implies p)$ *est restreinte aux paramètres de contrôle et en considérant la théorie* T *comme étant toujours vraie.*

Démonstration. Cette démonstration requiert de :

1. Transformer les conditions de stabilité $(\text{STBL}_{F_\mu}(s))$ et de propriété $(p(s))$ en formules propositionnelles (STBL_{F_U} et p).
2. Démontrer l'équivalence entre la formulation en logique des prédicats de la possibilité et de la nécessité avec leur formulation en logique propositionnelle.

En détail la preuve est comme suit. Le paragraphe 1.2.1 démontre la propriété (1) et le paragraphe 1.2.2 démontre la propriété (2).

1.2.1 Condition de stabilité et propriété en formule propositionnelle

La transformation sous forme de logique propositionnelle de la condition de stabilité peut s'exprimer de la façon suivante : l'ensemble d'équations étant fini, on peut réécrire la définition d'un état stable donnée au chapitre 1 (voir définition 1.2) par la

conjonction :

$$\bigwedge_{i=1}^{n} (x_i = f_i(x)).$$

L'équivalence $p_1 \iff p_2$ étant satisfaite si et seulement si $p_1(\mathrm{I}) = p_2(\mathrm{I})$, cet ensemble d'égalité peut être reformulé par un ensemble d'équivalence, définissant finalement la condition de stabilité par la formule propositonnelle suivante :

$$\mathrm{STBL_F} \stackrel{\text{def}}{=\!=} \bigwedge_{i=1}^{n} \left(x_i \iff f_i(x)\right).$$

Cette formulation peut être étendue aux RBC, où $\mathrm{STBL_{F_U}}$ contient des paramètres de contrôle.

La condition de propriété peut s'exprimer par le minterm de $p(s)$ noté p.

1.2.2 Formulation abductive de la possibilité et de la nécessité.

– $\Diamond p$ peut être reformulé en introduisant le symbole $\models$ de la façon suivante :

$$\exists s \in \mathrm{S_X} : (\mathrm{C}_s \wedge \mathrm{C}_\mu) \models \mathrm{STBL_{F_\mu}} \wedge p,$$

où C_s et C_μ sont les minterms d'un état s et d'une entrée de contrôle μ respectivement. Comme $\mathrm{STBL_{F_\mu}} \wedge p$ contient des variables d'états et des paramètres de contrôle, un impliquant C vérifiant $\mathrm{C} \models \mathrm{STBL_{F_\mu}} \wedge p$ peut-être divisé en $\mathrm{C} = \mathrm{C}_s \wedge \mathrm{C}_\mu$. Déterminer l'existence d'un impliquant C contenant des variables d'états X et de contrôle U est équivalent à prouver l'existence de son interprétation. Le quantificateur existentiel de s peut-être donc être supprimé et on obtient :

$$(\mathrm{C}_s \wedge \mathrm{C}_\mu) \models \mathrm{STBL_{F_\mu}} \wedge p.$$

– $\Box p$ peut être reformulé en introduisant le symbole $\models$ de la façon suivante :

$$\forall s \in \mathrm{S_X} : (\mathrm{C}_s \wedge \mathrm{C}_\mu) \models \mathrm{STBL_{F_\mu}} \wedge p,$$

où C_s et C_μ sont les minterms d'un état s et d'une entrée de contrôle μ respectivement.
($\Longrightarrow$) On définit $\mathrm{P} = (\mathrm{STBL_{F_\mu}} \implies p)$, s_{-i} un état dépourvu de son $i^{\text{ème}}$ composant et

$(s_{-i}, \neg s_i)$ l'état obtenu en inversant le $i^{\text{ème}}$ composant de l'état s. Soit C_μ une solution, la propriété étant vraie pour tous les états, pour tout état s il existe un état $(s_{-i}, \neg s_i)$ tels que s et $(s_{-i}, \neg s_i)$ vérifient $C_s \wedge C_\mu \wedge T \models P$ et $C_{(s_{-i}, \neg s_i)} \wedge C_\mu \wedge T \models P$ respectivement. Par application du théorème du consensus [83], on peut déduire que $C_{s_{-i}} \wedge C_\mu$ est également un impliquant. Cette simplification peut être appliquée à tous les composants i. On obtient donc finalement :

$$\forall s_{-i} : (C_{s_{-i}} \wedge C_\mu) \wedge T \models P.$$

En suivant le même raisonnement pour tous les états s_{-i} avec suppression du composant $j \neq i$, on peut en déduire la même conclusion pour l'état $s_{-\{i,j\}} i$ dépourvu de ses composants i et j. Cela peut être répété jusqu'à ce qu'aucune variable d'état ne reste $(C_\emptyset)$. Comme $C_\emptyset = 1$ par définition, on peut conclure que :

$$C_\mu \wedge T \models P.$$

($\Longleftarrow$) Inversement, supposons que $C_\mu \wedge T \models P$ avec $V(C_\mu) = U$ et $T = 1$. On déduit que, pour tout C tel que C est cohérent avec T, on a $C \wedge C_\mu \wedge T \models P$. Pour tout $s \in S_X$, C_s ne comporte que des variables d'état, C_μ des paramètres de contrôle et T ne contient aucune variable d'état. On a alors $V(C_s) \cap V(C_\mu) \cap V(T) = \emptyset$. Deux cubes satisfaisables composés de variables disjointes sont nécessairement cohérents. On peut donc conclure que : $\forall s \in S_X : (C_s \wedge C_\mu) \wedge T \models P$.

La résolution de $\Box$ peut donc être réalisée sur les paramètres de contrôle uniquement.

□

La possibilité et la nécessité sont initialement définies en logique des prédicats du premier ordre. Le lemme 1.1 fournit une formulation équivalente en logique propositionnelle. La résolution, exprimée sous forme abductive par : $C_\mu \wedge C_s \models \Diamond p$ et $C_\mu \models \Box p$, s'exprime à présent dans le cadre de la logique propositionnelle. Ceci permet de trouver des entrées de contrôle μ en exploitant les méthodes de résolution pour l'abduction dans cette logique. Ces entrées de contrôle correspondent à une interprétation dont l'impliquant C_μ est le minterm. Toutefois, les méthodes de calcul d'impliquants n'in-

tègrent pas la minimalité requise pour calculer le noyau par rapport à l'inclusion. En nous fondant sur les méthodes de calcul des impliquants premiers ou d'impliquants, nous caractérisons de nouvelles méthodes permettant de calculer les noyaux.

2 Méthodes de calcul d'impliquants premiers

Les impliquants premiers constituent les «briques élémentaires» des fonctions booléennes, le calcul de l'ensemble des impliquants premiers d'une fonction est un problème essentiel de l'abduction et a fait l'objet de nombreux travaux qui sont présentés dans [26]. Les méthodes de calcul de l'ensemble des impliquants premiers d'une formule propositionnelle sont à la base des méthodes développées dans cette thèse pour le calcul des noyaux de reprogrammation. Dans cette section, nous donnons d'abord des définitions sur les formes normales et les impliquants premiers des formules booléennes (sous-section 2.1). Puis, nous présentons des méthodes algorithmiques permettant le calcul de tous les impliquants premiers à partir de ces différentes représentations (sous-section 2.2).

2.1 Formes normales

Nous définissons des classes de formules propositionnelles qui ont un intérêt spécifique pour le calcul des impliquants premiers, soit les formes normales disjonctives et conjonctives. Puis, nous présentons les diagrammes de décision binaire (DDB), qui permettent une représentation compacte des formules booléennes.

Définition 2.1. Une *forme normale disjonctive* (FND) est une disjonction de cubes C_k ($k = (1,2,...,m)$), c'est-à-dire, une formule de la forme :

$$\bigvee_{k=1}^{m} C_k = \bigvee_{k=1}^{m} (\bigwedge_{i=1}^{n} l_i)_k,$$

et une *forme normale conjonctive* (FNC) est une conjonction de clauses D_k ($k = (1,2,...,m)$), c'est-à-dire, une formule de la forme :

$$\bigwedge_{k=1}^{m} D_k = \bigwedge_{k=1}^{m} (\bigvee_{i=1}^{n} l_i)_k.$$

Toute fonction booléenne peut être représentée par une FNC et une FND et toute formule propositionnelle peut être transformée en FND (ou en FNC) par éliminations des opérateurs autres que $\vee$, $\wedge$ et $\neg$ puis applications successives de la loi de De Morgan et les opérations de distributivité, de commutativité et d'associativité sur $\vee$ et $\wedge$ (voir théorèmes 1.4 et 1.5 de [26]).

Exemple 2.1. On peut transformer ϕ en une FND en appliquant successivement les opérations suivantes :

1. Élimination du $\veebar$ par l'identité $A \veebar B = (A \wedge \neg B) \vee (\neg A \wedge B)$, on obtient la formule :

$$\phi = (x_2 \wedge \neg(\neg x_1 \wedge \neg x_2 \wedge \neg x_3)) \vee (\neg x_2 \wedge (\neg x_1 \wedge \neg x_2 \wedge \neg x_3)).$$

2. Application de la loi de De Morgan :

$$\phi = (x_2 \wedge (x_1 \vee x_2 \vee x_3)) \vee (\neg x_2 \wedge (\neg x_1 \wedge \neg x_2 \wedge \neg x_3)).$$

3. Application de la distributivité du $\wedge$:

$$\phi_{\text{FND}} = (x_2 \wedge x_1) \vee x_2 \vee (x_2 \wedge x_3) \vee (\neg x_1 \wedge \neg x_2 \wedge \neg x_3). \tag{2.1}$$

De même, on peut transformer ϕ en une FNC en appliquant successivement les opérations suivantes :

1. Élimination du $\veebar$ par l'identité $A \veebar B = (\neg A \vee \neg B) \wedge (A \vee B)$, on obtient la formule :

$$\phi = (\neg x_2 \vee \neg(\neg x_1 \wedge \neg x_2 \wedge \neg x_3)) \wedge (x_2 \vee (\neg x_1 \wedge \neg x_2 \wedge \neg x_3)).$$

2. Application de la loi de De Morgan :

$$\phi = (\neg x_2 \vee (x_1 \vee x_2 \vee x_3) \wedge (x_2 \vee (\neg x_1 \wedge \neg x_2 \wedge \neg x_3)).$$

3. Application de la distributivité du $\vee$ (la première clause devient une tautologie) :

$$\phi_{\text{FNC}} = (x_2 \vee \neg x_1) \wedge (x_2 \vee \neg x_3). \tag{2.2}$$

Les formules ϕ, ϕ_{FND} et ϕ_{FNC} sont toutes équivalentes en terme d'évaluation : elles représentent la même fonction booléenne. ϕ_{FND} est formée de quatre cubes $(x_2 \wedge x_1)$, x_2, $(x_2 \wedge x_3)$ et $(\neg x_1 \wedge \neg x_2 \wedge \neg x_3)$ et ϕ_{FNC} est formée de deux clauses : $(x_2 \vee \neg x_1)$ et $(x_2 \vee \neg x_3)$.

Théorème 1. Toute fonction booléenne peut être représentée par la disjonction de tous ces impliquants premiers et cette formule est appelée *forme normale disjonctive minimale* de f (voir Théorème 1.13 [26]).

Diagramme de décision binaire. Les diagrammes de décision binaire (DDB) ont été introduits par Lee [58] afin de représenter de façon compacte des fonctions booléennes. Dans ce paragraphe, nous définissons d'abord les DDB et leur représentation sous forme de table d'association, puis nous montrons comment les construire à partir d'une formule propositionnelle et d'un ordre sur les variables de la formule.

Définition 2.2. Un *diagramme de décision binaire* est un graphe enraciné, acyclique et dirigé avec un ensemble de sommets composé de deux sommets terminaux étiquetés respectivement 0 et 1 et de sommets non terminaux identifiés par un indice u, une variable $var(u)$ et les indices de leurs deux successeurs appelés $low(u)$ et $high(u)$. Parmi ces sommets non terminaux, on trouve une racine qui n'a pas de parent. Un diagramme de décision binaire est *ordonné* si tous les chemins du graphe respectent le même ordre $\preceq$ sur les variables. Un diagramme de décision binaire est *réduit* si et seulement si :

— (unicité des sommets) il n'y a pas deux sommets distincts u et v qui ont le même nom de variable et les mêmes successeurs par 0 et par 1, c'est-à-dire, si $var(u) = var(v), low(u) = low(v), high(u) = high(v)$ alors $u = v$;

— (non redondance) il n'y a pas de sommets dont les successeurs low et $high$ sont identiques.

Exemple 2.2. La figure 2.2 montre un DDB ordonné et réduit à 5 sommets et sa représentation par une table d'association entre un indice et un triplet $var, low, high$. Il est important de remarquer qu'un FND peut être obtenue à partir d'un DDBOR par la disjonction de tous les chemins menant au sommet terminal 1, où chaque chemin correspond à un cube contenant le littéral négatif de la variable associée à un sommet

indice u	$var(u)$	low(u)	high(u)
0	0	-	-
1	1	-	-
2	x_3	1	0
3	x_2	2	1
4	x_2	0	1
5	x_1	3	4

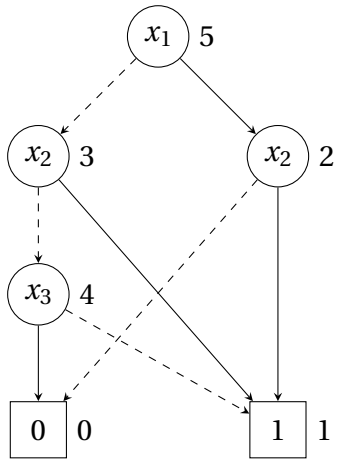


Figure 2.1 – Représentation graphique (à droite) et sous forme de table d'association (à gauche) d'un diagramme de décision binaire ordonné selon l'ordre $x_1 \prec x_2 \prec x_3$ et réduit. Les flèches en pointillés représentent la relation du sommet avec le successeur low et les flèches pleines à la relation avec le successeur $high$.

si l'on passe par son successeur low et le littéral positif de la variable associée à un sommet si l'on passe par son successeur $high$.

Les diagrammes de décision binaire ordonnés et réduits (notés DDBOR) permettent une représentation compacte des formules propositionnelles en supprimant des redondances au cours de leur construction. Les opérations logiques de conjonction, de disjonction et de négation peuvent être effectués directement sur ces graphes [16]. La construction d'un DDBOR à partir d'une formule se fait par application récursive de *l'expansion de Shannon* [26] sur la formule selon un ordre prédéfini sur les variables et des opérations de réduction.

Définition 2.3. Soit une formule booléenne ϕ et x une variable de cette formule. L'expansion de Shannon selon la variable x est définie par l'identité suivante :

$$\phi = (\neg x \wedge \phi_{\neg x}) \vee (x \wedge \phi_x),$$

où ϕ_x est le *co-facteur positif* de f par rapport à x, c'est-à-dire, la formule résultant du remplacement de x par 1 dans ϕ et $\phi_{\neg x}$ le *co-facteur négatif*, c'est-à-dire, le résultat du remplacement x par 0 dans ϕ.

Exemple 2.3. L'expansion de Shannon de la formule en FND (2.1) obtenue à la section

précédente selon x_1 est :

$$\begin{aligned} \phi &= (\neg x_1 \wedge (x_2 \vee (x_2 \wedge x_3) \vee (\neg x_2 \wedge \neg x_3))) && \text{co-facteur négatif} \\ &\quad \vee (x_1 \wedge (x_2 \vee (x_2 \wedge x_3)). && \text{co-facteur positif} \end{aligned}$$

La taille d'un DDBOR est fortement dépendante de l'ordre sur les variables. Dans la figure 2.2, nous donnons les DDBOR représentant la fonction ϕ, construit par application récursive de l'expansion de Shannon et des règles de réduction du graphe selon les règles précitées et selon deux ordres différents sur les variables. Nous donnons également les FND obtenues par disjonction des chemins menant à 1.

Théorème 2 (Théorème de canonicité). Il existe exactement un DDBOR avec un ordre donné représentant une fonction booléenne. Par extension, le DDBOR avec un ordre obtenu à partir de n'importe quelle formule propositionnelle d'une fonction est unique (pour preuve, voir [16] ou [7]).

$$\begin{aligned} \phi_{\text{FND}_1} &= (x_1 \wedge x_2) \\ &\vee (\neg x_1 \wedge x_2) \\ &\vee (\neg x_1 \wedge \neg x_2 \wedge \neg x_3) \end{aligned} \qquad \begin{aligned} \phi_{\text{FND}_2} &= x_2 \\ &\vee (\neg x_1 \wedge \neg x_2 \wedge \neg x_3) \end{aligned}$$

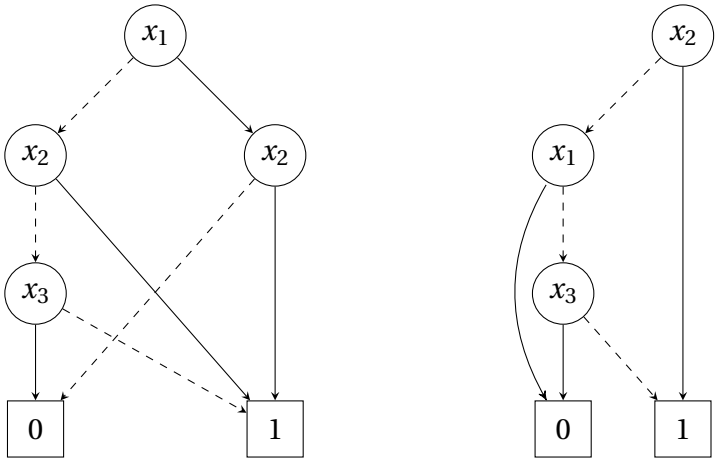


Figure 2.2 – Diagramme de décision binaire ordonné et réduit de la fonction f selon l'ordre $x_1 \prec x_2 \prec x_3$ (à gauche) et l'ordre $x_2 \prec x_1 \prec x_3$ (à droite) . Les flèches en pointillés représentent la valuation à 0 (co-facteur négatif) de la variable et les flèches pleines la valuation à 1 (co-facteur positif).

Dans la section suivante, nous montrons comment calculer les impliquants premiers d'une fonction booléenne à partir de différentes formes de la fonction.

2.2 Calcul des impliquants premiers

Ici, nous présentons les algorithmes permettant le calcul de l'ensemble des impliquants premiers d'une fonction booléenn,e soit par la procédure du consensus à partir de sa représentation en FND (sous-section 2.2.1) ou à partir d'une représentation compactée de la formule sous forme d'une diagramme de décision binaire (sous-section 2.2.2), soit par son expression comme un problème de couverture d'ensemble à partir de sa représentation en FNC (sous-section 2.2.3) .

2.2.1 Procédure du consensus

La procédure du consensus est une méthode permettant de générer tous les impliquants premiers d'une fonction booléenne à partir d'une forme normale disjonctive. Cette procédure a été découverte de façon répétée et indépendamment par différents chercheurs [83, 15, 89]. Elle consiste en la modification de la formule ϕ_{FND} par l'application répétée des opérations d'absorption et de consensus sur les cubes de ϕ. Nous définissons d'abord l'absorption et le consensus.

Définition 2.4. Les cubes des formes normales disjonctives peuvent être assimilés à des ensembles de littéraux et on dira qu'un cube C_1 *absorbe* un cube C_2 si $C_1 \subseteq C_2$.

Définition 2.5. Si $x_i \wedge C_1$ et $\neg x_i \wedge C_2$ sont deux cubes tels que $C_1 \wedge C_2$ ne contient pas de contradiction, alors on dit que $C_1 \wedge C_2$ est le *consensus* de ces deux conjonctions et que $C_1 \wedge C_2$ est *dérivé de* $x \wedge C_1$ *et* $\neg x \wedge C_2$ *par consensus sur* x.

L'algorithme du consensus est ensuite détaillé dans l'algorithme 1 où nous notons $\phi \setminus C$ la formule obtenue par la suppression du cube C de la formule ϕ.

Exemple 2.4. Considérons la formule FND (2.1) obtenue plus haut par transformation de ϕ :

$$\phi_{\text{FND}} = (x_2 \wedge x_1) \vee x_2 \vee (x_2 \wedge x_3) \vee (\neg x_1 \wedge \neg x_2 \wedge \neg x_3).$$

Algorithme 1 Procédure du Consensus

ENTRÉE: Une formule ϕ en FND
SORTIE: La formule ϕ en FND minimale, c'est-à-dire, la disjonction de tous les impliquants premiers de ϕ.

tantque une des conditions suivantes peut s'appliquer **faire**
  si il existe deux cubes C_1 et C_2 de ϕ tels que C_1 absorbe C_2 **alors**
    retirer C_2 de ϕ : $\phi := \phi \setminus C_2$
  fin si
  si il existe deux cubes $x_i \wedge C_1$ et $\neg x_i \wedge C_2$ de ϕ tels que $x_i \wedge C_1$ et $\neg x_i \wedge C_2$ ont un consensus et $C_1 \wedge C_2$ n'est absorbé par aucun autre cube de ϕ **alors**
    ajouter $C_1 \wedge C_2$ à ϕ : $\phi := \phi \vee C_1 \wedge C_2$
  fin si
fin tantque
Retourner ϕ.

L'application de l'algorithme du consensus transforme ϕ_{FND} par les étapes suivantes :

1. Élimination des cubes $(x_2 \wedge x_3)$ et $(x_2 \wedge x_1)$ par absorption par x_2. La formule devient :

$$\phi_{\text{FND}} = x_2 \vee (\neg x_1 \wedge \neg x_2 \wedge \neg x_3).$$

2. Ajout de la conjonction $\neg x_1 \wedge \neg x_3$ à ϕ_{DNF} dérivée par consensus sur x_2, la formule obtenue est :

$$\phi_{\text{FND}} = x_2 \vee (\neg x_1 \wedge \neg x_2 \wedge \neg x_3) \vee (\neg x_1 \wedge \neg x_3).$$

3. Élimination du cube $(\neg x_1 \wedge \neg x_2 \wedge \neg x_3)$ par absorption par $(\neg x_1 \wedge \neg x_3)$, puis la formule n'est plus réductible, la formule finale obtenue est la FND minimale :

$$\phi_{\text{FND}_{min}} = x_2 \vee (\neg x_1 \wedge \neg x_3).$$

Les impliquants premiers de la fontion ϕ sont donc x_2 et $(\neg x_1 \wedge \neg x_3)$.

2.2.2 DDBOR-IP

La FND minimale peut être obtenue à partir d'une formule sous forme DDBOR par la procédure $recurse$IP. Cette procédure correspond à la procédure du consensus dont les opérations de consensus et d'absorption sont réalisées variable par variable et récursivement sur le DDBOR représentant la formule. La procédure consiste en la décomposition (notée /) du DDBOR donné en entrée selon sa racine en deux DDBOR low et $high$ sur lesquels vont être réalisées les opérations de consensus et d'absorption. L'opération de consensus correspond à la construction d'un DDBOR représentant la conjonction des DDBOR low et $high$ (notée $\wedge$). La conjonction de deux DDBOR est réalisée par la procédure APPLY décrite dans [7]. L'opération d'absorption est réalisée par un opérateur de différence ensembliste (noté \) adapté à partir de [25]. On note $var(f.racine)$ la variable associée à la racine du DDBOR f. La fonction TS($\{var, low, high, dc\}$,A) est une fonction ajoutant à une table d'association A un sommet associé au descripteur $(var, low, high, dc)$ ou renvoyant ce sommet si il existe déjà dans A. La procédure est détaillée dans l'algorithme 2.

Algorithme 2 Calcul récursif des impliquants premiers sur un DDBOR

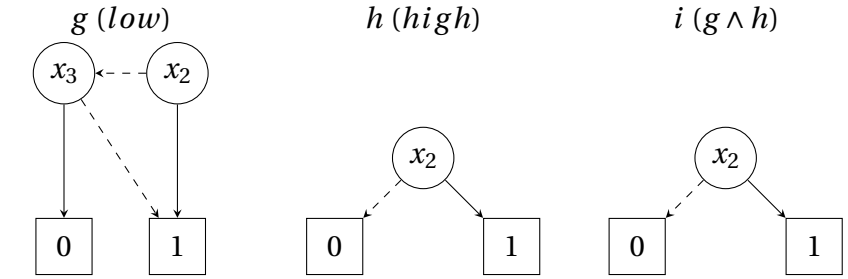
ENTRÉE: Un DDBOR f représentant une formule booléenne
SORTIE: Un graphe IP représentant la FND minimale

Fonction $recurse$IP(f)
  $a \leftarrow var(f.racine)$
  si $a = 1$ **alors**
    retourner 1 ▷ condition d'arrêt
  fin si
  si $a = 0$ **alors**
    retourner 0 ▷ condition d'arrêt
  fin si
  $g \leftarrow f/\neg a$ ▷ sous-DDBOR low de f
  $h \leftarrow f/a$ ▷ sous-DDBOR $high$ de f
  $i \leftarrow (g \wedge h)$ ▷ i est dérivé par consensus sur a
  $g \leftarrow (g \setminus i)$ ▷ Réduction de g par absorption par i
  $h \leftarrow (h \setminus i)$ ▷ Réduction de h par absorption par i
  IP $\leftarrow$ TS($\{a, recurse\text{IP}(g), recurse\text{IP}(h), recurse\text{IP}(i)\}$,IP) ▷ Construction du sommet du graphe IP
fin Fonction
retourner IP

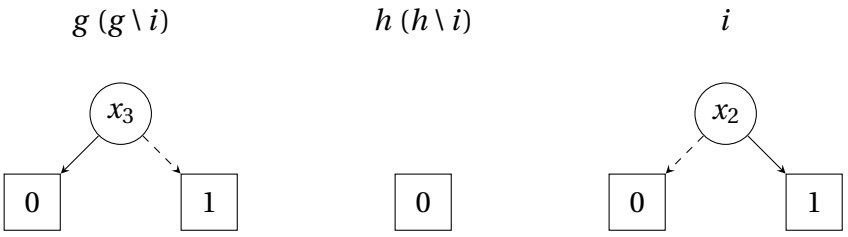
Le résultat est retourné sous forme d'un graphe dirigé acyclique (la table d'asso-

ciation A) réduit selon les mêmes règles qu'un DDBOR mais dont les sommets non-terminaux ont trois successeurs low, $high$ et dc (pour «don't care»). L'ensemble des impliquants premiers peut être obtenu à partir de ce graphe par la disjonction de tous les chemins menant au sommet terminal 1 et tel qu'un cube contient le littéral négatif (respectivement positif) de la variable associée à un sommet si le chemin passe par le successeur low (respectivement $high$) ou ne contient pas le littéral si le chemin passe par son successeur dc.

Exemple 2.5. La figure 2.3 présente la première étape récursive comprenant l'étape de calcul du consensus et l'étape de l'absorption sur la racine x_1 du DDBOR de la fonction f (présenté dans la figure 2.1) selon l'ordre $x_1 \prec x_2 \prec x_3$.



Décomposition selon x_1 et dérivation par consensus.



Réduction de g et h par absorption par i.

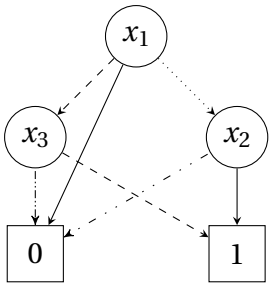


Figure 2.3 – Première étape de l'application de recurseIP sur le DDBOR représentant la fonction f selon l'ordre $x_1 \prec x_2 \prec x_3$ et graphe IP final représentant l'ensemble des impliquants premiers.

2.2.3 PLNE-IP

La programmation linéaire en nombres entiers résout des problèmes d'optimisation combinatoire où, étant donné un ensemble de variables et des contraintes sur ces variables, on doit trouver une assignation qui maximise (ou minimise) une certaine fonction de coût. On parle de problème linéaire lorsque les contraintes et la fonction de coût sont des combinaisons linéaires des variables et le problème est à nombres entiers si ces variables ne sont autorisées qu'à prendre des valeurs dans l'ensemble des entiers, et est binaire si les variables autorisées sont booléennes.

Définition 2.6. Soit y le vecteur binaire recherché, soit m un vecteur de taille h spécifiant la fonction de coût, soient v un vecteur d'entiers de taille r et W une matrice de taille (r,h) spécifiant les contraintes. Un problème de programmation linéaire en nombres entiers binaires (PLNE-{0,1}) s'exprime de la façon suivante :

$$\text{Minimiser } \sum_{j=1}^{h} m_j.y_j \text{ tel que } \sum_{j=1}^{h} \mathrm{W}_{i,j}.y_j \leq v_i \text{ avec } 1 \leq i \leq r, y \in \{0,1\}^h.$$

Soit une formule en FNC $\phi_{\mathrm{FNC}} = \mathrm{C}_1 \wedge \ldots \wedge \mathrm{C}_r$ où chaque C_i est une clause. On note $\mathrm{L}_\phi = \bigcup_{i=1}^{r} \mathrm{C}_i$ l'ensemble des littéraux apparaissant dans ϕ_{FNC}. Pour que ϕ_{FNC} soit satisfaite, il faut qu'au moins un littéral de chacune de ses clauses soit satisfait. Par conséquent, un impliquant de ϕ_{FNC} est un cube C satisfaisable (à savoir, sans contradiction) formé d'au moins un littéral de chaque clause de ϕ_{FNC}.

Définition 2.7. Soit $\mathrm{S} = (\mathrm{E}_1, ..., \mathrm{E}_n)$ une famille d'ensembles finis E_i. Une *frappe* de S est un ensemble T tel que T contient au moins un élément de chaque E_i. T est une *frappe minimale* si il n'existe pas de sous-ensembles de T qui soit une frappe de S.

Le calcul d'un impliquant consiste donc à trouver une frappe de la famille de clauses $(\mathrm{C}_1, ..., \mathrm{C}_r)$ d'une formule en FNC qui ne contienne pas de littéraux contradictoires et le calcul d'un impliquant premier correspond donc au clacul d'une frappe minimale de cette même famille et sans contradiction.

Formulation PLNE du calcul d'impliquant premier. On peut alors exprimer le problème d'inférence des impliquants premiers comme un problème de programmation

linéaire en nombres entiers de la façon suivante : on associe à chaque littéral $l_j \in \mathrm{L}_\phi$ une variable binaire notée y_{l_j}, qui représente la présence ($y_{l_j} = 1$) ou l'absence ($y_{l_j} = 0$) du littéral l_j. Le problème est alors de trouver un vecteur $y = (y_{l_1}, \ldots, y_{l_k})$ minimisant les littéraux (fonction objectif) qui soit à la fois une frappe d'ensembles (contrainte de frappe d'ensembles) de la FNC et satisfaisable (contrainte de satisfiabilité). En d'autres termes :

$$\text{Minimiser} \sum_{j=1}^{k} y_{l_j} \quad \text{avec} \quad (y_{l_1}, \ldots, y_{l_k}) \in \mathbb{B}^k.$$

La contrainte de frappe d'ensembles s'exprime par une contrainte sur chaque clause C_i de la FNC de façon à ce qu'au moins un des littéraux de la clause soit présent dans la solution y :

$$\forall \mathrm{C}_i : \sum_{l_j \in \mathrm{C}_i} y_{l_j} \geq 1.$$

Enfin, l'impliquant doit être satisfaisable : on ajoute donc la contrainte qu'il n'y ait pas de contradiction dans le cube et donc pas de littéraux de signes opposés. On dénote par $\mathrm{P} = \{v_j \in \mathrm{L}_\phi \mid v_j \in \mathrm{L}_\phi \wedge \neg v_j \in \mathrm{L}_\phi\}$ l'ensemble des variables dont les deux littéraux (positifs et négatifs) sont présents dans L_ϕ, la contrainte de satisfiabilité s'exprime alors par la contrainte linéaire suivante :

$$\forall\, v_j \in \mathrm{P} : y_{v_j} + y_{\neg v_j} \leq 1.$$

Algorithme de calcul de tous les impliquants premiers. Nous venons de présenter comment exprimer le problème de calcul d'un impliquant premier à partir d'une formule en FNC et d'une formulation en programmation linéaire en nombres entiers. Ici, nous montrons un algorithme permettant de calculer tous les impliquants premiers de la fonction par itération de la résolution PLNE. Pour calculer tous les impliquants premiers, il est nécessaire d'ajouter une contrainte de façon à ce que la solution qui vient d'être calculée ne soit pas recalculée, et de traduire les solutions y calculées en cube.

Conversion de y en un ensemble de littéraux C**.** y est une représentation binaire d'un cube C où $y_{l_j} = 0$ signifie que l_j n'appartient pas à C et $y_{l_j} = 1$ signifie que l_j est contenu dans C. Un impliquant premier correspond donc à la conjonction de l'ensemble des littéraux de C et dest éduit de y comme suit :

$$\mathrm{C} = \{l_j \in \mathrm{L}_\phi \mid y_{l_j} = 1\}.$$

Exclusion de y des solutions futures. Comme un impliquant premier est minimal par l'inclusion par définition, aucune nouvelle solution ne peut contenir les mêmes littéraux. On peut donc définir la contrainte suivante permettant d'exclure la solution y qui vient d'être calculée et représentée par un ensemble C de littéraux :

$$\sum_{l_j \in \mathrm{C}} y_{l_j} \leq |\mathrm{C}| - 1.$$

Algorithme PLNE-IP. L'algorithme permettant le calcul de tous les impliquants premiers d'une fonction booléenne f à partir d'une représentation de cette formule en FNC est donné dans l'algorithme 3.

Algorithme 3 Algorithme PLNE-IP

ENTRÉE: Une formule ϕ_{FNC} représentant une fonction booléenne f
SORTIE: L'ensemble IP de tous les impliquants premiers de f
Construction des contraintes linéaires à partir de ϕ_{FNC}
Initialisation de IP = $\emptyset$
tantque une solution y est trouvée **faire**
  $y \leftarrow$ Résoudre le problème PLNE avec un solveur
  si une solution y est trouvée **alors**
    Traduire y e unn ensemble de littéraux C
    Ajouter la contrainte d'exclusion de y
    IP = IP $\cup$ {C}
  fin si
fin tantque
Retourner IP.

Exemple 2.6. Considérons la FNC 2.2 que l'on a obtenue plus haut par transformation de ϕ :

$$\phi_{\mathrm{FNC}} = (x_2 \vee \neg x_1) \wedge (x_2 \vee \neg x_3).$$

L'ensemble des littéraux apparaissant dans ϕ_{FNC} est :

$$L_{\phi_{\text{FNC}}} = \{\neg x_1, x_2, \neg x_3\}.$$

Aucun littéral n'apparaît à la fois en positif et négatif dans ϕ_{FNC}, il n'y a donc pas de contrainte de satisfiabilité. Le problème en PLNE pour cet exemple s'exprime de la façon suivante :

$$\begin{array}{llllll}
\text{Minimiser} & y_{\neg x_1} + & y_{x_2} + & y_{\neg x_3} & & \text{Fonction objectif} \\
\text{tel que} & y_{\neg x_1} + & y_{x_2} & & \geq 1 & \text{Contrainte de frappe} \\
& & y_{x_2} + & y_{\neg x_3} & \geq 1 & \text{Contrainte de frappe} \\
& & & y & \in \mathbb{B}^3. &
\end{array}$$

Pour cet exemple, on peut facilement vérifier que la première solution (la plus petite) trouvée est $y = (0,1,0)$ conduisant à rajouter la contrainte suivante :

$$y_{x_2} \leq 0 \qquad \text{Contrainte d'exclusion de la solution } y = (0,1,0).$$

Cette contrainte exclut qu'un autre impliquant premier puisse contenir le littéral x_2. La seconde solution trouvée est $y = (1,0,1)$ conduisant à l'ajout de la contrainte :

$$y_{\neg x_1} + y_{\neg x_3} \leq 1 \qquad \text{Contrainte d'exclusion de la solution } y = (1,0,1).$$

Cette contrainte exclut qu'un autre impliquant premier puisse contenir à la fois le littéral $\neg x_1$ et le littéral $\neg x_3$. Puis, plus aucune solution n'est trouvée, l'algorithme s'arrête, convertit les deux solutions en l'ensemble des impliquants premiers de f, soit IP = $\{\{x_2\}, \{\neg x_1, \neg x_3\}$.

Dans la suite de ce chapitre, nous montrons comment ces algorithmes sont modifiés pour permettre le calcul des noyaux de la reprogrammation.

3 Méthodes d'inférence des noyaux

Dans cette section, nous allons montrer comment adapter les méthodes précédentes pour le calcul des noyaux de la reprogrammation.

3.1 Méthodes PLNE-Noyau

Algorithme 4 Algorithme PLNE-$\Diamond-$Noyau

ENTRÉE: Un RBC F_U et une propriété p
SORTIE: L'ensemble $\mathscr{K}^*$ de tous les noyaux de la reprogrammation pour $\Diamond p$
Assemblage de la condition de stabilité STBL_{F_U}
Calcul de ϕ_{FNC}, la conversion en FNC de $\text{STBL}_{F_U} \wedge p$
Construction des contraintes linéaires à partir de ϕ_{FNC}
Initialisation de $\mathscr{K}^* = \emptyset$

tantque une solution y est trouvée **faire**
  $y \leftarrow$ Résoudre le problème PLNE avec un solveur
  si une solution y est trouvée **alors**
    Traduire y en un ensemble de littéraux de contrôle négatifs C^*
    Ajouter la contrainte d'exclusion de y
    Ajouter C^* à l'ensemble des noyaux $\mathscr{K}^* = \mathscr{K}^* \cup \{C^*\}$
  fin si
fin tantque
Retourner $\mathscr{K}^*$.

La méthode PLNE-Noyau opère sur une formule ϕ en FNC, qui contient toutes les variables (d'états et de contrôle, requête $\Diamond$) ou seulement les variables de contrôle (requête $\Box$), et calcule l'ensemble des noyaux $\mathscr{K}^*$ à l'aide d'un solveur de programmation linéaire en nombres entiers. Les méthodes PLNE-$\Diamond-$Noyau et PLNE-$\Box-$Noyau sont détaillées dans l'algorithme 4 et l'algorithme 5 respectivement. Ces méthodes sont adaptées de l'algorithme 3 présenté dans la section précédente et sont fondées sur la traduction des contraintes liées à la définition des noyaux en contraintes linéaires, de telle façon que la solution calculée y soit la représentation binaire d'un impliquant C^* minimisant les paramètres de contrôle négatifs par l'inclusion (c'est-à-dire, un noyau).

La construction des contraintes linéaires diffère de PLNE-IP parce que l'on ne minimise pas tous les littéraux mais seulement les littéraux de contrôle négatifs. De fait, les contraintes de frappe et de satisfiabilité sont identiques, puisque l'on cherche également un impliquant de la fonction, mais la fonction objectif, la fonction de traduction

d'une solution y en noyau et la contraintes d'exclusion sont modifiées par rapport à l'algorithme 3.

Algorithme 5 Algorithme PLNE-$\Box$−Noyau

ENTRÉE: Un RBC $\mathrm{F_U}$ et une propriété p
SORTIE: L'ensemble $\mathscr{K}^*$ de tous noyaux de la reprogrammation pour $\Box p$
Assemblage de la condition de stabilité $\mathrm{STBL_{F_U}}$
Conversion en FNC de $\mathrm{STBL_{F_U}} \implies p$
Suppression des variables d'états des cubes pour obtenir ϕ_{FNC}
Construction des contraintes linéaires à partir de ϕ_{FNC}
Initialisation de $\mathscr{K}^* = \emptyset$

tantque une solution y est trouvée **faire**
  $y \leftarrow$ Résoudre le problème PLNE avec un solveur
  si une solution y est trouvée **alors**
    Traduire y en un ensemble de littéraux de contrôle négatifs C^*
    Ajouter la contrainte d'exclusion de y
    Ajouter C^* à l'ensemble des noyaux $\mathscr{K}^* = \mathscr{K}^* \cup \{\mathrm{C}^*\}$
  fin si
fin tantque
Retourner $\mathscr{K}^*$.

Théorème 3. La méthode PLNE-Noyau calcule tous les noyaux de reprogrammation et seulement ceux-ci.

Démonstration. La preuve du théorème 3 réside dans la traduction des contraintes liées à la définition des noyaux en contraintes linéaires.

Soit la formule d'entrée en FNC $\phi_{\mathrm{FNC}} = \mathrm{C}_1 \wedge \ldots \wedge \mathrm{C}_r$ spécifiant la requête de reprogrammation et où chaque C_i est une clause. On note $\mathrm{L}_\phi = \bigcup_{i=1}^{r} \mathrm{C}_i$ l'ensemble des littéraux apparaissant dans ϕ_{FNC}. On associe à chaque littéral $l_j \in \mathrm{L}_f$ une variable binaire notée y_{l_j}, c'est-à-dire que le vecteur y est indexé par les littéraux de L_f.

Fonction objectif. Soit $\bar{\mathrm{N}}$ l'ensemble des paramètres de contrôle apparaissant négativement dans L_f, c'est-à-dire, $\bar{\mathrm{N}} = \{u_j \in \mathrm{U} \mid \neg u_j \in \mathrm{L}_f\}$. Comme chaque noyau doit, par définition, minimiser les paramètres de conrtôle négatif, la fonction objectif correspond à la somme des variables $y_{\neg u_j}$ associées à ces paramètres de contrôle négatifs :

$$\sum_{u_j \in \bar{\mathrm{N}}} y_{\neg u_j}.$$

Conversion d'une solution y en un noyau C^*. y est la représentation binaire d'un cube C avec $y_{l_j} = 0$ signifiant que l_j n'appartient pas à C et $y_{l_j} = 1$ signifiant que l_j est contenu dans C, soit $C = \{l_i \mid y_{l_i} = 1\}$. Un noyau C^* étant l'ensemble des paramètres de contrôle contenus dans C, C^* est déduit de y comme suit :

$$C^* = \{u_j \in U \mid y_{\neg u_j} = 1\}.$$

Exclusion de tous les ensembles incluant le noyau calculé. L'exclusion des noyaux des solutions futures suit la même logique que la contrainte d'exclusion de l'algorithme 3, excepté qu'elle ne s'applique elle aussi qu'à l'ensemble des paramètres de contrôle négatifs. Cette contrainte s'exprime donc à partir de C^* comme suit :

$$\sum_{u_j \in C^*} y_{\neg u_j} \leq |C^*| - 1.$$

La méthode est ensuite itérée jusqu'à ce qu'aucune solution ne soit trouvée et retourne finalement l'ensemble de tous les noyaux.

□

Minimisation de poids. Les médicaments existants peuvent avoir des coûts plus ou moins élevés. Le développement de médicaments peut également avoir des coûts variables en fonction de l'accessibilité des molécules. Les méthodes PLNE-Noyau permettant de calculer les solutions une par une, il peut être intéressant de «prioriser» leur calcul en fonction de ces coûts. Les méthodes PLNE-Noyau peuvent être aisément enrichies de façon à «prioriser» le calcul des solutions minimisant un poids. Les valeurs des variables $y_{\neg u_j}$ correspondent alors au coût associé à l'action réalisée par le contrôle u_j. Quand aucun coût n'est spécifié, chaque variable $y_{\neg u_j}$ a la valeur 1 et l'algorithme calcule en priorité les noyaux ayant le plus petit nombre de paramètres actifs de contrôle. Quand un coût est défini, l'algorithme calcule en priorité les noyaux minimisant la somme des coûts des paramètres de contrôle actifs.

3.2 Méthodes DDB-Noyau

Les méthodes DDB-Noyau calculent les impliquants premiers à partir de la formule spécifiant les conditions de stabilité et de propriété représentées par un diagramme de décision binaire ordonné et réduit. Ces méthodes sont donc fondées sur la procédure du consensus appliquée aux DDBOR présentée dans l'algorithme 2. Pour la méthode DDB-$\Diamond$-Noyau, le DDBOR est construit directement à partir de l'assemblage de la condition de la stabilité et de propriété, sans nécessité de convertir la formule en FNC au préalable. Cependant, les impliquants premiers calculés ne sont pas les noyaux : des variables d'états sont présentes, il est donc nécessaire de filtrer les ensembles calculés en en supprimant les variables d'états puis en sélectionnant les ensembles de paramètres de contrôle qui ne sont inclus dans aucun autre. Pour la méthode DDB-$\Box$-Noyau, la formule est d'abord convertie en FNC de façon à supprimer les variables d'état de la formule, puis le DDBOR est construit et les impliquants premiers calculés.

3.3 Remarques sur les méthodes PLNE-Noyau et DDB-Noyau

Le calcul de l'ensemble des noyaux de reprogrammation est un problème NP-difficile, comme le calcul des impliquants premiers d'une fonction booléenne (par réduction au problème de frappe minimale d'ensembles).

Les méthodes PLNE-Noyau sont des méthodes itératives qui permettent de calculer un noyau de reprogrammation à la fois et d'implémenter aisément la notion de coût des actions.

Les méthodes DDB-Noyau sont des méthodes récurvises qui calculent tous les impliquants premiers à la fois. Elles permettent d'améliorer les performances de la procédure du consensus, qui est limitée par le nombre de redondances dans le calcul quand elle est appliquée directement sur les formules. En effet, l'application de la procédure du consensus sur les DDB permet de diviser le calcul en deux étapes : 1) construction du DDB représentant la formule puis 2) calcul des impliquants premiers de la formule. Des pistes sont envisageables pour l'optimisation des méthodes sur les DDBOR :

1. optimisation de construction du DDB par le choix d'une heuristique sur l'ordre des variables ;

2. changement des règles de construction du DDB de façon à supprimer les variables d'état au cours de la construction, sans passer par une conversion en FNC pour la méthode DDB-□-Noyau.

4 Comparaison de méthodes

Les méthodes algorithmiques permettant l'identification de stratégies de contrôle de réseaux booléens dans le but d'identifier des cibles thérapeutiques en se fondant sur l'analyse de ces réseaux a fait l'objet de publications ces dernières années. Ces méthodes correspondent à des approches par simulation [57, 60], par recherche de motifs stables [109] ou par inférence du contrôle par calcul algébrique [73]. Nous décrivons dans cette section ces différentes approches puis nous les comparons à nos méthodes.

4.1 Approche par simulation

Dans l'article [57], les auteurs assimilent les perturbations cancéreuses et interventions thérapeutiques à des défaillances des réseaux de régulation génique et introduisent les notions de *défaillance par blocage de valeurs* («stuck-at-fault») et de *défaillance par pontage* («bridging faults») par analogie aux erreurs des circuits électroniques. La défaillance par blocage de valeurs correspond à la fixation d'une variable du réseau à une valeur particulière (0 ou 1) et les défaillances par pontage correspondent à la suppression ou à l'incorporation de nouvelles interactions entre les variables du système. Une méthode d'implémentation et d'identification des défaillances par blocage de valeurs reposant sur la simulation est proposée.

Les auteurs se concentrent sur des réseaux *acycliques*, c'est-à-dire dont le graphe d'interaction ne contient pas de boucle de rétro-action. Les réseaux acycliques ont des *variables d'entrée* et des *variables de sortie*. À un état x sur les variables d'entrée est associé à un état $z(x)$ sur les variables de sortie de façon déterministe par la fonction z. La présence d'une défaillance f transforme le réseau booléen et f est dite *détectable* si, pour un même état d'entrée x, l'état de sortie du réseau défaillant $z_f(x)$ diffère de l'état de sortie du réseau sans défaillance, c'est-à-dire, $z(x) \neq z_f(x)$. Une défaillance est donc indétectable, si, pour tout état d'entrée, on a $z(x) = z_f(x)$.

Les défaillances uniques responsables de la prolifération des cellules cancéreuses sont calculées et classées selon le protocole suivant :

1. Définir un état d'entrée x et calculer $z(x)$ assimilé à l'état de sortie *sain*.
2. Simuler les n défaillances uniques par blocage de valeurs possibles du réseau (n étant le nombre de variables du réseau).
3. Classer les défaillances selon leur effet par rapport à $z(x)$, cette classification repose sur différents aspects :
   — certaines défaillances sont détectables, d'autres non ;
   — les différents états de sortie possibles sont associés à un «degré de gravité» du phénotype, les défaillances sont donc plus ou moins «graves» selon l'état de sortie qu'elles engendrent ;
   — deux défaillances f_1 et f_2 sont dites *équivalentes* si leur activation est associée au même état de sortie, c'est-à-dire, $z_{f_1}(x) = z_{f_2}(x)$.

Les défaillances assimilées à des effets de médicaments sont identifiées suivant le protocole suivant :

1. Modélisation à partir de la connaissance biologique de l'effet de médicaments existants.
2. Simulation exhaustive des 2^n vecteurs de médicaments possibles : un vecteur de n médicaments est un vecteur binaire de taille n où la défaillance est active (1) ou inactive (0). Les médicaments efficaces sont assimilés à des défaillances uniques et des combinaisons de défaillances dont l'application permet de récupérer $z(x)$.

La méthode proposée dans cet article repose donc sur une méthode «force brute» de simulation exhaustive de toutes les défaillances uniques possibles pour des réseaux booléens acycliques.

Cette approche est améliorée par [60], qui :

1. présente avec plus de précision l'implémentation de l'ajout de paramètres modélisant les défaillances par blocage du réseau booléen ;
2. propose une approche par simulation symbolique plutôt qu'une simulation explicite en se fondant sur des méthodes SAT-ATPG, qui traduisent les conditions

de tests par simulation en une formule et permettent l'identification des défaillances directement à partir de la formule ;

3. introduit des notions d'optimalité des thérapies (minimalité des solutions, poids/coûts) ;
4. propose d'associer la «sévérité» du phénotype malade à la mesure de la distance de Hamming entre un état malade $z_f(x)$ et l'état sain $z(x)$.

Formalisation de la simulation de défaillances. À chaque gène g_j du réseau est associé à une formule en FNC G_j, qui est évaluée à 1 si et seulement si les variables représentant ses entrées et ses sorties prennent des valeurs cohérentes avec sa table de vérité. Par exemple, à un gène g_1 défini par la formule $a \wedge b$ est associé la formule $\mathrm{G}_1 = (\neg g_1 \vee a) \wedge (\neg g_1 \vee b) \wedge (g_1 \vee \neg a \vee \neg b)$. Les défaillances f sont ajoutées à cette formule en FNC de façon à ce que f soit une entrée du gène g_1. La simulation de l'activation de la défaillance consiste à ajouter une clause f à cette FNC tandis que la simulation de la non-activation de la défaillance consiste à ajouter une clause $\neg f$ à la FNC.

Formulation SAT de la simulation de défaillances. La formulation du problème d'identification des défaillances est réalisée par la construction d'une formule en FNC composée :

1. d'une formule décrivant le réseau booléen acyclique et ses possibles défaillances (la formule en FNC précédente) ;
2. d'une clause spécifiant que la sortie du réseau doit être différente de l'état sain ;
3. d'une formule spécifiant le nombre de défaillances pouvant être activées simultanément.

Le calcul des défaillances est réalisé à partir de cette formule en calculant toutes les valuations des variables satisfaisant cette FNC. Les résultats obtenus sont donc des listes des ensembles de défaillances détectables (qu'ils appellent les défaillances non-redondantes) d'une taille donnée et des états de sortie auxquelles elle sont associées.

Formulation Max-SAT pour trouver des traitements optimaux. Les traitements optimaux définis par les auteurs ont deux caractéristiques :

1. Ils doivent rectifier au mieux la sortie du réseau (minimiser la distance de Hamming par rapport à l'état sain).

2. Ils sont composés d'un nombre minimal de médicaments et/ou d'un ensemble de médicaments dont le coût est minimal.

Pour calculer ces traitements optimaux, les auteurs proposent une méthode fondée sur une résolution Max-SAT partielle à poids (WPMS : «Weighted Partial Max-SAT») d'une formule en FNC dont les clauses sont classées en *clauses dures* («hard clauses») et *clauses douces* («soft clauses»). Des poids sont associés aux clauses douces et l'objectif est de calculer une valuation des variables qui satisfait toutes les clauses dures et maximise les clauses douces satisfaites.

Afin de sélectionner les médicaments garantissant la meilleure sortie du réseau, des poids sont assignés aux clauses représentant des états de sortie dans la FNC de façon à ce que l'état sain soit associé au plus grand poids et que les états défaillants aient un poids d'autant plus bas que leur distance de Hamming de l'état sain est grande. Afin de «prioriser» les traitements qui minimisent le nombre et le coût des médicaments, des poids positifs proportionnels au coût d'un médicament sont assignés aux défaillances non sélectionnées correspondant à l'action du médicament.

4.2 Approche par recherche de motifs stables

L'approche proposée par [109] est une méthode hybride reposant à la fois sur la structure et la dynamique de réseaux booléens (aussi bien acycliques que cycliques) et permettant d'identifier des stratégies de contrôle qui font converger la dynamique vers un attracteur d'intérêt. Leur algorithme de *contrôle de motifs stables* se déroule en quatre étapes :

1. Identification de l'ensemble des *motifs stables* du réseau booléen.
2. Construction d'un *diagramme de succession des motifs stables* qui associe, à chaque motif stable du réseau booléen, une séquence de motifs stables aboutissant à un attracteur.
3. Minimisation du diagramme.
4. Détermination du contrôle à partir du diagramme minimal.

Identification de l'ensemble des motifs stables. Un motif stable est un composant dynamico-structurel du réseau défini comme un ensemble de sommets du graphe

d'interaction et de leurs états associés et tel que :

— le sous-graphe d'interaction composé de ces sommets forme une composante fortement connexe minimale du graphe d'interaction ;
— les états de ces variables forment un état stable partiel.

Construction du diagramme de succession. Un diagramme de succession correspond à un ensemble de séquences de motifs stables qui déterminent un attracteur. Le diagramme de succession des motifs stables est un arbre construit à partir du réseau initial par application itérative, pour chaque motif stable et jusqu'à ce que le réseau obtenu n'ait plus de motifs stables, de :

— la réduction du réseau booléen par fixation des valeurs des variables d'un motif stable ;
— l'identification de motifs stables dans le réseau booléen réduit.

Minimisation du diagramme de succession. La minimisation d'un diagramme de succession a pour but de simplifier les séquences de motifs stables déterminant un attracteur de façon à minimiser les motifs stables à contrôler pour atteindre un attracteur d'intérêt. Cette minimisation des séquences est réalisée par identification de motifs dont tous les motifs suivants dans la séquence sont suivis d'un même attracteur et qui peuvent donc être supprimés de la séquence.

Détermination du contrôle Le diagramme de succession minimal détermine une séquence de motifs stables à contrôler pour «pousser» le système vers un attracteur. La détermination d'un contrôle minimal consiste alors à déterminer un ensemble minimal de sommets du motif stable qui «vérouille» l'ensemble du motif, et ce, pour chaque motif de la séquence d'un diagramme menant à l'attracteur d'intérêt.

4.3 Approche par calcul algébrique

Les auteurs de [73] proposent une modélisation proche de celle présentée dans cette thèse, avec une implémentation des réseaux booléens contrôlés permettant le contrôle des arcs et des sommets du graphe d'interaction par gel à l'aide de paramètres

de contrôle introduits dans les formules booléennes. Trois types de reprogrammation sont étudiés :

— la génération de nouveaux états stables ;
— la destruction d'états stables pré-existants ;
— le bloquage de régions de l'espace d'états.

Les méthodes reposent sur la réécriture du réseau booléen et du type de reprogrammation sous forme d'un système d'équations polynomiales. La résolution repose sur le calcul des *bases de Gröbner* du système d'équations.

4.4 Synthèse

Les méthodes algorithmiques développées pour calculer des stratégies de contrôle des réseaux booléens diffèrent notamment par :

la topologie des réseaux qu'elles permettent de traiter : les méthodes force brute et Max-SAT sont plus contraignantes car elles ne permettent de traiter que des réseaux acycliques ;

les perturbations modélisées : toutes les méthodes modélisent les perturbations par gel des variables mais les méthodes force brute, Max-SAT et par motifs stables ne prennent pas en comptent les perturbations par gel des interactions ;

les objectifs de la reprogrammation correspondent à la génération de nouveaux états et à la destruction d'états pré-existants. Dans le cas des méthodes traitant exclusivement les réseaux acycliques, ces états finaux sont déterminés de façon déterministe à partir d'un état d'entrée tandis que pour les méthodes traitant des réseaux cycliques avec un mode de mise à jour asynchrone, les états finaux doivent être des attracteurs stables ;

la minimalité des solutions concerne la minimalité du contrôle à effectuer pour garantir les objectifs de la reprogrammation. Certaines méthodes permettent également la minimisation du coût du contrôle ;

l'exhaustivité des solutions calculées.

Le tableau 2.2 récapitule les méthodes présentées et leurs caractéristiques. Le formalisme et les méthodes de résolution les plus proches de ceux que nous décrivons

dans cette thèse sont les travaux portant sur les bases de Gröbner, qui utilisent également le formalisme des réseaux booléens contrôlés. Cependant, nos méthodes reposent sur le calcul d'impliquants minimaux. Par ailleurs, notre méthode par résolution en programmation linéaire en nombres entiers permet d'implémenter aisément des méthodes de minimisation de coût des actions. De plus, nous spécifions formellement un langage de requêtes fondé sur des modes de repogrammation. Ces éléments permettent d'étendre les définitions des modes à des logiques plus riches tout en fournissant un cadre formel d'abduction pour la résolution.

	Méthode	Force brute	Max-Sat	Motifs stables	Bases de Gröbner	PLNE-Noyau	DDB-Noyau
Réseaux	Acycliques	+	+	+	+	+	+
	Avec cycles	−	−	+	+	+	+
Perturbations	Gel des sommets	+	+	+	+	+	+
	Gel d'arcs	−	−	−	+	+	+
Reprogrammation	Générer de nouveaux états	−	−	+	+	+	+
	Détruire des états pré-existants	−	−	+	+	+	+
	Application aux états stables	−	−	+	+	+	+
	Bloquer des régions de l'espace d'états	−	−	−	+	+	+
	Définition de modes	−	−	−	−	+	+
Solutions	Minimalité	−	−	+	-?	+	+
	Exhaustivité	−	−	−	+	+	+
	Minimisation du coût (poids)	−	+	−	−	+	−
Année de publication		2011 [57]	2012 [60]	2015 [109]	2016 [73]	2016/2017 [13, 11]	2018 [12]

Tableau 2.2 – Comparaison des méthodes d'identification de stratégies de contrôle des réseaux booléens. Le symbole + signifie que la méthode possède la caractéristique, − qu'elle ne possède pas cette caractéristique et −? signifie que la caractéristique n'est pas explicitement décrite dans l'article de référence.

Chapitre 3

Application au Cancer

Dans les chapitres précédents nous avons défini un cadre théorique de reprogrammation comportementale des systèmes dynamiques booléens (chapitre 1) et démontré que l'inférence des perturbations causales responsables de cette reprogrammation peut se résoudre par abduction (chapitre 2). Nous présentons dans ce chapitre les principes d'application du cadre théorique et d'interprétation des solutions pour l'inférence des perturbations causales du cancer et des cibles thérapeutiques (section 1). Ces principes seront illustrés ensuite sur deux modèles biologiques. Le premier porte sur le cancer du sein et permet une preuve de concept de l'approche (section 2) en récupérant des perturbations causales du cancer compatibles avec la connaissance biologique et récupère le phénomène de *létalité synthétique*, qui est un cas particulier d'épistasie négative permettant de prédire des cibles thérapeutiques. La seconde application (section 3) a été réalisée sur un modèle du cancer de la vessie issu de la littérature [84] et montre que les résultats obtenus pour l'inférence de perturbations causales du cancer sont reproductibles sur un modèle dont la conception est indépendante de l'application de la méthode.

1 Principes d'application

Les principes d'application du cadre théorique établissent un parallèle entre les concepts formels et leur interprétation en regard de la connaissance du fonctionnement des systèmes biologiques. Ils définissent donc une procédure d'application et

de validation des méthodes développées 1) pour l'inférence de perturbations responsables de la transformation d'un modèle booléen sain en réseau booléen malade et 2) pour l'inférence de cibles thérapeutiques.

1.1 Dynamique booléenne des réseaux biologiques

Le formalisme des réseaux booléens a été introduit pour la modélisation de la dynamique des *réseaux de régulation génique* par Stuart Kauffman [**?**] et René Thomas [97]. Les réseaux booléens sont aujourd'hui utilisés également pour la modélisation des *voies de signalisation*. Cette modélisation concerne divers systèmes biologiques (plantes [10, 28], cellules mammifères [84, 36, 18, 111, 23], levures [59], etc.). Les attracteurs (stables et cycliques) de la dynamique booléenne permettent la représentation de *fonctionnalités cellulaires* et de *types cellulaires* à partir d'information qualitative sur l'interaction des molécules composant ces cellules.

Réseaux de signalisation et réseaux de régulation génique. Les réseaux de signalisation correspondent à des chaînes de réactions biochimiques (phosphorylation ou ubiquitination par exemple) ayant lieu dans le cytoplasme et résultant de l'activation ou de l'inhibition de récepteurs présents au niveau de la membrane de la cellule. Les réseaux de régulation génique concernent la régulation transcriptionnelle de l'expression des gènes par des facteurs de transcription ou des ARNs, suite à une stimulation par les réseaux de signalisation, et a lieu dans le noyau par fixation aux promoteurs des gènes. Les réseaux de signalisation et de régulation génique sont fortement interconnectés : la régulation de la transcription peut concerner des molécules de signalisation et les réseaux de signalisation peuvent activer ou inhiber des facteurs de transcription. La modélisation booléenne consiste donc à abstraire les réactions biochimiques ainsi que les fixations et les mécanismes de régulation génique en ne considérant que leur effet sur l'activité et/ou l'expression des molécules sans prendre en compte leurs paramètres cinétiques [71, 101].

Phénotypes, biomarqueurs et attracteurs. Les attracteurs (états stables et attrateurs cycliques) des réseaux booléens représentent des phénotypes cellulaires correspondant à :

— des fonctionnalités cellulaires comme l'apoptose, la prolifération, la survie, la nécrose qui sont contrôlés par ces réseaux [84, 36, 18, 59] ;
— des types cellulaire spécialisés : cellules myéloïdes et lymphoïdes [23], lymphocytes T CD4 ou CD8 [88], par exemple.

Un état stable est associé à la signature moléculaire d'un phénotype tandis qu'un attracteur cyclique représente l'oscillation entre différents phénotypes.

Les attracteurs sont associés à un phénotype à partir des états d'un sous-ensemble de variables du réseau. Ces variables peuvent être assimilées à la notion de *biomarqueurs* en médecine, qui est définie dans [14] comme *une caractéristique qui peut être objectivement mesurée et évaluée comme un indicateur d'un processus biologique*. Dans notre cadre théorique, la propriété p sur laquelle sont définies les requêtes d'inférence peut représenter une formule définissant un état sur les biomarqueurs du réseau.

Pluripotence et différenciation. Les cellules capables de se différencier en différents types cellulaires sont dites *pluripotentes* par opposition aux cellules d'un type cellulaire précis qui sont dites *différenciées*. Un système multistable dont les états stables représentent différents phénotypes correspond donc à un modèle ayant la potentialité d'atteindre ces différents phénotypes, tandis qu'un système dont les états stables représentent tous le même phénotype est un modèle différencié de la cellule. Nous parlerons de modèle pluripotent pour un modèle multistable dont les états stables représentent différentes fonctionnalités cellulaires et de modèle spécialisé pour un modèle dont les états stables représentent tous le même phénotype.

Dans le cadre théorique, nous définissons deux modalités de reprogrammation d'un système dynamique : la possibilité et la nécessité. Ces deux modalités peuvent s'interpréter sur les systèmes biologiques de la façon suivante. Considérons un système initial F multistable dont l'un des états stables possède la propriété p. Une requête de reprogrammation visant à obtenir un système G dont tous les états stables ont la propriété p (requête $\Box p$) peut s'interpréter comme une différenciation de la cellule pour le phénotype représenté par la propriété p. Considérons maintenant la requête $\Box\neg p$ qui vise à obtenir un système dont aucun des états stables n'a la propriété p, cette requête peut être interprétée comme la perte de la potentialité d'un phénotype. La requête $\Diamond p$ ne peut s'appliquer que sur un système initial F dont aucun état stable ne

possède la propriété p (sinon l'ensemble des noyaux de reprogrammation est vide) et correspond au gain d'un phénotype cellulaire ayant la propriété p. Ainsi, ces requêtes de reprogrammation peuvent s'assimiler à des modes de spécialisation cellulaire d'un modèle pluripotent, de perte ou de gain de phénotypes cellulaires.

1.2 Causes et modélisation du cancer

Les cellules cancéreuses. Le cancer est défini comme un groupe de maladies partageant des caractéristiques communes en termes de phénotypes cellulaires. Ruddon, dans son livre Cancer Biology [87] en donne la définition suivante : *«Le cancer est une croissance anormale des cellules causée par des changements multiples de l'expression génique, ayant pour conséquence la dérégulation de la balance entre prolifération et mort cellulaire, et évoluant vers une population de cellules qui peuvent envahir les tissus et métastaser dans différents sites distants du site primaire»*. Les cellules cancéreuses sont représentées par un réseau booléen dont le comportement global est reprogrammé par des perturbations des valeurs de ses variables, représentant les *changements d'expression génique*. Dans notre cadre théorique, les perturbations de l'expression génique sont réalisées par l'activation des paramètres de contrôle. La surexpression d'un gène correspond au blocage de la valeur d'un paramètre de contrôle à 1 tandis que sa sous-expression correspond au blocage de sa valeur à 0. Les conséquences de ces perturbations sont formulées par des requêtes de reprogrammation d'inférence des perturbations nécessaires à la transformation d'un modèle sain pluripotent (capable de proliférer et de mourir) en un modèle cancéreux spécialisé dans la division cellulaire.

Gènes pilotes et gènes passagers. De nombreux gènes peuvent être perturbés dans les cellules cancéreuses, en partie à cause de l'instabilité génomique qui les caractérise. Parmi ces gènes, on distingue des *gènes pilotes* qui sont définis comme des gènes dont la mutation confère un avantage sélectif à la cellule porteuse, des *gènes passagers* qui peuvent être mutés mais ne confèrent pas d'avantage sélectif à la cellule [100]. Dans notre cadre théorique, les gènes pilotes peuvent être assimilés aux facteurs causaux, c'est-à-dire, aux paramètres de contrôle d'un noyau de reprogrammation.

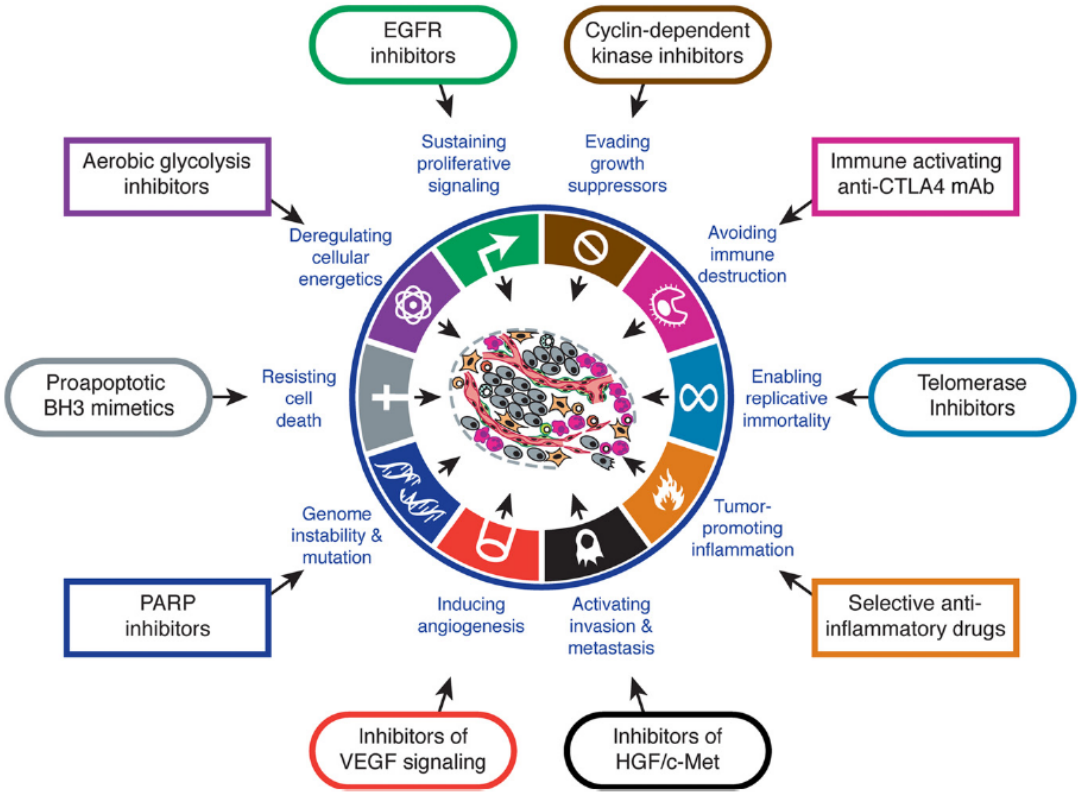


Figure 3.1 – Caractéristiques des cellules cancéreuses et exemples de thérapies les ciblant [41].

.

Oncogènes et gènes suppresseurs de tumeurs. Les gènes causaux de cancer sont divisés en 2 classes selon l'effet de leur mutation dans les cellules cancéreuses. Les cellules cancéreuses contiennent des *perturbations activatrices* (amplifications/mutations entraînant une surexpression de la protéine) des *oncogènes* et des *perturbations inhibitrices* (délétion/mutations entraînant une sous-expression de la protéine) des *gènes suppresseurs de tumeurs (GST)*. Dans notre cadre théorique, les oncogènes (respectivement les GST) peuvent être assimilés aux paramètres de gel à 1 (respectivement à 0) dans les noyaux de reprogrammation permettant la transformation d'une cellule saine en cellule cancéreuse.

1.3 Épistasie et cancer

Le nombre de perturbations nécessaires à la transformation d'une cellule saine en cellule cancéreuse est évaluée entre 2 et 8 par [100] et les auteurs de [102] suggèrent

que l'épistasie est un phénomène répandu entre les perturbations causales du cancer. L'existence de relations épistatiques entre les perturbations observées dans les cellules cancéreuses est suggérée par l'observation de schémas d'*exclusion mutuelle* et de *co-occurence* entre les mutations présentes dans les cellules cancéreuses [48, 108]. L'exclusion mutuelle entre deux perturbations correspond à l'observation que ces deux perturbations ont statistiquement tendance à ne pas apparaître dans les mêmes tumeurs. La co-occurence entre deux perturbations correspond à l'observation que ces deux perturbations ont statistiquement tendance à apparaître dans les mêmes tumeurs. Par ailleurs, la *létalité synthétique*, qui est un phénomène d'épistasie négative, est particulièrement importante pour la découverte de thérapies ciblées [47]. Il n'existe pas, à notre connaissance, de méthode causale permettant d'identifier des relations épistatiques entre perturbations génétiques.

Coopération. L'interprétation biologique de la co-occurence repose sur la nécessaire *coopération* de différentes perturbations pour l'émergence d'un comportement cancéreux. La coopération entre deux gènes α et β est un phénomène d'épistasie positive où la perturbation d'un seul de ces deux gènes n'est pas suffisante pour faire émerger un phénotype cancéreux prolifératif mais où la combinaison des deux perturbations le permet. L'hypothèse de coopération est corroboré par les travaux de [24] qui ont montré que les perturbations des deux gènes MYC et RAS *coopèrent* dans le processus de transformation cancéreuse. Dans notre cadre théorique, le phénomène de coopération peut s'interpréter à partir des noyaux de reprogrammation contenant plusieurs perturbations : la combinaison de ces perturbations est nécessaire à la reprogrammation comportementale du modèle.

Redondance. En ce qui concerne l'exclusion mutuelle, elle pourrait s'expliquer par les phénomènes suivants :

— la *redondance* : deux perturbations sont redondantes si leur perturbation simultanée dans une cellule cancéreuse n'augmente pas son potentiel de prolifération. Cette hypothèse est corroborée par l'observation que les paires de perturbations mutuellement exclusives ont tendance à concerner des molécules des mêmes voies de signalisation [21].

β \ α	A	*a*
B	**Sain**	**Viable**
b	**Viable**	**Non viable**

Figure 3.2 – Illustration schématique du phénomène de létalité synthétique entre deux gènes α et β.

— la *létalité synthétique*, dont nous parlons au paragraphe suivant.

Dans notre cadre théorique, le phénomène de redondance peut s'interpréter par comparaison des noyaux de reprogrammation (qui sont minimaux) : les perturbations d'un noyau sont suffisantes à la transformation à la reprogrammation comportementale du modèle et l'ajout de perturbations d'un autre noyau serait redondants avec l'effet des perturbations du noyau.

Létalité synthétique et thérapies ciblées. Dans [41], les auteurs décrivent les principales caractéristiques des cellules cancéreuses par opposition aux cellules normales. Les thérapies dites ciblées ont pour but de cibler de façon précise ces caractéristiques dans le but d'éviter les effets secondaires importants liés aux chimiothérapies traditionnelles. La figure 3.1, empruntée à [41], montre ces caractéristiques et des exemples de thérapies les ciblant. La découverte de partenariats de létalité synthétique est particulièrement important dans le cadre des thérapies ciblées anti-cancéreuses puisqu'elle permettrait de provoquer la mort cellulaire des cellules cancéreuses mutées sans affecter les cellules saines [47].

La *létalité synthétique* est un phénomène que l'on peut classer dans le cadre de l'épistasie négative : la présence d'un gène *a* muté dans la cellule ou d'un gène *b* muté n'a pas d'incidence sur la capacité de cette cellule à survivre, voire peut améliorer sa capacité à croître. En revanche, la perturbation simultanée de *a* et de *b* dans la cellule la conduit à la mort cellulaire, les deux gènes α et β sont alors dits *synthétiques létaux* (voir figure 3.2).

2 Cancer du sein

Le cancer du sein est un problème majeur de santé publique représentant 10% des cancers et un quart des nouveaux cas de cancer diagnostiqués chez les femmes [46]. C'est également un modèle en termes de thérapies ciblées et de connaissance des gènes perturbés dans les cellules cancéreuses : le cancer du sein est divisé en divers sous-types moléculaires selon les gènes perturbés dans la tumeur. L'identification de ces sous-types permet le choix d'une thérapie en se fondant sur ces perturbations [33]. Par ailleurs, le cancer du sein lié à une déficience du gène BRCA1 est, à notre connaissance, le seul exemple de cancer pour lequel existe actuellement une thérapie ne reposant pas sur le fait de contrer les actions de la maladie mais sur le concept de létalité synthétique. Nous avons donc conçu un modèle centré sur le cancer du sein afin d'établir une preuve de concept de notre cadre théorique de modélisation.

Nous décrivons d'abord les étapes de la conception d'un modèle booléen représentant une cellule saine (sous-section 2.1), puis nous spécifions des requêtes de reprogrammation permetttant la transformation de ce modèle en modèle de cellule cancéreuse (sous-section 2.2). Enfin, nous analysons les résultats de l'inférence au regard de la connaissance sur le cancer du sein issue de la littérature et prédisons des cibles thérapeutiques (sous-section 2.3).

2.1 Conception du modèle

La conception d'un modèle booléen à partir de la connaissance biologique requiert plusieurs étapes :

1. Identification des fonctions cellulaires perturbées par la maladie et des *graines* du modèle, c'est-à-dire de molécules considérées comme importantes pour le cancer d'intérêt et les fonctions cellulaires que l'on veut modéliser.
2. Construction d'un graphe d'interaction reliant les graines par différents intermédiaires en s'aidant de bases de données d'interactions biologiques et de la littérature.
3. Définition de formules propositionnelles pour chaque sommet du graphe de façon à ce que chaque sommet dépende de ses interacteurs (les sommets qui

l'influencent au niveau dynamique), choix des biomarqueurs et interprétation des phénotypes cellulaires à partir des attracteurs de la dynamique du modèle.

Dans la suite, nous détailllons ces étapes pour la conception d'un modèle booléen pour le cancer du sein.

2.1.1 Fonctions cellulaires et graines du cancer du sein

Les caractéristiques principales et les mieux décrites des cellules cancéreuses sont leurs capacités à proliférer continuellement et à éviter d'entrer en mort cellulaire par apoptose. Le gène BRCA1 et les gènes de la famille des récepteurs aux facteurs de croissance (EGFR pour «epidermal growth factor receptors», dont fait partie le gène HER2 qui est parfois noté ERBB2) sont des acteurs majeurs du cancer du sein [8, 45, 75]. BRCA1 ainsi que son partenaire synthétique léthal PARP1 sont impliqués dans la réponse aux dommages de l'ADN et la régulation de l'entrée en mitose ou du déclenchement de la mort cellulaire si les dommages de l'ADN sont irréparables. La famille EGFR est impliquée dans la stimulation de la division cellulaire par des facteurs de croissance. Nous avons donc développé un modèle se focalisant sur la prise de décision apoptose/division cellulaire à partir des graines BRCA1, PARP1 et EGFR.

2.1.2 Construction du graphe d'interaction

Le graphe d'interaction présenté dans la figure 3.3 comporte 13 sommets et 21 interactions. Il a été construit à partir des bases de données KEGG [49] et SIGNOR [81] et de recherche bibliographique. Le tableau 3.1 donne le mécanisme de régulation (physique, biochimique ou biologique) décrit dans les références pour chaque interaction. Le modèle se concentre sur le contrôle de la division cellulaire et de l'apoptose par les voies de signalisation et de régulation génique intracellulaires [1]. Ce contrôle s'effectue, d'une part, par la voie de signalisation EGFR et, d'autre part, par la réponse aux dommages à l'ADN par les gènes BRCA1/PARP1. EGFR est un membre de la famille des récepteurs aux facteurs de croissance, qui activent l'entrée en division cellulaire par les voies suivantes :

1. Dans la suite de ce manuscrit et pour des raisons de simplicité, les noms de gènes (par convention écrits en lettres majuscules) peuvent aussi dénommer les protéines qu'ils codent.

Interaction	Mécanisme moléculaire	Référence
EGFR $\longrightarrow$ PI3K	Fixation	[33, 5, 81, 27]
PI3K $\longrightarrow$ AKT	Inconnu/Phosphorylation	[33, 5, 81]
PTEN $\dashv$ PI3K	Catalyzation de la réaction inverse	[33, 5, 81, 27]
AKT $\dashv$ GSK3β	Phosphorylation	[33, 5, 81]
AKT $\longrightarrow$ MDM2	Phosphorylation	[33, 5, 81]
AKT $\longrightarrow$ BCL-2	Activation par phosphorylation de CREB1, qui active la transcription de BCL-2	[5, 81]
GSK3β $\dashv$ CYCD1	Phosphorylation	[5, 81]
TP53 $\longrightarrow$ MDM2	Transcriptionnel	[81, 4]
MDM2 $\dashv$ TP53	Ubiquitination (signal de dégradation par le protéasome)	[33, 81, 4]
TP53 $\longrightarrow$ PTEN	Transcriptionnel	[27, 4, 65, 94]
TP53 $\longrightarrow$ BAX	Fixation/Transcriptionnel	[81, 4]
BRCA1 $\longrightarrow$ TP53	Transcriptionnel	[72, 110]
ERK1/2 $\longrightarrow$ PARP1	Phosphorylation	[81, 27, 22, 95]
BCL2 $\dashv$ BAX	Fixation/Localisation subcellulaire	[81, 99]
PARP1 $\dashv$ TP53	Transcriptionnel	[63, 77]
EGFR $\longrightarrow$ ERK1/2	Chaîne d'activation enzymatique	[33, 3, 112]
PARP1 $\longrightarrow$ CYCD1	Transcriptionnel	[92]
ERK1/2 $\longrightarrow$ CYCD1	Transcriptionnel/Post-transcriptionnel	[54, 68]
BRCA1 $\dashv$ CYCD1	Transcriptionnel/Régulation de la transition G1/S	[104, 29]
CYCD1 $\dashv$ BRCA1	Phosphorylation par le complexe CyclinD/CDK4	[81, 52]
BRCA1 $\dashv$ EGFR	Transcriptionnel, post-transcriptionnel et post-traductionnel	[17, 55]

Tableau 3.1 – Mécanismes moléculaires et références du graphe d'interaction du modèle du cancer du sein.

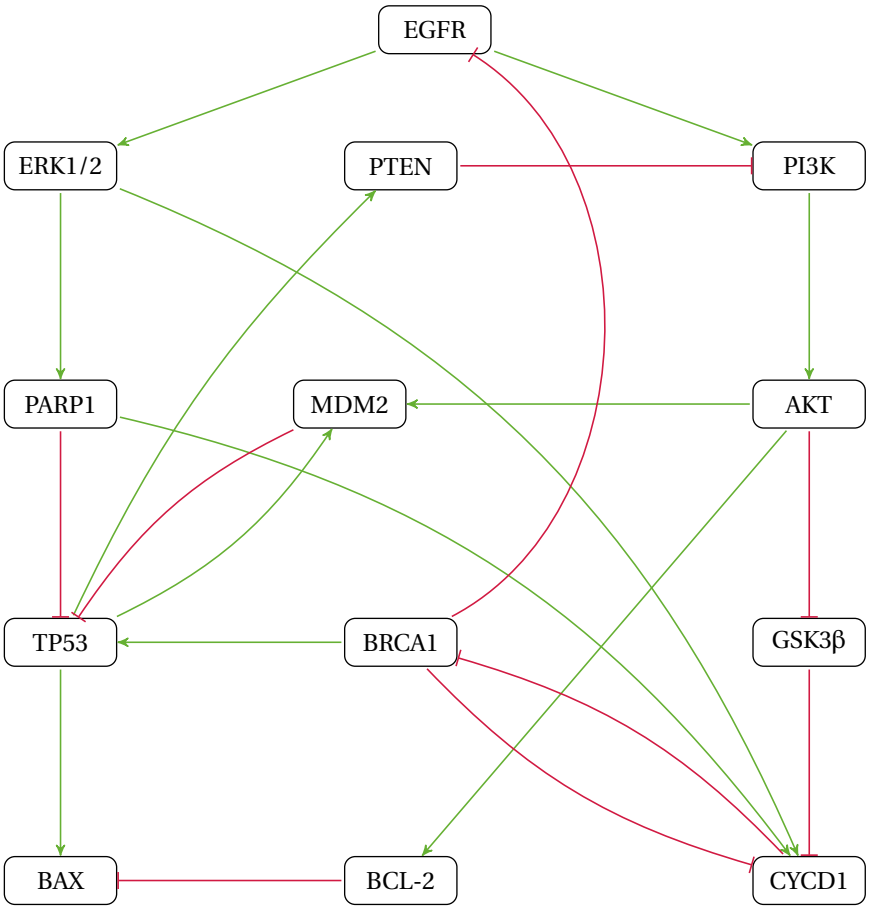


Figure 3.3 – Graphe d'interaction du modèle du cancer du sein. Les interactions positives, représentant l'activation, sont en vert et les interactions négatives, représentant l'inhibition, sont en rouge.

1. la voie de signalisation des protéines MAPK (pour Mitogen-activated protein kinases), représentée dans le graphe d'interaction par son composant ERK1/2 ;
2. la voie de signalisation PI3K (pour phosphatidyl-inositol-3 kinase).

Ce deux voies de signalisation stimulent l'entrée en division cellulaire :

1. ERK1/2 active la protéine CYCLIND1, qui est un effecteur et le régulateur majeur de la transition G1/S du cycle cellulaire (elle représente donc l'entrée en division) ;
2. le médiateur de la voie de signalisation de PI3K est AKT qui libère CYCLIND1 de son inhibition par GSK3β.

Le gène BRCA1 est impliqué dans le déclenchement de l'apoptose à travers l'activation de TP53 qui active BAX, un effecteur de l'apoptose et l'arrêt de la division du point de contrôle de la transition G1/S par une boucle négative avec CYCLIND1. PARP1 est

un gène impliqué dans la réparation de l'ADN, il est activé par la voie MAPK, inhibe TP53 et active CYCLIND1.

Les voies de l'apoptose et de la division cellulaire s'entrecroisent à différents niveaux :

1. ERK1/2 inhibe l'apoptose à travers l'activation de PARP1 ;
2. AKT active BCL-2 libérant ainsi BAX, l'effecteur de l'apoptose ;
3. TP53 inhibe la voie PI3K à travers l'activation de PTEN, qui supprime la signalisation PI3K en catalysant la réaction inverse (déphosphorylation) ;
4. TP53 est inhibé par la voie PI3K, qui active son inhibiteur MDM2 ;
5. BRCA1 bloque la signalisation d'entrée dans le cycle cellulaire en inhibant EGFR.

2.1.3 Définition des formules et interprétation des phénotypes

Variable	Formule	Référence
PI3K	$\neg$PTEN $\wedge$ EGFR	[33, 27]
MDM2	AKT $\wedge$ TP53	[65, 70]
TP53	$\neg$MDM2 $\wedge$ (BRCA1 $\vee$ $\neg$PARP1)	[65, 72, 110, 63, 77, 70]
BAX	$\neg$BCL-2 $\wedge$ TP53	[99, 98]
CYCD1	($\neg$GSK3β $\wedge$ ERK1/2) $\vee$ ($\neg$BRCA1 $\wedge$ PARP1)	[72, 77, 29, 6]

Tableau 3.2 – Formules propositionnelles des variables du graphe d'interaction de la figure 3.3 dépendants de plusieurs variables.

La définition des formules concerne les sommets du graphe d'interaction étant influencés par plusieurs autres sommets. C'est le cas pour les molécules suivantes : PI3K, MDM2, TP53, BRCA1, BAX et CYCD1. Les formules pour chacune de ces variables sont données dans le tableau 3.2.

Les états stables de la dynamique aynchrone du modèle ainsi défini sont donnés dans le tableau 3.3. L'effecteur de la transition G1/S d'entrée dans le cycle cellulaire étant la cycline D1, on définit la variable CYCD1 comme la marqueur de la prolifération. BAX étant un effecteur de l'apoptose, on le définit comme le marqueur de l'apop-

#	EGFR	ERK1/2	PI3K	AKT	GSK3β	MDM2	TP53	PTEN	PARP1	BRCA1	BCL-2	CYCD1	BAX
1	1	1	1	1	0	0	0	0	1	0	1	1	0
2	0	0	0	0	1	0	1	1	0	1	0	0	1

Tableau 3.3 – États stables du réseau booléen modèle du cancer du sein.

tose. Ainsi, on peut définir 4 phénotypes selon les signatures définies par ces deux marqueurs :

— le phénotype *apoptotique* correspond à la signature suivante :

$$\{\text{BAX} \mapsto 1, \text{CYCD1} \mapsto 0\}$$

— le phénotype *prolifératif* correspond à la signature suivante :

$$\{\text{BAX} \mapsto 0, \text{CYCD1} \mapsto 1\}$$

— le phénotype *quiescent* est assimilé à une cellule qui n'entre ni en division cellulaire ni en apoptose [93] et correspond à la signature suivante :

$$\{\text{BAX} \mapsto 0, \text{CYCD1} \mapsto 0\}$$

— le phénotype *dormant* est assimilé à une cellule dont la signalisation apoptotique est compensé par la signalisation d'entrée en division [93] et correspond à la signature suivante :

$$\{\text{BAX} \mapsto 1, \text{CYCD1} \mapsto 1\}$$

L'état stable #1 correspond donc à un phénotype prolifératif et l'état stable #2 à un phénotype d'apoptose.

2.2 Requêtes et solutions

Les cellules cancéreuses étant caractérisées par la dérégulation de la balance apoptose/prolifération avec perte de la capacité d'entrée en apoptose et maintien de la signalisation de prolifération, l'inférence des perturbations causales de la transforma-

#	Noyaux de reprogrammation de $\Box p \wedge \Diamond p$
1	PI3KCA = 1
2	AKT = 1
3	BRCA1 = 0
4	EGFR = 1, MDM2 = 1
5	EGFR = 1, TP53 = 0
6	EGFR = 1, PTEN = 0
7	EGFR = 1, GSK3β = 0
8	ERK1/2 = 1, GSK3β = 0

Tableau 3.4 – Solutions de l'inférence des perturbations causales de la transformation de modèle sauvage à modèle cancéreux avec paramètres de contrôle en gel-D.

tion du modèle sauvage en modèle cancéreux s'exprime par une requête spécifiant que tous les états stables et au moins un état stable doivent avoir une signature de phénotype prolifératif. Cette requête de reprogrammation s'exprime donc de la façon suivante : $\Box p \wedge \Diamond p$ avec p correspondant à la signature du phénotype prolifératif, soit :

$$p = \text{CYCD1} \wedge \neg \text{BAX}.$$

.

Le réseau booléen contrôlé est construit de façon à ce que toutes les variables sauf les marqueurs soient contrôlées par des paramètres de contrôle en gel-D (D^0 et D^1, voir tableau 3.4 pour les résultats de l'inférence en Gel-D). Les noyaux de reprogrammation en Gel-U pour l'inférence des solutions de la requête $\Box p$ n'étant pas analysés, ici nous les donnons à titre informatif dans le tableau 3.9 de l'annexe de ce chapitre.

2.3 Analyse des solutions

L'analyse des solutions de l'inférence a pour but d'évaluer la pertinence des noyaux de reprogrammation au regard de la connaissance biologique. Cette pertinence sera évaluée en trois étapes :

1. Comparaison des valeurs du gel affectant individuellement les variables dans les solutions à leur rôle attribué dans la littérature biologique (oncogène ou

gène supresseur de tumeur).

2. Comparaison des solutions dans leur globalité avec la recherche de schémas de co-occurence et d'exclusion mutuelle issus de l'analyse des paysages mutationnels.
3. Prédiction de partenariats de létalité synthétique.

2.3.1 Oncogènes *versus* GST

Variable	Symbole	Oncogène (%)	GST (%)	Rôle	Table
Gel à 0					
TP53	TP53	73	20	GST	S2A
BRCA1	BRCA1	0	69	GST	S2A
PTEN	PTEN	30	55	GST	S2A
GSK3β	-	-	-	-	-
Gel à 1					
PI3K	PI3KCA	95	1	oncogène	S2A
AKT	AKT1	93	1	oncogène	S2A
EGFR	EGFR	97	0	oncogène	S2A
ERK1/2	MAP2K1	67	0	oncogène	S2A
MDM2	-	-	-	oncogène	S2B

Tableau 3.5 – Comparaison de la valeur du gel-D de la molécule dans les solutions de l'inférence aux score oncogènes et GST et au rôle attribué au gène dans [100].

Dans l'article [100], les auteurs proposent une classification oncogène/GST de gènes fréquemment altérés dans les cellules cancéreuses. Cette classification repose sur des mesures réalisées à partir de données de séquençage issues de la base de données COSMIC [2, 39]. Le tableau S2A de leur matériel supplémentaire présente une liste de gènes mutés auxquels sont attribués un score oncogène et un score GST (ces scores sont en pourcentage) à partir desquels sont déduits leur classification en tant que GST ou oncogène de la façon suivante : un gène est classé GST s'il a un score GST supérieur à 20% et oncogène s'il a un score oncogène supérieur à 20%. Quand un gène a un score oncogène supérieur à 20% et un score GST supérieur à 5%, il est classé comme GST. Le tableau S2B présente une liste de gènes affectés par des amplifications ou des délétions

homozygotes. Ces gènes sont classés comme oncogènes s'ils respectent les deux conditions suivantes : ils sont inclus dans le Cancer Gene Census [1] et respectent les critères énoncés décrits dans [91]. Ils sont classifiés comme GST s'ils respectent les deux conditions suivantes : ils ont au moins 10 délétions homozygotes référencées dans la base de données COSMIC et ne sont co-délétés avec aucun autre gène ayant également 10 instances de délétion homozygote.

Nous avons donc comparé les valeurs de gel des variables dans les solutions inférées à leur classification en tant qu'oncogène ou GST. Dans le tableau 3.5, nous donnons, pour chaque variable affectée par un gel dans au moins une des solutions de l'inférence, le symbole par lequel il est identifié dans les tableaux de [100], ses scores oncogène et GST, sa classification et le tableau de référence dans lequel peut être retrouvée cette classification. GSK3β n'étant par répertorié, nous avons effectué une recherche bibliographique sur son rôle dans le cancer du sein. GSK3β semble avoir un rôle complexe dans le développement des cancers : il est identifié comme GST pour les cancers du sein [34] et de la peau et comme oncogène pour les cancers du foie et des ovaires [62]. Ce rôle dual de GSK3β dans la tumorigenèse est passé en revue dans [62]. On remarque donc que *tous les gènes gelés à* 0 *dans au moins une solution sont des GST et tous les gènes gelés à* 1 *dans au moins une solution sont des oncogènes. L'inférence des gènes causaux sur le modèle du cancer du sein ségrège donc les oncogènes et les GST.*

2.3.2 Combinaisons de perturbations

Coopération. Dans [74], les auteurs étudient les réponses des cellules cancéreuses du sein à un traitement par le trastuzumab, un anticorps inhibant la croissance des cellules cancéreuses mammaires surexprimant ERBB2 (membre de la famille des récepteurs de croissance EGFR), et montrent que des délétions de PTEN sont associées à la résistance à ce médicament. Il a été montré [19] que les mutations de TP53 sont fréquentes dans les cellules cancéreuses surexprimant ERBB2. Ces observations statistiques sont corroborées par les expériences de [106], qui montrent que la mutation de TP53 amplifie la signalisation EGFR et favorise la prolifération cellulaire.

Ces résultats suggèrent que les combinaisons de perturbations EGFR, TP53 et EGFR, PTEN coopèrent dans les cellules cancéreuses mammaires. Ces deux combinaisons

sont représentées dans les noyaux de reprogrammation 5 et 6.

Redondance. Dans [80], les auteurs rapportent l'exclusion mutuelle entre les mutations de PI3KCA et de AKT1 ainsi qu'entre l'amplification de ERBB2 et AKT1 dans les données issues de tumeurs mammaires. Les délétions de PTEN et les mutations activatrices de PI3KCA ne sont pas mutuellement exclusives, contrairement à ce qui est observé par [19], qui rapportent que les mutations de PI3KCA sont associées à un sous-type du cancer du sein différent du sous-type associé à des délétions de PTEN, ce qui les rend mutuellement exclusives. On peut remarquer que les perturbations de PI3KCA, AKT et EGFR, d'une part, et de PI3KCA et PTEN, d'autre part, sont présentes dans des noyaux différents de reprogrammation. Ces résultats pourraient suggérer que les différents noyaux soient associés à différents sous-types de cancer.

#	Noyaux de reprogrammation de $\Diamond q$
1	BRCA1 = 1
2	PARP1 = 0
3	ERK1/2 = 0
4	EGFR = 0

Tableau 3.6 – Inférence de cibles thérapeutiques pour la cancer BRCA1.

Létalité synthétique. Afin d'inférer des cibles thérapeutiques potentielles, nous simulons le comportement d'une cellule cancéreuse en appliquant la mutation BRCA1 = 0 au modèle sauvage et exprimons la requête de reprogrammation pour le gain de l'apoptose, soit $\Diamond q$ avec $q = \neg$CYCD1 $\wedge$ BAX. De même que pour l'inférence des gènes causaux du cancer, toutes les variables sont contrôlées, les solutions de l'inférence avec contrôle en Gel-D sont données dans le tableau 3.6 et celles avec contrôle en Gel-U sont données en annexe dans le tableau 3.10.

L'inférence de cibles thérapeutiques dans le cas de cellules cancéreuses déficientes en BRCA1 proposent 4 solutions de traitement : rétablir l'activité de BRCA1, ou inhiber l'une des trois molécules suivantes : BRCA1, PARP1 ou ERK1/2. L'inférence du gain de l'apoptose sur le modèle permet de récupérer le partenariat de létalité synthétique entre BRCA1 et PARP1 qui est un phénomène connu [47, 42]. L'utilisation d'inhibiteur

de PARP1 chez les patients présentant une déficience en BRCA1 empêche toute possibilité de réparation de l'ADN, ce qui entraîne des dommages permanents de l'ADN et induit l'apoptose de la cellule cancéreuse [35, 61]. L'inférence prévoit également que l'inhibition de l'EGFR serait létale synthétique avec des mutations BRCA1. Ceci est corroboré par l'observation que les propriétés de prolifération des cellules déficientes en BRCA sont sensibles à l'inhibition de l'EGFR par l'erlotinib [17]. Nous n'avons trouvé aucun travail publié suggérant que l'inhibition de l'ERK1/2 dans de telles cellules serait létale synthétique.

3 Cancer de la vessie

Le cancer de la vessie est un cancer fréquent dans les pays occidentaux. Les thérapies utilisées actuellement sont la chirurgie, la radiation et des chimiothérapies non ciblées [51]. Ce cancer est associé à un fort taux de récidive. Des thérapies ciblées sont actuellement en développement ou au stade d'essai clinique sans démonstration claire de leur efficacité d'après [40].

3.1 Présentation du modèle

Le modèle sauvage du cancer de la vessie développé dans [84] est un réseau multivalué décrit par un graphe d'interaction composé de 30 sommets et 84 interactions. Il contient les voies et gènes principaux impliqués dans la régulation du cycle cellulaire et de l'apoptose dans les cellules cancéreuses de la vessie. Le réseau contient 5 variables multivaluées (pouvant prendre les valeurs 0,1 et 2) (E2F3, E2F1, CHEK1/2, ATM and Apoptosis), 4 variables d'*entrée* (EGFRstimulus, FGFR3stimulus, DNAdamage and Growthinhibitor) et 3 variables de *sortie* : Proliferation, Apoptosis et GrowthArrest. Les variables d'entrée sont des variables dont la fonction associée est une constante booléenne assimilée à des stimuli extracellulaires (EGFRstimulus, FGFR3stimulus and GrowthInhibitor) ou intracellulaires (DNAdamage) tandis que les variables de sortie sont des variables «conceptuelles» n'agissant sur aucune autre variable et qui correspondent à des «observateurs» de l'activité des gènes représentatifs d'une fonctionnalité cellulaire (c'est à dire les biomarqueurs). Notre cadre théorique s'appliquant aux

réseaux booléens exclusivement, nous avons d'abord traduit ce réseau multivalué en réseau booléen. Pour cela, nous avons utilisé l'algorithme décrit par [31] consistant à encoder une variable entière m par m variables booléennes. Cet encodage permet de préserver le comportement dynamique du réseau. Nous obtenons un réseau booléen à 35 variables et 117 interactions, où chaque variable multivaluée x est représentée par deux variables x et $x\lambda 2$. La sortie Apoptosis du réseau de référence étant multivalué, le réseau «booléanisé» a 4 lecteurs de comportement cellulaire : Proliferation, Apoptosis, Apoptosisλ2 et GrowthArrest.

Afin d'avoir un modèle générique sans variable à valeur constante, nous ajoutons des boucles d'auto-régulation positives sur les variables d'entrée. Le réseau booléen sauvage ainsi défini a 20 états stables correspondant aux 20 états stables obtenus dans l'article de référence (Table S2 du Matériel supplémentaire) par simulation de toutes les interprétations possibles sur les variables d'entrée. Les 20 états stables peuvent être classé en 3 phénotypes sur la base de l'état des observateurs :

— Les états stables correspondant au *phénotype prolifératif* ont la signature suivante :

$$\{\text{Proliferation} \mapsto 1, \text{Apoptosis} \mapsto 0, \text{Apoptosis}\lambda 2 \mapsto 0, \text{GrowthArrest} \mapsto 0\}.$$

— Les états stables correspondant au *phénotype apoptotique* ont la signature suivante :

$$\{\text{Proliferation} \mapsto 0, \text{Apoptosis} \mapsto 1, \text{Apoptosis}\lambda 2 \mapsto 0, \text{GrowthArrest} \mapsto 1\}.$$

— Les états stables correspondant au *phénotype d'arrêt de la croissance* ont la signature suivante :

$$\{\text{Proliferation} \mapsto 0, \text{Apoptosis} \mapsto 0, \text{Apoptosis}\lambda 2 \mapsto 0, \text{GrowthArrest} \mapsto 1\}.$$

Le réseau sauvage possède 9 états stables apoptotiques, 7 états stables d'arrêt de la croissance et 4 états stables prolifératifs. Le graphe d'interaction issu de l'article, les formules booléennes obtenues par traduction du réseau multivalué et les états stables colorés selon le phénotype qu'ils représentent sont donnés dans l'annexe de ce cha-

pitre (section 4).

3.2 Requêtes et solutions

En suivant le même raisonnement que pour la spécification des requêtes sur le modèle du cancer du sein décrite plus haut, l'inférence des perturbations causales de la transformation du modèle sauvage en modèle cancéreux s'exprime par une requête spécifiant que tous les états stables et au moins un état stable doivent avoir une signature de phénotype prolifératif. Cette requête de reprogrammation s'exprime donc de la façon suivante : $\Box p \wedge \Diamond p$ avec p correspondant au marquage de la signature du phénotype prolifératif, soit :

$$p = \text{Proliferation} \wedge \neg\text{Apoptosis} \wedge \neg\text{Apoptosis}\lambda 2 \wedge \neg\text{GrowthArrest}.$$

#	Noyaux de reprogrammation de $\Box p \wedge \Diamond p$
1	E2F3 = 0, CyclinD1 = 1, EGFR = 1
2	E2F3 = 0, CyclinD1 = 1, FGFR3 = 1, PI3K = 1
3	E2F3 = 0, CyclinD1 = 1, RAS = 1, PI3K = 1
4	E2F3 = 0, p16INK4a = 0, EGFR = 1, MDM2 = 1
5	E2F3 = 0, p16INK4a = 0, EGFR = 1, PI3K = 1
6	E2F3 = 0, p16INK4a = 0, EGFR = 1, TP53 = 0
7	E2F3 = 0, p16INK4a = 0, FGFR3 = 1, PI3K = 1
8	E2F3 = 0, p16INK4a = 0, RAS = 1, PI3K = 1
9	E2F3 = 0, RBL2 = 0, EGFR = 1, RB1 = 0
10	E2F3 = 0, RBL2 = 0, EGFR = 1, MDM2=1
11	E2F3 = 0, RBL2 = 0, EGFR = 1, p16INK4a = 1
12	E2F3 = 0, RBL2 = 0, EGFR = 1, PI3K = 1
13	E2F3 = 0, RBL2 = 0, EGFR = 1, TP53 = 0
14	E2F3 = 0, RBL2 = 0, FGFR3 = 1, PI3K = 1
15	E2F3 = 0, RBL2 = 0, RAS = 1, PI3K = 1

Tableau 3.7 – Solutions de l'inférence pour l'inférence des gènes causaux de la transformation de modèle sauvage à modèle prolifératif.

Les algorithmes d'inférence PLNE utilisés pour l'inférence des noyaux de reprogrammation ne pouvant traiter un réseau booléen de cette taille contrôlé sur toutes les

variables, nous avons sélectionné les gènes d'intérêt de façon à pouvoir comparer les résultats obtenus par inférence aux résultats obtenus par simulation dans l'article de référence. Les gènes d'intérêt sont les 11 gènes de l'ensemble suivant :

{TP53, E2F3, FGFR3, p16INK4a, RAS, CyclinD1, PI3K, MDM2, EGFR, RBL2, RB1}.

Ils correspondent à 12 variables dans le réseau booléen, E2F3 étant un gène multivalué et donc représenté par 2 variables E2F3 et E2F3λ2. De façon à réduire encore le temps de calcul, nous fixons le nombre de variables pouvant être gelées simultanément à 7 et inférons les noyaux de reprogrammation sur tous les ensembles de 7 variables contrôlées parmi les 12 variables d'intérêt (soit 792 combinaisons de 7 variables contrôlées, temps de calcul de 2079 secondes/35 minutes). L'union de toutes les solutions trouvées est donnée dans le tableau 3.7.

3.3 Analyse des solutions

L'analyse des solutions de l'inférence des gènes causaux du cancer sur le modèle de la vessie suit le même schéma que l'analyse des solutions pour le modèle du cancer du sein : nous comparons d'abord la valeur du gel pour chaque gène individuellement, de façon à vérifier que l'interprétation du rôle du gène *i.e.*, oncogène *versus* gène suppresseur de tumeur (GST) est conforme à la connaissance biologique. Puis, nous comparons les solutions dans leur globalité. Cette comparaison n'est pas faite sur la base de la littérature comme pour le cancer du sein mais en utilisant les données patients utilisées dans l'article de référence.

3.3.1 Oncogènes *versus* GST

Comme pour les solutions du cancer du sein, nous avons comparé les valeurs de gel des variables dans les solutions inférées avec leur classification en tant qu'oncogène ou GST à partir des tableaux de [100]. Dans le tableau 3.8, nous donnons, pour chaque variable affectée par un gel dans au moins une des solutions de l'inférence, le symbole par lequel il est identifié dans les tableaux de [100], ses scores oncogène et GST, sa classification et le tableau de référence dans lequel peut être retrouvée cette classi-

Variable	Symbole	Oncogène (%)	GST (%)	Rôle	Table
Gel à 0					
TP53	TP53	73	20	GST	S2A
RB1	RB1	4	80	GST	S2A
Gel à 1					
PI3K	PI3KCA	95	1	oncogène	S2A
EGFR	EGFR	97	0	oncogène	S2A
FGFR3	FGFR3	99	0	oncogène	S2A
CyclinD1	CCND1	-	-	oncogène	S2B
	HRAS	96			
RAS	KRAS	100	0	oncogene	S2A
	NRAS	99			
MDM2	-	-	-	oncogène	S2B
Gel à 0 ou à 1 selon les solutions					
p16INK4a	CDKN2A	32	49	GST	S2A

Tableau 3.8 – Comparaison de la valeur du gel dans les solutions de l'inférence aux score oncogènes et GST et au rôle attribué aux gènes dans [100].

fication. Pour les deux gènes qui ne sont pas représentés dans les tableaux de l'article de référence (soit E2F3 et RBL2), nous avons réalisé une recherche bibliographique. E2F3 est identifié comme oncogène dans toutes les références trouvées [38, 44, 79], le rôle de RBL2 est moins documenté mais il est identifié comme GST par sa similarité structurelle avec RB1 [30, 20, 107].

On observe donc que les gènes TP53, RB1 et RBL2 sont correctement assimilés à des GST par leur gel à 0 dans les solutions et que les gènes PI3K, EGFR, FGFR3 et CCND1 sont correctement assimilés à des oncogènes par leur gel à 1 dans les solutions. E2F3 par contre doit être gelé à 0 dans toutes les solutions alors qu'il est clairement identifié comme un oncogène dans la littérature. CDKN2A peut être gelé à 1 ou à 0 selon les solutions. De plus ses scores oncogène/GST sont plus partagés que pour les autres gènes. Nous avons donc réalisé une recherche bibliographique concernant son rôle. Deux articles [105, 86] suggèrent que le rôle de CDKN2A est plus complexe que celui d'un oncogène ou d'un GST strict et remarquent que ce gène peut être sous-exprimé dans certaines tumeurs ou sur-exprimé dans d'autres selon le contexte mutationnel.

Sur les 11 gènes étudiés, les solutions de l'inférence sont consistantes en termes de rôle attribué aux gènes pour 10 de ces gènes. De plus, les solutions suggèrent que les rôles des gènes peuvent être plus complexes, ce qui est confirmé par la littérature.

3.3.2 Comparaison globale

La comparaison globale a pour objectif de déterminer si les différentes solutions de reprogrammation sont représentatives des profils de perturbations observés chez les patients à partir des données issues de l'article [84].

Précision et sensibilité. Une solution de reprogrammation est *valide* si elle est incluse dans au moins un des profils de perturbation. Un profil est *représenté* par l'ensemble des solutions de la reprogrammation s'il inclut au moins une solution. La *précision* d'un ensemble de solutions de reprogrammation correspond au ratio de solutions valides sur le nombre total de solutions de la reprogrammation. La *sensibilité* correspond au ratio de profils de perturbation représentés sur le nombre total de profils. Afin de vérifier que la précision des solutions de la reprogrammation est supérieure à l'aléatoire, nous l'avons comparé à des ensembles de solutions générés aléatoirement. De même, nous avons comparé la sensibilité des solutions pour un ensemble de profils à des ensembles de profils générés aléatoirement.

Formatage des données et génération aléatoire de profils et de solutions. Les données concernent 178 échantillons prélevés chirurgicalement des tumeurs de patients et dont les perturbations des 11 gènes précédemment cités ont été évaluées en termes de présence/absence de mutations ou de variation du nombre de copies du gène (noté **VNC**) dans le génome. Dans les échantillons considérés, les gènes TP53, FGFR3, RAS, PI3KCA peuvent être affectés par des mutations et les gènes CCND1, MDM2, E2F3, EGFR, RBL2, CDKN2A et RB1 par des VNC. Dans les données issues de [84], chaque patient est associé à un *profil de perturbation*. Un profil de perturbation associe à chaque gène de l'ensemble $\{TP53, FGFR3, RAS, PI3KCA\}$ une valeur 1 si le gène est muté chez le patient ou une valeur 0 si ce gène n'est pas muté et à chaque gène de l'ensemble $\{CCND1, MDM2, E2F3, EGFR, RBL2, CDKN2A, RB1\}$ une valeur de 0 si le gène n'est pas affecté par une VNC, ou une valeur de -1 si une des deux copies du gène est man-

quante, de -2 si les deux copies sont manquantes, de 1 si on observe la présence d'une copie supplémentaire du gène ou de 2 si on oberve deux copies supplémentaires du gène.

Afin de pouvoir comparer les profils de perturbation aux solutions de la reprogrammation, nous avons formaté ces données de la façon suivante. N'ayant pas d'information sur l'effet fonctionnel des mutations (baisse ou augmentation de l'expression du gène), un profil patient contient le gène uniquement si celui-ci est muté chez le patient. En ce qui concerne les VNC, les VNC négatives étant souvent associées à une sous-expression de la protéine codée par le gène et les VNC positives à une surexpression de la protéine, un profil patient ne contient pas la perturbation si le gène n'est pas affecté, contient la perturbation $g = 0$ si le gène g a une VNC négative et contient la perturbation $g = 1$ si le gène g a une VNC positive. Par exemple, un patient dont le génome de la cellule cancéreuse contient une mutation du gène TP53, une VNC de CDKN2A de -2 et une VNC de RB1 de 1 sera représenté par le profil de perturbations $\{\text{TP53}, \text{CDKN2A} = 0, \text{RB1} = 1\}$. Les données formatées correspondent à un ensemble de profils de perturbations. 12 de ces profils sont vides et la suppression des duplicats génère un ensemble de 121 profils différents. Nous supprimons également les 20 profils contenant deux perturbations ou moins. La perturbation $\text{E2F3} = 0$ présente dans tous les noyaux n'étant pas pertinente (E2F3 est un oncogène), nous l'avons supprimée de toutes les solutions.

La génération aléatoire de solutions est réalisée par échantillonnage (tirage sans remise) selon une loi uniforme à partir de l'ensemble de toutes les perturbations (excepté celles affectant E2F3) et en respectant les contraintes suivantes :

— les solutions aléatoires ont le même nombre de gènes mutés et de gènes atteints par des VNC que les solutions calculées ;
— une solution ne peut contenir à la fois $g = 0$ et $g = 1$.

De même, la génération aléatoire de profils est réalisée par échantillonnage selon une loi uniforme à partir de l'ensemble de toutes les perturbations et en respectant les contraintes suivantes :

— les profils aléatoires ont le même nombre de gènes mutés et de gènes atteints par des VNC que les profils patients issus des données ;

— un profil ne peut contenir à la fois $g = 0$ et $g = 1$.

Résultats. Dix des quinze solutions inférées sont incluses dans au moins un profil, ce qui fait une précision de 66,67%. La précision moyenne pour 500 ensembles de 15 solutions aléatoires est de 42,92%. Trente-six des quatre-vingt onze profils patients incluent au moins une solution de reprogrammation, ce qui fait une sensibilité de 39,56%. La sensibilité moyenne pour 500 ensembles de 91 profils aléatoires est de 18,19%. Ces résultats suggèrent que les solutions de la reprogrammation du cancer de la vessie ont un pouvoir prédictif supérieur à l'aléatoire puisqu'elles sont à la fois plus précises et plus sensibles que le hasard.

Redondance et coopération. Dans l'article de référence, les auteurs ont évalué les relations d'exclusion mutuelle et de co-occurence existant entre les différentes paires de perturbations dans les données, leurs résultats sont donnés sur la figure 3.4. Leurs résultats, évalués statistiquement, ne peuvent être exhaustifs : certaines relations ne peuvent être établies du fait du manque de données. La relation d'exclusion mutuelle entre FGFR3 et RAS est représentée dans les solutions, dont aucune ne contient simultanément des perturbations des deux gènes. Il en est de même pour la relation d'exclusion mutuelle entre TP53 et FGFR3. La co-occurence de perturbations de FGFR3 et CDKN2A est retrouvée dans la solution 7, la co-occurence de FGFR3 et PIK3CA dans les solutions 2, 7 et 1 et la co-occurence de CDKN2A et PIK3CA dans la solution 7. La relation de co-occurence entre CDKN2A et cycline D1 n'est représentée. Les auteurs ont observés que leur perturbation simultanée est responsable d'une baisse du nombre d'états stables associés à l'arrêt de la croissance mais n'a pas d'incidence sur les états stables apoptotiques. La relation de co-occurence entre FGFR3 et Cyclin D1 n'est pas non plus représentée. Les auteurs avaient remarqué que le modèle ne représentait pas cette co-occurence. Les relations concernant E2F3 ne peuvent être analysées car la perturbation de E2F3 dans les solutions de reprogrammation n'est pas conforme aux observations biologiques.

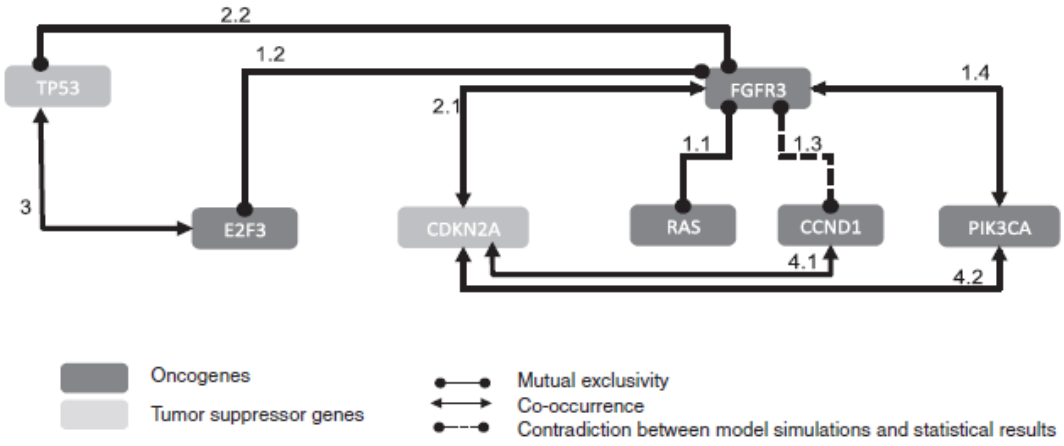


Figure 3.4 – Relations d'exclusion mutuelle et de coopération entre les gènes perturbés dans les données patients (reproduit à partir de [84]).

4 Annexe du chapitre

Cette annexe contient :

1. les tableaux présentant les noyaux de reprogrammation avec paramètres de contrôle en Gel-U pour l'inférence des gènes causaux (tableau 3.9) et des cibles thérapeutiques (tableau 3.10) pour le cancer du sein ;
2. le graphe d'interaction (figure 3.5) du modèle multivalué du cancer de la vessie présenté dans l'article [84] ;
3. les tableaux 3.11, 3.12, et 3.13 contenant les formules propositionnelles après transformation du réseau multivalué en réseau booléen ;
4. les tableaux 3.14 et 3.15 présentant les états stables du réseau booléen.

#	Noyaux de reprogrammation de $\Box p$ en Gel-U
1	BRCA1 $\dashv$ EGFR = 0, GSK3β $\dashv$ CYCD1 = 0
2	BRCA1 $\dashv$ EGFR = 0, PTEN $\dashv$ PI3KCA = 0
3	BRCA1 $\dashv$ EGFR = 0, BRCA1 $\dashv$ TP53 = 0
4	BRCA1 $\dashv$ EGFR = 0, BRCA1 $\dashv$ CYCD1 = 0
5	BRCA1 $\dashv$ EGFR = 0, TP53 $\rightarrow$ PTEN = 0

Tableau 3.9 – Solutions de l'inférence des perturbations causales de la transformation de modèle sauvage à modèle cancéreux avec des paramètres de contrôle en Gel-U pour le modèle du cancer du sein.

#	Noyaux de reprogrammation de $\Diamond q$ en Gel-U
1	ERK1/2 $\longrightarrow$ PARP1 = 0
2	EGFR $\longrightarrow$ ERK1/2 = 0
3	PARP1 $\longrightarrow$ CYCD1 = 0, PARP1 $\dashv$ TP53 = 0

Tableau 3.10 – Solutions de l'inférence de cibles thérapeutiques pour le cancer de type BRCA1 avec des paramètres de contrôle en Gel-U pour le modèle du cancer du sein.

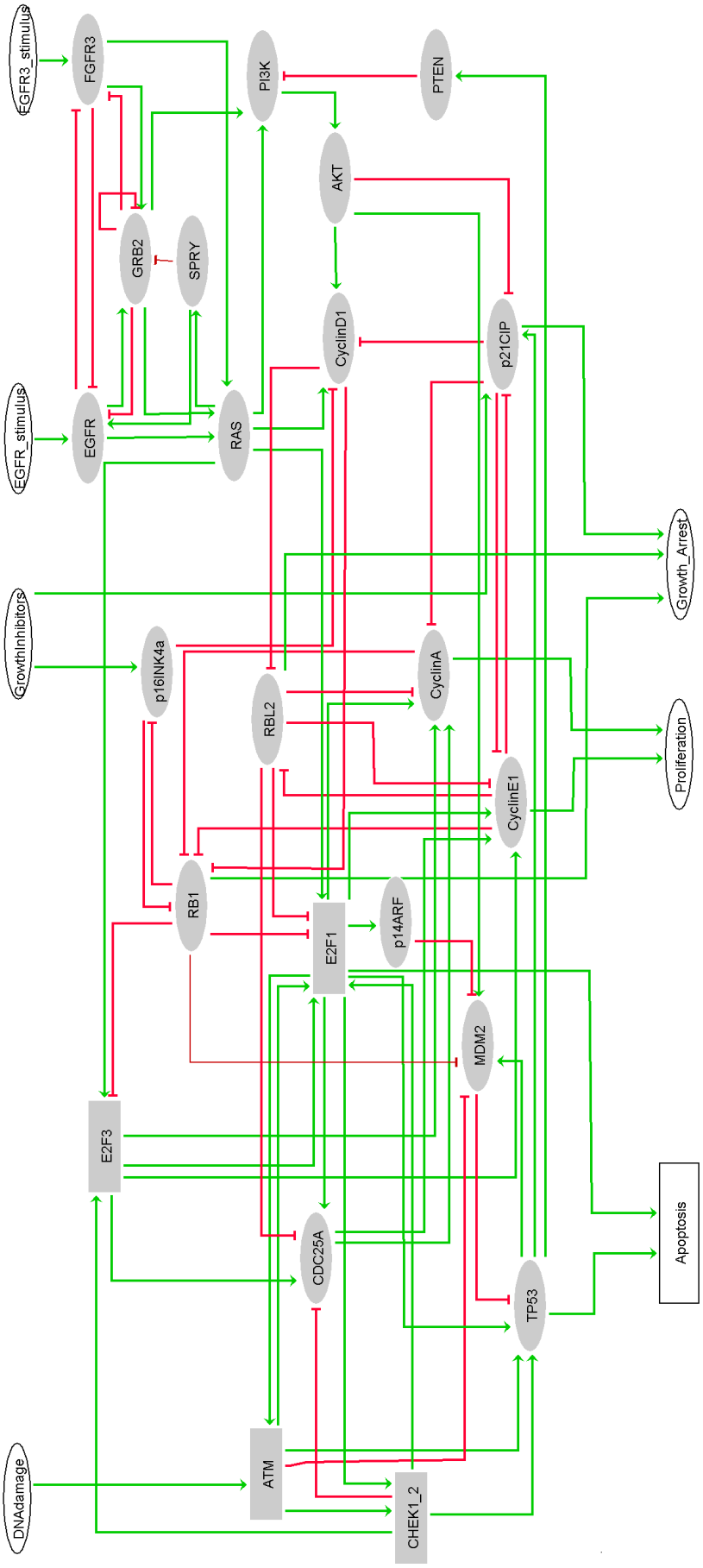


Figure 3.5 – Graphe d'interaction du modèle du cancer de la vessie issu de [84].

Variable	Règle d'activation
EGFR.stimulus	EGFR.stimulus
FGFR3.stimulus	FGFR3.stimulus
DNAdamage	DNAdamage
GrowthInhibitors	GrowthInhibitors
Proliferation	CyclinA ∨ CyclinE1
Apoptosis	(E2F1 ∨ ¬E2F1λ2) ∧ (E2F1λ2 ∨ TP53)
Apoptosisλ2	E2F1 ∧ E2F1λ2
GrowthArrest	p21CIP ∨ RB1 ∨ RBL2
EGFR	(EGFRstimulus ∨ SPRY) ∧ ¬FGFR3 ∧ ¬GRB2
FGFR3	¬EGFR ∧ FGFR3stimulus ∧ ¬GRB2
E2F1	(ATM ∨ ¬ATMλ2) ∧ (CHEK1/2 ∨ ¬CHEK1/2λ2) ∧ (E2F3 ∨ ¬E2F3λ2) ∧ (E2F3 ∨ RAS) ∧ ¬RB1 ∧ ¬RBL2
E2F1λ2	ATM ∧ ATMλ2 ∧ CHEK1/2 ∧ CHEK1/2λ2 ∧ (E2F3 ∨ ¬E2F3λ2) ∧ (E2F3λ2 ∨ RAS) ∧ ¬RB1 ∧ ¬RBL2

Tableau 3.11 – Ensemble de formules du réseau booléen obtenu par transformation du réseau multivalué de l'article de référence auquel on a ajouté des boucles de régulation positive sur les entrées du graphe d'interaction.

Variable	Règle d'activation
RAS	EGFR ∨ FGFR3 ∨ GRB2
E2F3	(CHEK1/2 ∨ ¬CHEK1/2λ2) ∧ RAS ∧ ¬RB1
E2F3λ2	CHEK1/2 ∧ CHEK1/2λ2 ∧ RAS ∧ ¬RB1
CyclinD1	(AKT ∨ RAS) ∧ ¬p16INK4a ∧ ¬p21CIP
CyclinE1	CDC25A ∧ (E2F1 ∨ ¬E2F1λ2) ∧ (E2F1 ∨ E2F3) ∧ (E2F3 ∨ ¬E2F3λ2) ∧ ¬p21CIP ∧ ¬RBL2
CyclinA	CDC25A ∧ (E2F1 ∨ ¬E2F1λ2) ∧ (E2F1 ∨ E2F3) ∧ (E2F3 ∨ ¬E2F3λ2) ∧ ¬p21CIP ∧ ¬RBL2
CDC25A	¬CHEK1/2 ∧ ¬CHEK1/2λ2 ∧ (E2F1 ∨ ¬E2F1λ2) ∧ (E2F1 ∨ E2F3) ∧ (E2F3 ∨ ¬E2F3λ2) ∧ ¬RBL2
p16INK4a	GrowthInhibitors ∧ ¬RB1
RB1	¬CyclinA ∧ ¬CyclinD1 ∧ ¬CyclinE1 ∧ ¬p16INK4a
RBL2	¬CyclinD1 ∧ ¬CyclinE1
p21CIP	¬AKT ∧ ¬CyclinE1 ∧ (GrowthInhibitors ∨ TP53)
ATM	DNAdamage ∧ (E2F1 ∨ ¬E2F1λ2)

Tableau 3.12 – Ensemble de formules du réseau booléen obtenu par transformation du réseau multivalué de l'article de référence auquel on a ajouté des boucles de régulation positive sur les entrées du graphe d'interaction (suite).

Variable	Règle d'activation
ATMλ2	DNAdamage ∧ E2F1
CHEK1/2	ATM ∧ (E2F1 ∨ ¬E2F1λ2)
CHEK1/2λ2	ATM ∧ E2F1
MDM2	(AKT ∨ TP53) ∧ ¬ATM ∧ ¬ATMλ2 ∧ ¬p14ARF ∧ ¬RB1
TP53	(ATM ∨ ¬ATMλ2) ∧ (ATM ∨ E2F1λ2) ∧ (CHEK1/2 ∨ ¬CHEK1/2λ2) ∧ (CHEK1/2 ∨ E2F1λ2) ∧ (E2F1 ∨ ¬E2F1λ2) ∧ ¬MDM2
p14ARF	E2F1
PTEN	TP53
PI3K	GRB2 ∧ ¬PTEN ∧ RAS
AKT	PI3K
GRB2	(EGFR ∨ FGFR3) ∧ (EGFR ∨ ¬GRB2) ∧ (EGFR ∨ ¬SPRY)
SPRY	RAS

Tableau 3.13 – Ensemble de formules du réseau booléen obtenu par transformation du réseau multivalué de l'article de référence auquel on a ajouté des autoboucles sur les entrées du graphe d'interaction (suite bis).

#	EGFR.stimulus	FGFR3.stimulus	DNAdamage	GrowthInhibitors	Proliferation	Apoptosis	ApoptosisΛ2	GrowthArrest	EGFR	FGFR3	RAS	E2F1	E2F1Λ2	E2F3	E2F3Λ2	CyclinD1	CyclinE1	CyclinA	CDC25A	p16INK4a	RB1	RBL2	p21CIP	ATM	ATMΛ2	CHEK1/2	CHEK1/2Λ2	MDM2	TP53	p14ARF	PTEN	PI3K	AKT	GRB2	SPRY
1	1	1	1	1	0	1	0	1	0	1	1	0	0	1	0	0	0	0	0	1	0	1	1	1	0	1	0	0	1	0	1	0	0	0	1
2	1	1	1	1	0	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	1	0	1	0	0	0	1
3	1	1	1	0	0	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	1	0	1	0	0	0	1
4	1	1	0	1	1	0	0	0	0	1	1	1	0	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
5	1	1	0	1	0	0	0	1	0	1	1	0	0	1	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1
6	1	1	0	1	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1
7	1	1	0	0	1	0	0	0	0	1	1	1	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
8	0	1	1	1	0	1	0	1	0	1	1	0	0	1	0	0	0	0	0	1	0	1	1	1	0	1	0	0	1	0	1	0	0	0	1
9	0	1	1	1	0	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	1	0	1	0	0	0	1
10	0	1	1	0	0	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	1	0	1	0	0	0	1

Tableau 3.14 – États stables du réseau booléen modèle du cancer de la vessie. Les colonnes correspondant au sommets d'entrée (stimuli) du réseau sont colorées en gris, les états stables correspondant à un phénotype apoptotique sont colorés en rouge, ceux correspondant à un phénotype d'arrêt de la croissance en jaune et ceux correspondant à un phénotype prolifératif en orange.

#	EGFR.stimulus	FGFR3.stimulus	DNAdamage	Growthinhibitors	Proliferation	Apoptosis	ApoptosisA2	GrowthArrest	EGFR	FGFR3	RAS	E2F1	E2F1A2	E2F3	E2F3A2	CyclinD1	CyclinE1	CyclinA	CDC25A	p16INK4a	RB1	RBL2	p21CIP	ATM	ATMA2	CHEK1/2	CHEK1/2A2	MDM2	TP53	p14ARF	PTEN	PI3K	AKT	GRB2	SPRY
11	0	1	0	1	1	0	0	0	0	1	1	1	0	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
12	0	1	0	1	0	0	0	1	0	1	1	0	0	1	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1
13	0	1	0	1	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1
14	0	1	0	0	1	0	0	0	0	1	1	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
15	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0
16	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0
17	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0
18	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Tableau 3.15 – États stables du réseau booléen modèle du cancer de la vessie (suite). Les colonnes correspondant au sommets d'entrée (stimuli) du réseau sont colorées en gris, les états stables correspondant à un phénotype apoptotique sont colorés en rouge, ceux correspondant à un phénotype d'arrêt de la croissance en jaune et ceux correspondant à un phénotype prolifératif en orange

Chapitre 4

Comparaison de performances

Le calcul des noyaux de reprogrammation est un problème NP-difficile, dont la résolution est présentée au chapitre 2 par deux méthodes, PLNE-Noyau et DDB-Noyau. Dans ce chapitre, nous comparons les performances de ces deux méthodes en évaluant différentes techniques heuristiques permettant de réduire le temps de calcul. Ces techniques ont pour objectif de réduire la profondeur des DDBOR... concernent la spécification de *contraintes sur les paramètres de contrôle* pour PLNE et DDB et des tests sur des techniques de réduction de la taille du DDB avec des règles sur l'*ordre des variables* pour la construction du DDB et la *réduction des paramètres de contrôle positifs*.

1 PLNE-Noyau

PLNE-Noyau est une méthode itérative permettant le calcul des noyaux de reprogrammation un par un selon un mode de reprogrammation donné. La minimalité de ces impliquants implique qu'ils ne peuvent contenir à la fois un paramètre de contrôle actif u_i^0 gelant une variable à 0 et un paramètre de contrôle actif u_i^1 gelant la même variable à 1. Cela peut être décrit dans la formule spécifiant les observations par l'ajout de clauses $u_i^1 \vee u_i^0$ car la seule configuration pour laquelle cette contrainte est fausse est que les deux paramètres de contrôle soient actifs (donc égaux à 0 par convention). Nous avons testé, pour les modes de possibilité et de nécessité, les performances de la méthodes PLNE-Noyau avec ou sans cette contrainte. Les calculs sont réalisés sur

# agents	Temps de calcul		Nombre de clauses	
	Moyenne	Ecart-type	Moyenne	Ecart-type
10	0.0296402	0.0150994	26.1	13.1385
20	0.0530403	0.0415523	30.8	42.8186
30	0.767525	2.24556	167.1	8.32814
40	0.0795605	0.0220037	227.25	7.34757
50	1.38763	4.48424	287.9	10.9732
60	1.35175	5.7259	347.3	13.6925

Tableau 4.1 – Temps de calcul moyen (en secondes) et nombre moyen de clauses de la FNC représentant les observations pour les requêtes de Possibilité.

des réseaux booléens monotones de taille croissante générés aléatoirement et dont les variables sont contrôlées par des paramètres de contrôle en Gel-D (définition). La génération aléatoire des réseaux booléens est réalisée selon les étapes suivantes :

1. génération aléatoire d'un graphe d'interaction non dirigé de type scale-free selon la méthode décrite par Albert Barabasi [9] avec l'ajout d'un sommet connecté par deux interactions à chaque étape ;
2. affectation aléatoire équiprobable d'une direction et d'un signe à chaque interaction ;
3. affectation aléatoire d'une fonction booléenne à chaque sommet du graphe et dépendant des interacteurs du sommet.

De plus, un marquage est défini arbitrairement sur deux variables (x_1, x_2), qui sont les seules à ne pas être contrôlées. Le calcul des noyaux est arrêté lorsque le temps de calcul excède 10 minutes. Pour chaque taille, 10 réseaux sont générés. Les tests sont réalisés sur un ordinateur HP EliteBook 820 G2 Intel Core (2.30 GHz i5-5300U CPU avec 16 Go de mémoire.

Pour les requêtes de Possibilité, le temps de calcul des noyaux est compris entre 0 et 1 seconde pour les réseaux jusqu'à 50 variables et entre 1 et 2 secondes pour les réseaux de 50 à 60 variables. La méthode passe donc à l'échelle pour les tailles de modèles biologiques publiés. La moyenne du temps de calcul pour les 10 réseaux en fonction de la taille du réseau booléen pour les requêtes de nécessité est donné dans la figure 4.2.

La méthode PLNE-Noyau est donc beaucoup plus efficace pour le calcul des noyaux

# agents	Temps de calcul		Nombre de clauses	
	Moyenne	Ecart-type	Moyenne	Ecart-type
6	0.0826805	0.0472362	34.9	14.9628
7	0.146641	0.13613	71.55	34.2936
8	0.634924	0.629443	197.6	106.937
9	3.69098	3.64535	534.2	185.472
10	18.1445	18.1445	1272.9	810.262
11	188.473 (1)	294.761	3539.7	1996.91
12	-	-	-	-

Tableau 4.2 – Temps de calcul moyen (en secondes) et nombre moyen de clauses de la FNC représentant les observations pour les requêtes de Nécessité

de la reprogrammation pour les requêtes de possibilité que pour les requêtes de nécessité. En effet, pour les requêtes de nécessité, le temps de calcul devient supérieur à 10 minutes à partir de réseaux à 11 variables. Par ailleurs, la contrainte sur les paramètres de contrôle n'affecte pas le temps de calcul.

2 DDB-Noyau

DDB-Noyau est une méthode récursive calculant tous les noyaux de reprogrammation à la fois sur un diagramme de décision binaire ordonné et réduit (DDBOR) représentant la formule spécifiant les observations selon un mode de reprogrammation donné. Nous avons donc testé les performances de DDB-Noyau avec diverses options. Ces options concernent :

1. les contraintes sur les contrôles par ajout de clauses à la formule spécifiant les observations ;
2. la définition d'un ordre sur les variables pour la construction du DDBOR ;
3. l'ajout d'une règle de réduction des paramètres de contrôle positifs au cours de la construction du DDBOR.

Les deux dernières options dépendent de la méthode DDB-Noyau puisqu'elles concernent la construction d'un DDBOR à partir d'une formule.

Contraintes sur les contrôles. Le nombre d'opérations à réaliser pour le calcul des impliquants minimaux sur les DDBOR est d'autant plus réduit que la taille des chemins menant aux sommets terminaux 0 et 1 est petite. L'ajout des contraintes $u_i^1 \vee u_i^0$ dans la formule spécifiant les observations augmente la taille de la formule en FNC mais réduit dans certains cas la taille du DDBOR représentant la formule en le dirigeant plus rapidement vers les sommets terminaux.

$$
\begin{aligned}
f = {} & (\neg d_1^0 \vee d_2^1 \vee d_3^1) \wedge (\neg d_1^0 \vee \neg d_2^0 \vee d_3^1) \wedge (\neg d_1^0 \vee \neg d_2^0 \vee \neg d_3^1) \\
& \wedge (\neg d_1^0 \vee d_2^1 \vee \neg d_3^1) \wedge (\neg d_2^0 \vee d_1^1 \vee \neg d_3^1) \wedge (d_1^1 \vee d_2^1 \vee \neg d_3^1) \\
& \wedge (\neg d_1^0 \vee \neg d_3^0 \vee \neg d_2^1) \wedge (\neg d_1^0 \vee \neg d_2^1 \vee d_3^1) \wedge (\neg d_3^0 \vee d_1^1 \vee \neg d_2^1) \\
& \wedge (d_1^1 \vee \neg d_2^1 \vee d_3^1) \wedge (d_2^0 \vee d_3^0 \vee d_1^1 \vee \neg d_2^1 \vee \neg d_3^1) \\
& \wedge (d_1^0 \vee d_1^1) \wedge (d_2^0 \vee d_2^1) \wedge (d_3^0 \vee d_3^1)
\end{aligned}
$$

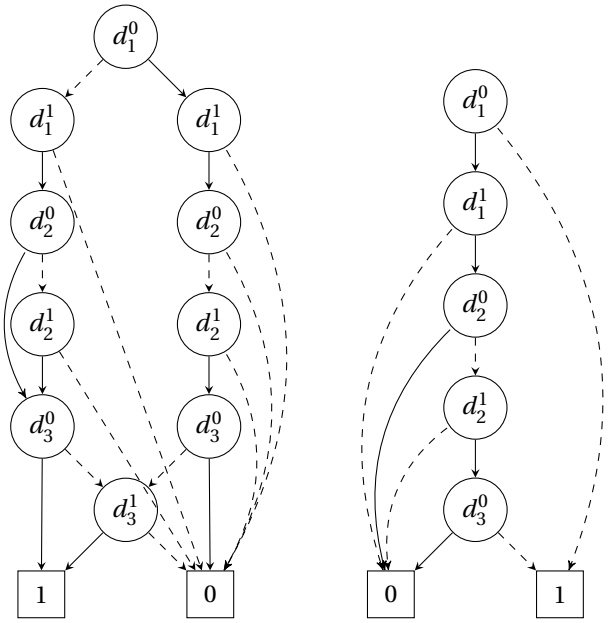


Figure 4.1 – Formule f (haut), sa représentation par un DDBOR (gauche) et un DDBPR (droite).

Ordre des variables. L'ordre des variables pour la construction a une grande influence sur la taille du DDBOR. La détermination de l'ordre des variables générant un DDBOR de taille minimale est un problème NP-complet [16]. Nous utilisons une heuristique se fondant le graphe d'interaction du réseau. En effet, la formule représentant la condition de stabilité dépend de la spécification du réseau et par conséquent, de son graphe d'interaction. L'heuristique repose sur la détermination d'un ordre selon

«l'importance» des variables. L'importance est mesurée par le calcul d'une centralité du sommet du graphe d'interaction représentant la variable. Après des tests préliminaires, la mesure de centralité de proximité («closeness centralité») apparait la plus efficace. Pour un graphe à N sommets, la centralité de proximité d'un sommet x, notée $\mathrm{C}(x)$ correspond à l'inverse de la somme de la longueur du plus court chemin de x à chaque autre sommet y du graphe, soit :

$$\mathrm{C}(x) = \frac{\mathrm{N}}{\sum_y d(x,y)}$$

Selon cette définition, un sommet x est d'autant plus «central» que $\mathrm{C}(x)$ est petit. Une variable précèdera une autre dans l'ordre si sa centralité de proximité est plus petite. L'hypothèse expliquant cette efficacité est que le contrôle des variables ayant les plus faibles mesure de centralité de proximité conduira en moins d'étape aux sommets terminaux : ces sommets sont plus déterminant dans l'évaluation à 0 ou 1 de la fonction booléenne.

Réduction des paramètres de contrôle positifs. La réduction des paramètres de contrôle positifs consiste à ajouter une règle de réduction supplémentaire à la construction du DDBOR. Cette règle supprime les sommets dont le successeur *high* est le sommet terminal 1 et le successeur *low* est le sommet terminal 0. En effet, ces sommets s'interprètent comme des littéraux de contrôle positifs qui ne font pas partie des noyaux (qui ne contiennent, par définiton, que des paramètres de contrôle négatifs). Dans le pire des cas, le DDBOR réduit des paramètres de contrôle positifs (noté DDBPR) et le DDBOR sont équivalents. Par exemple, dans la figure 4.1, le DDBOR de gauche représente la formule f sans cette réduction et le DDBOR de droite représente le DDBPR obtenu avec la règle de réduction des paramètres de contrôle positifs.

Temps de calcul. La figure 4.3 représente les moyennes et écart-types des temps de calcul de DDB-□-Noyau avec les différentes options. La courbe rouge correspond à la méthode intégrant les contraintes sur les paramètres de contrôle et la réduction des paramètres de contrôle positifs. On peut observer que ces deux options améliorent sensiblement le temps de calcul et permettent donc de traiter des réseaux de plus

grande taille. Enfin, sans heuristique sur l'ordre des variables pour la construction du DDBOR, ce dernier ne peut être construit en dix minutes pour des réseaux dépassant 8 à 9 variables.

3 Comparaison de PLNE-Noyau et DDB-Noyau

La figure 4.4 récapitule les résultats expérimentaux réalisés sur les mêmes réseaux booléens pour DDB-Noyau et PLNE-Noyau. Elle montre que, pour la Nécessité, la méthode DDB-Noyau est plus efficace. En effet, DDB-Noyau peut gérer des réseaux booléens de taille 16 avec un contrôle de toutes les variables en gel-D, ce qui représente une augmentation de 45% par rapport à la méthode PLNE.

En comparant le nombre de clauses des FNC pour les requêtes de Nécessité et de possibilité (tableaux 4.1 et 4.2), on constate que le nombre de clauses explose rapidement pour les requêtes de Nécessité, affectant les performances de PLNE-Noyau mais pas celles de DDB-Noyau car la représentation d'un DDBOR est canonique. Pour conclure, il apparaît que PLNE-Noyau est plus performant pour la résolution des requêtes de Possibilité et DDB-Noyau pour la résolution des requêtes de nécessité.

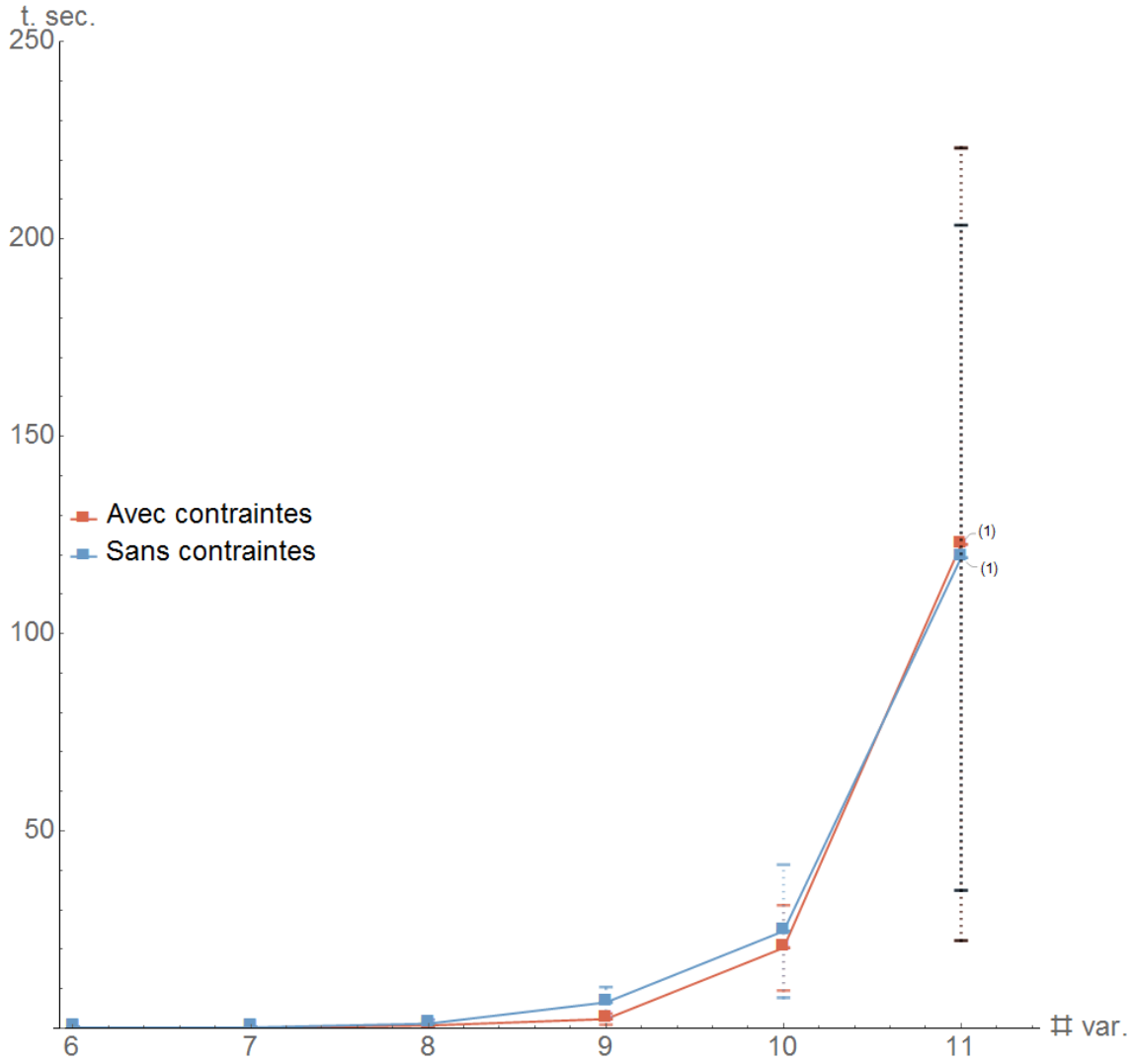


Figure 4.2 – Moyenne et écart-types des temps de calcul (en secondes) des noyaux de la reprogrammation par la méthode PLNE-Noyau pour des requêtes de Nécessité en fonction de la taille du réseau booléen. Les chiffres entre parenthèses correspondent au nombre de réseaux pour lesquels le calcul n'a pu s'achever avant les 10 minutes. Pour chaque taille, 10 réseaux sont générés.

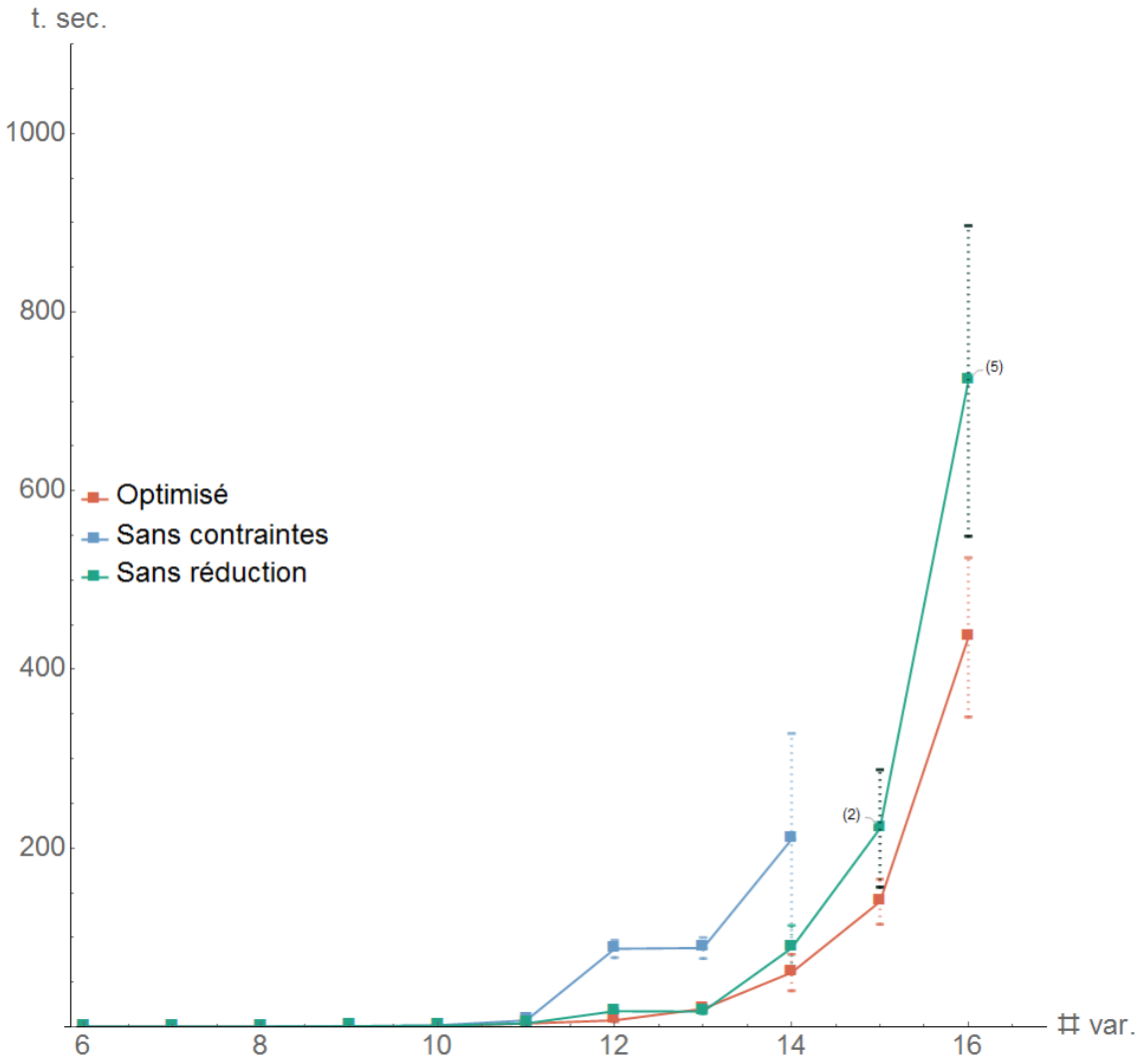


Figure 4.3 – Moyennes et écart-types des temps de calcul (en secondes) des noyaux de la reprogrammation par la méthode BDD-Noyau pour des requêtes de Nécessité en fonction de la taille du réseau booléen (10 réseaux/taille). Les chiffres entre parenthèses correspondent au nombre de réseaux pour lesquels le calcul n'a pu s'achever avant les 10 minutes. L'ordre des variables pour la construction du DDBOR est déterminé selon la centralité de proximité des variables contrôlées.

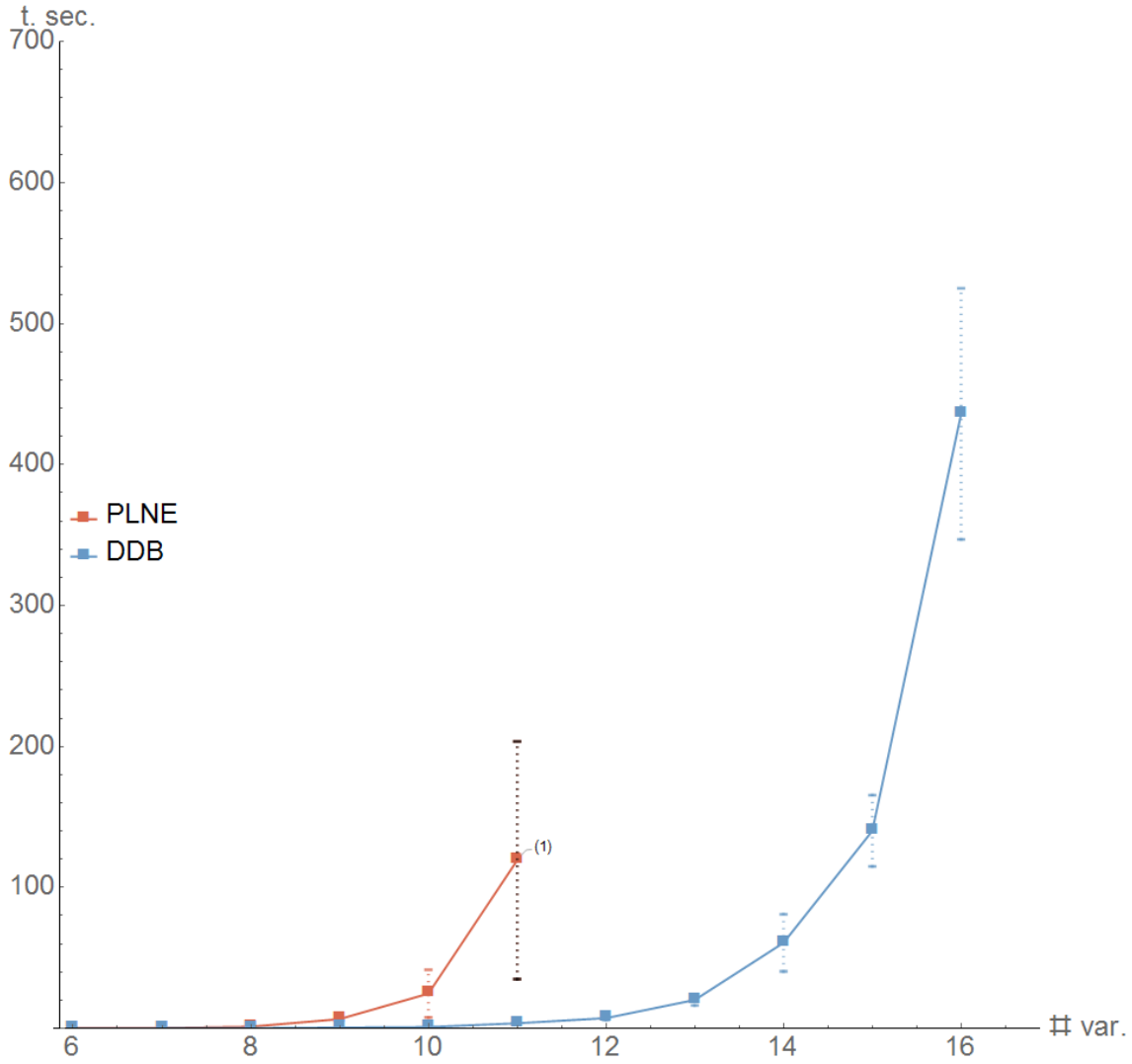


Figure 4.4 – Moyennes et écart-types des temps de calcul (en secondes) des noyaux de la reprogrammation par les méthodes PLNE-Noyau sans contraintes et BDD-Noyau optimisé pour des requêtes de Nécessité en fonction de la taille des réseaux booléens (10 réseaux/taille). Les chiffres entre parenthèses correspondent au nombre de réseaux pour lesquels le calcul n'a pu s'achever avant les 10 minutes.

Conclusion et perspectives

La médecine de précision est un champ de recherche en biologie moléculaire et médicale dont un enjeu majeur est la définition des maladies à l'échelle moléculaire et la découverte de nouvelles cibles thérapeutiques pour les traiter. Or, les maladies humaines les plus courantes sont des maladies dites complexes, qui sont dues à des perturbations multiples et diverses des systèmes moléculaires. Une des difficultés de cet enjeu réside donc dans la complexité de la relation entre l'information génétique et, plus largement, moléculaire, contenue dans les systèmes biologiques et le comportement de ces systèmes (la relation «génotype-phénotype»). La complexité de cette relation est notamment mise en lumière par la pléiotropie et l'épistasie des perturbations moléculaires. Dans cette thèse, nous avons soutenu que l'élaboration de cadres de modélisation fondés sur l'analyse de la dynamique des réseaux moléculaires constitue un élément essentiel pour déchiffrer cette relation et identifier les perturbations causales des maladies complexes et les modes d'action de leur traitement.

Nous proposons un cadre de modélisation de reprogrammation comportementale des systèmes dynamiques booléens qui assimile les causes des maladies complexes à des perturbations des fonctions dynamiques locales dont la conséquence est la reprogrammation de la dynamique globale de ces systèmes. Ce cadre de modélisation s'appuie, d'une part, sur les réseaux booléens contrôlés et, d'autre part, sur la définition de modes (Possibilité, Nécessité) permettant d'exprimer les objectifs de la reprogrammation. Les réseaux booléens contrôlés sont des réseaux booléens dans lesquels sont insérés des paramètres de contrôle modélisant les perturbations associées aux causes des maladies et aux modes d'action des médicaments. Les actions de contrôle sont de deux types : le Gel-D, qui gèle les valeurs des variables du réseaux et le Gel-U qui gèle les valeurs des interactions entre les variables du réseau. Les contrôles par gel de valeurs

s'assimilent aux perturbations des réseaux d'interaction biologiques provoquées par des mutations ou des médicaments. Les deux modes de reprogrammation : la possibilité et la nécessité, définissent les conséquences des perturbations sur les états stables du réseau booléen et modélisent donc les effets des maladies et des médicaments sur le comportement du système biologique.

À partir de ce cadre, nous décrivons des méthodes permettant le calcul exhaustif des noyaux de la reprogrammation, c'est-à-dire des ensembles minimaux d'action permettant la reprogrammation selon une requête d'objectif exprimée en fonction du mode. Nous démontrons que ce calcul s'exprime comme un problème d'inférence abductive en logique propositionnelle qui consiste à inférer les causes les plus explicatives de faits observables représentés par une formule booléenne au regard d'une théorie.

En nous appuyant sur les méthodes historiques permettant le calcul des impliquants premiers des fonctions booléennes, à savoir la procédure du consensus et son expression en problème de minimisation logique, nous développons deux méthodes permettant le calcul de l'ensemble des noyaux de la reprogrammation. La première méthode, appelée PLNE-Noyau, est un algorithme itératif fondé sur la programmation linéaire en nombres entiers et permettant le calcul des noyaux de reprogrammation un par un. Cette méthode permet également d'assigner des coûts aux perturbations et de prioriser ainsi le calcul des noyaux de reprogrammation minimisant ce coût. La seconde méthode, appelée DDB-Noyau, est un algorithme récursif dérivé de la procédure du consensus et permettant le calcul de tous les noyaux de reprogrammation à la fois à partir d'une représentation compactée de la formule booléenne en arbre de décision binaire ordonné et réduit [13, 11, 12].

Enfin, nous évaluons la pertinence du cadre de modélisation pour l'identification des perturbations causales responsables de la transformation d'une cellule saine en cellule cancéreuse et pour la découverte de cibles thérapeutiques permettant de pousser la cellule cancéreuse vers l'apoptose. Dans un premier temps, nous réalisons une preuve de concept de l'approche sur un modèle conçu pour le cancer du sein. Le cancer du sein est un modèle de maladie, dans le sens où il est classé en sous-types selon la cause moléculaire, et, dans le cas du cancer de sous-type BRCA1, il existe un trai-

tement chimiothérapeutique ciblé consistant à inhiber la molécule PARP1 qui est le partenaire synthétique létal de BRCA1. Nous montrons notamment que les perturbations causales de la transformation du modèle sain en modèle cancéreux sont compatibles avec la connaissance biologique : la valeur de gel des variables du réseau booléen discrimine les rôles des gènes en oncogènes et gènes suppresseurs de tumeurs et récupère la mutation causale du gène BRCA1. De plus, la récupération de l'apoptose à partir du modèle malade avec une mutation de BRCA1 récupère le phénomène de létalité synthétique entre PARP1 et BRCA1. La létalité synthétique constitue un traitement anti-cancéreux optimal car il cible spécifiquement les celllules tumorales. Dans un deuxième temps, nous appliquons ce même protocole à un modèle publié du cancer de la vessie afin de garantir la reproductibilité des résultats. Nous montrons que la ségrégation oncogènes et gènes supresseur de tumeurs est également observée dans les noyaux de reprogrammation. Nous avons également proposé d'autres prédictions à valider notamment concernant le rôle du gène CDNK2A.

Pour conclure, cette thèse présente donc une nouvelle approche théorique et algorithmique originale pour l'étude des causes des maladies complexes et de leur thérapie. Son application à des modèles de cellules donne des résultats cohérents avec la connaissance biologique dans le cas du cancer.

Perspectives. Fearon et Vogelstein ont démontré en 1990 que la tumorigenèse du cancer colorectal repose sur une séquence de perturbations de trois gènes [37], appelée «séquence de Vogelstein». Cette séquence suggère que l'acquisition d'un phénotype cancéreux nécessite de passer par différents stades intermédiaires et que, si ces mêmes perturbations arrivent dans un ordre différent de celui observé, la cellule meure avant de développer son potentiel cancéreux. Une perspective intéressante d'élargissement du cadre théorique consisterait à définir de nouveaux modes de reprogrammation permettant la prise en compte de la temporalité des perturbations. Ces nouveaux modes de reprogrammation pourraient être spécifiés soit en spécifiant une séquence de contrôles correspondant à une application successive des différents contrôles, soit par un contrôle rétro-actif se modifiant selon l'état des variables obtenues après une précédente reprogrammation.

Par ailleurs, les cellules cancéreuses peuvent être résistantes aux traitements et les

traitements peuvent être efficaces mais générer des effets secondaires importants de par leurs effets sur les cellules saines de l'organisme. Les auteurs de [53] associent la résistance aux médicaments aux propriétés de robustesse des cellules cancéreuses et les effets secondaires à la fragilité des cellules saines. Modéliser ces mécanismes de résistance et d'effets secondaires permettraient d'affiner les prédictions de cibles thérapeutiques en favorisant les cibles générant moins d'effets secondaires et/ou moins susceptibles de provoquer des mécanismes de résistance dans les cellules cancéreuses. Un des mécanisme de robustesses concerne l'hétérogénéite des cellules cancéreuses : les cellules constituant une tumeur peuvent contenir des mutations différentes. Une solution serait alors de définir les noyaux de reprogrammation de façon à ce qu'ils constituent une stratégie commune à différents modèles malades. Pour éviter les effets secondaires, les noyaux de reprogrammation devraient garantir que les perturbations proposées maintiennent les états stables du système sain.

Bibliographie

[1] Cancer Gene Census. https://cancer.sanger.ac.uk/census.

[2] COSMIC Database. https://cancer.sanger.ac.uk/cosmic.

[3] KEGG MAPK signalling pathway. https://www.genome.jp/kegg-bin/show_pathway?map04010+M00687.

[4] KEGG p53 signalling pathway. https://www.genome.jp/kegg-bin/show_pathway?ko04115+K04451.

[5] KEGG PI3K/AKT signalling pathway. https://www.genome.jp/kegg-bin/show_pathway?ko04151+K04451.

[6] J. P. Alao. The regulation of cyclin D1 degradation : roles in cancer development and the potential for therapeutic invention. *Molecular Cancer*, 6(1) :1, 2007.

[7] H. R. Andersen. An introduction to binary decision diagrams. *Department of Information Technology, Technical University of Denmark, Lyngby, Denmark, 1997. Lecture notes for 49285 Advanced Algorithms E97, http ://www.itu.dk/people/hra/notes-index.html.*, 1997.

[8] S. Azab and A. Al-Hendy. Signal transduction pathways in breast cancer–drug targets and challenges. In *Breast Cancer-Carcinogenesis, Cell Growth and Signalling Pathways*. InTech, 2011.

[9] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439) :509–512, 1999.

[10] M. Benítez, C. Espinosa-Soto, P. Padilla-Longoria, and E. R. Alvarez-Buylla. Interlinked nonlinear subnetworks underlie the formation of robust cellular patterns in arabidopsis epidermis : a dynamic spatial model. *BMC Systems Biology*, 2(1) :98, 2008.

[11] C. Biane and F. Delaplace. Abduction based drug target discovery using Boolean control network. In *International Conference on Computational Methods in Systems Biology*, pages 57–73. Springer, 2017.

[12] C. Biane and F. Delaplace. Causal reasoning on Boolean control networks based on abduction : theory and application to cancer drug discovery. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2018.

[13] C. Biane, F. Delaplace, and T. Melliti. Abductive network action inference for targeted therapy discovery. *Electronic Notes in Theoretical Computer Science*, 335 :4–26, 2018.

[14] Biomarkers Definitions Working Group, A. J. Atkinson Jr, W. A. Colburn, V. G. DeGruttola, D. L. DeMets, G. J. Downing, D. F. Hoth, J. A. Oates, C. C. Peck, R. T. Schooley, et al. Biomarkers and surrogate endpoints : preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, 69(3) :89–95, 2001.

[15] A. Blake. *Canonical Expressions in Boolean Algebra*. University of Chicago Press, 1938.

[16] R. E. Bryant. Graph-based algorithms for Boolean function manipulation. *IEEE Transactions on Computers*, 100(8) :677–691, 1986.

[17] L. N. Burga, H. Hu, A. Juvekar, N. M. Tung, S. L. Troyan, E. W. Hofstatter, and G. M. Wulf. Loss of BRCA1 leads to an increase in epidermal growth factor receptor expression in mammary epithelial cells, and epidermal growth factor receptor inhibition prevents estrogen receptor-negative cancers in BRCA1-mutant mice. *Breast Cancer Research*, 13(2) :R30, 2011.

[18] L. Calzone, L. Tournier, S. Fourquet, D. Thieffry, B. Zhivotovsky, E. Barillot, and A. Zinovyev. Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLoS Computational Biology*, 6(3) :e1000702, 2010.

[19] Cancer Genome Atlas Network et al. Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418) :61, 2012.

[20] P.-Y. Cheah. The emerging role of RBL2/p130 in multi-step retinoblastoma tumorigenesis. *Cancer Biology & Therapy*, 8(8) :718–719, 2009.

[21] G. Ciriello, E. Cerami, C. Sander, and N. Schultz. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Research*, 22(2) :398–406, 2012.

[22] M. Cohen-Armon, L. Visochek, D. Rozensal, A. Kalal, I. Geistrikh, R. Klein, S. Bendetz-Nezer, Z. Yao, and R. Seger. DNA-Independent PARP-1 Activation by Phosphorylated ERK2 Increases Elk1 Activity : A Link to Histone Acetylation. *Molecular Cell*, 25(2) :297–308, 2007.

[23] S. Collombet, C. van Oevelen, J. L. S. Ortega, W. Abou-Jaoudé, B. Di Stefano, M. Thomas-Chollier, T. Graf, and D. Thieffry. Logical modeling of lymphoid and myeloid cell specification and transdifferentiation. *Proceedings of the National Academy of Sciences of the USA*, 114(23) :5792–5799, 2017.

[24] S. J. Compere, P. Baldacci, A. H. Sharpe, T. Thompson, H. Land, and R. Jaenisch. The ras and myc oncogenes cooperate in tumor induction in many tissues when introduced into midgestation mouse embryos by retroviral vectors. *Proceedings of the National Academy of Sciences of the USA*, 86(7) :2224–2228, 1989.

[25] O. Coudert and J. C. Madre. A unified framework for the formal verification of sequential circuits. In *IEEE International Conference on Computer-Aided Design. Digest of Technical Papers*, pages 126–129, 1990.

[26] Y. Crama and P. L. Hammer. *Boolean Functions : Theory, Algorithms, and Applications*. 2011.

[27] M. Cully, H. You, A. J. Levine, and T. W. Mak. Beyond PTEN mutations : the PI3K pathway as an integrator of multiple inputs during tumorigenesis. *Nature Reviews Cancer*, 6(3) :184–192, 2006.

[28] J. Demongeot, E. Goles, M. Morvan, M. Noual, and S. Sené. Attraction basins as gauges of robustness against boundary conditions in biological complex systems. *PloS One*, 5(8) :e11793, 2010.

[29] C.-X. Deng. BRCA1 : cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution. *Nucleic Acids Research*, 34(5) :1416–1426, 2006.

[30] F. A. Dick. Structure-function analysis of the retinoblastoma tumor suppressor protein–is the whole a sum of its parts? *Cell division*, 2(1) :26, 2007.

[31] G. Didier, E. Remy, and C. Chaouiya. Mapping multivalued onto boolean dynamics. *Journal of theoretical biology*, 270(1) :177–184, 2011.

[32] A. E. Emery, F. Muntoni, and R. C. Quinlivan. *Duchenne muscular dystrophy*. Oxford University Press, 2015.

[33] P. Eroles, A. Bosch, J. Alejandro Pérez-Fidalgo, and A. Lluch. Molecular biology in breast cancer : Intrinsic subtypes and signaling pathways. *Cancer Treatment Reviews*, 38(6) :698–707, 2012.

[34] M. Farago, I. Dominguez, E. Landesman-Bollag, X. Xu, A. Rosner, R. D. Cardiff, and D. C. Seldin. Kinase-inactive glycogen synthase kinase 3β promotes wnt signaling and mammary tumorigenesis. *Cancer Research*, 65(13) :5792–5801, 2005.

[35] H. Farmer, N. McCabe, C. J. Lord, A. N. Tutt, D. A. Johnson, T. B. Richardson, M. Santarosa, K. J. Dillon, I. Hickson, C. Knights, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, 434(7035) :917–921, 2005.

[36] 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) :124–131, 2006.

[37] E. R. Fearon and B. Vogelstein. A genetic model for colorectal tumorigenesis. *Cell*, 61(5) :759–767, 1990.

[38] A. Feber, J. Clark, G. Goodwin, A. R. Dodson, P. H. Smith, A. Fletcher, S. Edwards, P. Flohr, A. Falconer, T. Roe, et al. Amplification and overexpression of E2F3 in human bladder cancer. *Oncogene*, 23(8) :1627, 2004.

[39] S. A. Forbes, D. Beare, H. Boutselakis, S. Bamford, N. Bindal, J. Tate, C. G. Cole, S. Ward, E. Dawson, L. Ponting, et al. COSMIC : somatic cancer genetics at high-resolution. *Nucleic Acids Research*, 45(D1) :D777–D783, 2016.

[40] Y. H. Goh, J. Yoo, J. H. Noh, and C. Kim. Emerging targeted therapies in advanced bladder cancer. *Translational Cancer Research*, 6(4) :S666–S676, 2017.

[41] D. Hanahan and R. A. Weinberg. Hallmarks of cancer : The next generation. *Cell*, 144(5) :646–674, 2011.

[42] T. Helleday. The underlying mechanism for the parp and brca synthetic lethality : clearing up the misunderstandings. *Molecular oncology*, 5(4) :387–393, 2011.

[43] L. Hood, J. R. Heath, M. E. Phelps, and B. Lin. Systems biology and new technologies enable predictive and preventative medicine. *Science*, 306(5696) :640–643, 2004.

[44] C. D. Hurst, D. C. Tomlinson, S. V. Williams, F. M. Platt, and M. A. Knowles. Inactivation of the Rb pathway and overexpression of both isoforms of E2F3 are obligate events in bladder tumours with 6p22 amplification. *Oncogene*, 27(19) :2716, 2008.

[45] N. E. Hynes and H. A. Lane. ERBB receptors and cancer : the complexity of targeted inhibitors. *Nature Reviews Cancer*, 5(5) :341–354, 2005.

[46] International Agency for Research on Cancer et al. World cancer factsheet. *Cancer Research UK, London*, 2014.

[47] W. G. Kaelin. The concept of synthetic lethality in the context of anticancer therapy. *Nature Reviews Cancer*, 5(9) :689–98, 2005.

[48] C. Kandoth, M. D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J. F. McMichael, M. A. Wyczalkowski, et al. Mutational landscape and significance across 12 major cancer types. *Nature*, 502(7471) :333–339, 2013.

[49] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima. KEGG : new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1) :D353–D361, 2017.

[50] S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22(3) :437–467, 1969.

[51] D. S. Kaufman, W. U. Shipley, and A. S. Feldman. Bladder cancer. *The Lancet*, 374(9685) :239–249, 2009.

[52] K. Kehn, R. Berro, A. Alhaj, M. Bottazzi, W. Yeh, Z. Klase, R. Van Duyne, S. Fu, and F. Kashanchi. Functional consequences of cyclin D1/BRCA1 interaction in breast cancer cells. *Oncogene*, 26(35) :5060–5069, 2007.

[53] H. Kitano. A robustness-based approach to systems-oriented drug design. *Nature Reviews Drug Discovery*, 6(3) :nrd2195, 2007.

[54] E. A. Klein and R. K. Assoian. Transcriptional regulation of the cyclin D1 gene at a glance. *Journal of Cell Science*, 121(Pt 23) :3853–3857, 2008.

[55] E. Kumaraswamy, K. L. Wendt, L. A. Augustine, S. R. Stecklein, E. C. Sibala, D. Li, S. Gunewardena, and R. A. Jensen. BRCA1 regulation of epidermal growth factor receptor (EGFR) expression in human breast cancer cells involves microRNA-146a and is critical for its tumor suppressor function. *Oncogene*, 34(33) :4333–4346, 2015.

[56] W. L. Lanier and S. V. Rajkumar. Empiricism and rationalism in medicine : Can 2 competing philosophies coexist to improve the quality of medical care ? *Mayo Clinic Proceedings*, 88(10) :1042–1045, 2013.

[57] R. Layek, A. Datta, M. Bittner, and E. R. Dougherty. Cancer therapy design based on pathway logic. *Bioinformatics*, 27(4) :548–555, 2011.

[58] C. Y. Lee. Representation of switching circuits by binary-decision programs. *The Bell System Technical Journal*, 38(4) :985–999, July 1959.

[59] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences of the USA*, 101(14) :4781–4786, 2004.

[60] P.-C. K. Lin and S. P. Khatri. Application of Max-SAT-based ATPG to optimal cancer therapy design. *BMC Genomics*, 13 Suppl 6(Suppl 6) :S5, 2012.

[61] L. Livraghi and J. E. Garber. PARP inhibitors in the management of breast cancer : current data and future prospects. *BMC Medicine*, 13(1) :1, 2015.

[62] J. Luo. Glycogen synthase kinase 3β (GSK3β) in tumorigenesis and cancer chemotherapy. *Cancer Letters*, 273(2) :194–200, 2009.

[63] M. Malanga, J. M. Pleschke, H. E. Kleczkowska, and F. R. Althaus. Poly (ADP-ribose) binds to specific domains of p53 and alters its DNA binding functions. *Journal of Biological Chemistry*, 273(19) :11839–11843, 1998.

[64] P. Marquis. Extending abduction from propositional to first-order logic. In *Fundamentals of artificial intelligence research*, pages 141–155. Springer, 1991.

[65] L. D. Mayo and D. B. Donner. The PTEN, Mdm2, p53 tumor suppressor-oncoprotein network. *Trends in Biochemical Sciences*, 27(9) :462–467, 2002.

[66] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4) :115–133, 1943.

[67] V. A. McKusick. *Mendelian inheritance in man : a catalog of human genes and genetic disorders*, volume 1. JHU Press, 1998.

[68] S. Meloche and J. Pouysségur. The ERK1/2 mitogen-activated protein kinase pathway as a master regulator of the G1-to S-phase transition. *Oncogene*, 26(22) :3227, 2007.

[69] I. Miko. Genetic dominance : genotype-phenotype relationships. *Nature Education*, 1(1) :140, 2008.

[70] U. M. Moll and O. Petrenko. The MDM2-p53 Interaction. *Molecular Cancer Research*, 1(14) :1001–1008, 2004.

[71] M. K. Morris, J. Saez-Rodriguez, P. K. Sorger, and D. A. Lauffenburger. Logic-based models for the analysis of cell signaling networks. *Biochemistry*, 49(15) :3216–3224, 2010.

[72] P. Mullan, J. Quinn, and D. Harkin. The role of BRCA1 in transcriptional regulation and cell cycle control. *Oncogene*, 25(43) :5854–5863, 2006.

[73] D. Murrugarra, A. Veliz-Cuba, B. Aguilar, and R. Laubenbacher. Identification of control targets in Boolean molecular network models via computational algebra. *BMC Systems Biology*, 10(1) :94, 2016.

[74] Y. Nagata, K.-H. Lan, X. Zhou, M. Tan, F. J. Esteva, A. A. Sahin, K. S. Klos, P. Li, B. P. Monia, N. T. Nguyen, et al. PTEN activation contributes to tumor inhibition by trastuzumab, and loss of PTEN predicts trastuzumab resistance in patients. *Cancer Cell*, 6(2) :117–127, 2004.

[75] S. A. Narod and W. D. Foulkes. BRCA1 and BRCA2 : 1994 and beyond. *Nature Reviews Cancer*, 4(9) :665–676, 2004.

[76] National Research Council. *Toward precision medicine : building a knowledge network for biomedical research and a new taxonomy of disease*. National Academies Press, 2011.

[77] D. Nguyen, M. Zajac-Kaye, L. Rubinstein, D. Voeller, J. E. Tomaszewski, S. Kummar, A. P. Chen, Y. Pommier, J. H. Doroshow, and S. X. Yang. Poly (ADP-ribose) polymerase inhibition enhances p53-dependent and-independent DNA damage responses induced by DNA damaging agent. *Cell Cycle*, 10(23) :4074–4082, 2011.

[78] M. Noual, D. Regnault, and S. Sené. About non-monotony in boolean automata networks. *Theoretical Computer Science*, 504 :12–25, 2013.

[79] M. Oeggerli, S. Tomovska, P. Schraml, D. Calvano-Forte, S. Schafroth, R. Simon, T. Gasser, M. J. Mihatsch, and G. Sauter. E2F3 amplification and overexpression is associated with invasive tumor growth and rapid tumor cell proliferation in urinary bladder cancer. *Oncogene*, 23(33) :5616, 2004.

[80] B. Pereira, S.-F. Chin, O. M. Rueda, H.-K. M. Vollan, E. Provenzano, H. A. Bardwell, M. Pugh, L. Jones, R. Russell, S.-J. Sammut, et al. The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nature Communications*, 7 :11479, 2016.

[81] L. Perfetto, L. Briganti, A. Calderone, A. C. Perpetuini, M. Iannuccelli, F. Langone, L. Licata, M. Marinkovic, A. Mattioni, T. Pavlidou, D. Peluso, L. L. Petrilli, S. Pirro, D. Posca, E. Santonico, A. Silvestri, F. Spada, L. Castagnoli, and G. Cesareni. SIGNOR : A database of causal relationships between biological entities. *Nucleic Acids Research*, 44(D1) :D548–D554, 2016.

[82] P. C. Phillips. Epistasis - the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, 9(11) :855–867, 2008.

[83] W. Quine. On cores and prime implicants of truth functions. *The American Mathematical Monthly*, 66(9) :755–760, 1959.

[84] E. Remy, S. Rebouissou, C. Chaouiya, A. Zinovyev, F. Radvanyi, and L. Calzone. A modelling approach to explain mutually exclusive and co-occurring genetic alterations in bladder tumorigenesis. *Cancer Research*, pages canres–0602, 2015.

[85] F. Robert. *Discrete iterations : a metric study*. Springer-Verlag, 1986.

[86] C. Romagosa, S. Simonetti, L. Lopez-Vicente, A. Mazo, M. Lleonart, J. Castellvi, and S. R. y Cajal. p16 Ink4a overexpression in cancer : a tumor suppressor gene associated with senescence and high-grade tumors. *Oncogene*, 30(18) :2087, 2011.

[87] R. W. Ruddon. *Cancer Biology*. Oxford University Press, 2007.

[88] J. Saez-Rodriguez, L. Simeoni, J. A. Lindquist, R. Hemenway, U. Bommhardt, B. Arndt, U.-U. Haus, R. Weismantel, E. D. Gilles, S. Klamt, et al. A logical mo-

del provides insights into T cell receptor signaling. *PLoS Computational Biology*, 3(8) :e163, 2007.

[89] E. Samson and B. Mills. *Circuit Minimization : Algebra and Algorithms for New Boolean Canonical Expressions.* Air Force Cambridge Research Center Technical report. U.S. Department of Commerce, Office of Technical Services, 1954.

[90] C. Sanchez, C. Lachaize, F. Janody, B. Bellon, L. Röder, J. Euzenat, F. Rechenmann, and B. Jacq. Grasping at molecular interactions and genetic networks in Drosophila melanogaster using FlyNets, an Internet database. *Nucleic Acids Research*, 27(1) :89–94, 1999.

[91] T. Santarius, J. Shipley, D. Brewer, M. R. Stratton, and C. S. Cooper. A census of amplified and overexpressed human cancer genes. *Nature Reviews Cancer*, 10(1) :59, 2010.

[92] L. Shan, X. Li, L. Liu, X. Ding, Q. Wang, Y. Zheng, Y. Duan, C. Xuan, Y. Wang, F. Yang, et al. GATA3 cooperates with PARP1 to regulate CCND1 transcription through modulating histone H1 incorporation. *Oncogene*, 33(24) :3205, 2014.

[93] M. Spiliotaki, D. Mavroudis, K. Kapranou, H. Markomanolaki, G. Kallergi, F. Koinis, K. Kalbakis, V. Georgoulias, and S. Agelaki. Evaluation of proliferation and apoptosis markers in circulating tumor cells of women with early breast cancer who are candidates for tumor dormancy. *Breast Cancer Research*, 16(6) :485, 2014.

[94] V. Stambolic, D. MacPherson, D. Sas, Y. Lin, B. Snow, Y. Jang, S. Benchimol, and T. W. Mak. Regulation of PTEN transcription by p53. *Molecular Cell*, 8(2) :317–325, 2001.

[95] M. Strnisková, M. Barančík, and T. Ravingerová. MItogen-activated protein kinases and their role in regulation of cellular processes. *General Physiology and Biophysics*, 21(3) :231–255, 2002.

[96] D. C. Swinney and J. Anthony. How were new medicines discovered? *Nature Reviews Drug Discovery*, 10(7) :507–519, 2011.

[97] R. Thomas. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, 42(3) :563–585, 1973.

[98] M. Toshiyuki and J. C. Reed. Tumor suppressor p53 is a direct transcriptional activator of the human bax gene. *Cell*, 80(2) :293 – 299, 1995.

[99] F. Tsuruta, N. Masuyama, and Y. Gotoh. The phosphatidylinositol 3-kinase (PI3K)-Akt pathway suppresses Bax translocation to mitochondria. *Journal of Biological Chemistry*, 277(16) :14040–14047, 2002.

[100] B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, and K. W. Kinzler. Cancer genome landscapes. *Science*, 339(6127) :1546–1558, 2013.

[101] R.-S. Wang, A. Saadatpour, and R. Albert. Boolean modeling in systems biology : an overview of methodology and applications. *Physical Biology*, 9(5) :055001, 2012.

[102] X. Wang, A. Q. Fu, M. E. McNerney, and K. P. White. Widespread genetic epistasis among cancer genes. *Nature Communications*, 5, 2014.

[103] W. M. Webb. Rationalism, Empiricism, and Evidence-Based Medicine : A Call for a New Galenic Synthesis. *Medicines*, 5(2) :40, 2018.

[104] P. L. Welcsh, M. K. Lee, R. M. Gonzalez-Hernandez, D. J. Black, M. Mahadevappa, E. M. Swisher, J. A. Warrington, and M.-C. King. BRCA1 transcriptionally regulates genes involved in breast tumorigenesis. *Proceedings of the National Academy of Sciences of the USA*, 99(11) :7560–7565, 2002.

[105] A. K. Witkiewicz, K. E. Knudsen, A. P. Dicker, and E. S. Knudsen. The meaning of p16ink4a expression in tumors : functional significance, clinical associations and future developments. *Cell Cycle*, 10(15) :2497–2503, 2011.

[106] A. R. Yallowitz, D. Li, A. Lobko, D. Mott, A. Nemajerova, and N. Marchenko. Mutant p53 amplifies epidermal growth factor receptor family signaling to promote mammary tumorigenesis. *Molecular Cancer Research*, pages molcanres–0360, 2015.

[107] R. Yao, W. J. Lemon, Y. Wang, C. J. Grubbs, R. A. Lubet, and M. You. Altered gene expression profile in mouse bladder cancers induced by hydroxybutyl (butyl) nitrosamine. *Neoplasia*, 6(5) :569–577, 2004.

[108] C.-H. Yeang, F. McCormick, and A. Levine. Combinatorial patterns of somatic gene mutations in cancer. *The FASEB Journal*, 22(8) :2605–2622, 2008.

[109] J. G. Zanudo and R. Albert. Cell fate reprogramming by control of intracellular network dynamics. *PLoS Computational Biology*, 11(4) :e1004193, 2015.

[110] H. Zhang, K. Somasundaram, Y. Peng, H. Tian, H. Zhang, D. Bi, B. L. Weber, and W. S. El-Deiry. BRCA1 physically associates with p53 and stimulates its transcriptional activity. *Oncogene*, 16(13) :1713–1721, 1998.

[111] R. Zhang, M. V. Shah, J. Yang, S. B. Nyland, X. Liu, J. K. Yun, R. Albert, and T. P. Loughran. Network model of survival signaling in large granular lymphocyte leukemia. *Proceedings of the National Academy of Sciences of the USA*, 2008.

[112] W. Zhang and H. T. Liu. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Research*, 12(1) :9–18, 2002.

[113] Q. Zhong, N. Simonis, Q.-R. Li, B. Charloteaux, F. Heuze, N. Klitgord, S. Tam, H. Yu, K. Venkatesan, D. Mou, V. Swearingen, M. a. Yildirim, H. Yan, A. Dricot, D. Szeto, C. Lin, T. Hao, C. Fan, S. Milstein, D. Dupuy, R. Brasseur, D. E. Hill, M. E. Cusick, and M. Vidal. Edgetic perturbation models of human inherited disorders. *Molecular Systems Biology*, 5(321) :321, 2009.

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Reprogrammation comportementale : modèles, algorithmes et application aux maladies complexes

Mots clés : Réseaux booléens contrôlés ; Abduction ; Reprogrammation comportementale ; Maladies complexes

Résumé : Les maladies complexes comme le Cancer et la maladie d'Alzheimer sont causées par des perturbations moléculaires multiples responsables d'un comportement cellulaire pathologique. Un enjeu majeur de la médecine de précision est l'identification des perturbations moléculaires induites par les maladies complexes et les thérapies à partir de leurs conséquences sur les phénotypes cellulaires.
Nous définissons un modèle des maladies complexes, appelé la reprogrammation comportementale, assimilant les perturbations moléculaires à des altérations des fonctions dynamiques locales de systèmes dynamiques discrets induisant une reprogrammation de la dynamique globale du réseau. Ce cadre de modélisation s'appuie d'une part, sur les réseaux booléens contrôlés, qui sont des réseaux booléens dans lesquels sont insérés des paramètres de contrôle modélisant les perturbations et, d'autre part, sur la définition de modes (Possibilité, Nécessité) permettant d'exprimer les objectifs de cette reprogrammation.
À partir de ce cadre, nous démontrons que le calcul des noyaux, c'est-à-dire, des ensembles minimaux d'actions permettant la reprogrammation selon un mode s'exprime comme un problème d'inférence abductive en logique propositionnelle. En nous appuyant sur les méthodes historiques de calcul d'impliquants premiers des fonctions booléennes, nous développons deux méthodes permettant le calcul exhaustif des noyaux de la reprogrammation.
Enfin, nous évaluons la pertinence du cadre de modélisation pour l'identification des perturbations responsables de la transformation d'une cellule saine en cellule cancéreuse et la découverte de cibles thérapeutiques sur un modèle du cancer du sein. Nous montrons notamment que les perturbations inférées par nos méthodes sont compatibles avec la connaissance biologique en discriminant les oncogènes des gènes suppresseurs de tumeurs et en récupérant la mutation du gène BRCA1. De plus, la méthode récupère le phénomène de létalité synthétique entre PARP1 et BRCA1, qui constitue un traitement anticancéreux optimal car il cible spécifiquement les cellules tumorales.

Behavioral reprogramming : models, algorithms and application to complex diseases

Keywords : Boolean Control Networks ; Abduction ; Behavioral Reprogramming ; Complex Diseases

Abstract : Complex diseases such as cancer and Alzheimer's are caused by multiple molecular perturbations responsible for pathological cellular behavior. A major challenge of precision medicine is the identification of the molecular perturbations induced by the diseases and the therapies from their consequences on cell phenotypes.
We define a model of complex diseases, called behavioral reprogramming, that assimilates the molecular perturbations to alterations of the dynamic local functions of discrete dynamical systems inducing a reprogramming of the global dynamics of the network. This modeling framework relies on the one hand, on Control Boolean networks, which are Boolean networks containing control parameters modeling the perturbations and, on the other hand, the definition of reprogramming modes (Possibility, Necessity) expressing the objective of the behavioral reprogramming.
From this framework, we prove that the computation of the cores, namely, the minimal sets of action enabling reprogramming is a problem of abductive inference in propositional logic. Using historical methods computing the prime implicants of Boolean functions, we develop two methods computing all the reprogramming cores.
Finally, we evaluate the modeling framework for the identification of perturbations responsible for the transformation of a healthy cell into a cancer cell and the discovery of therapeutic targets on a model of breast cancer. In particular, we show that the perturbations inferred by our methods are compatible with biological knowledge by discriminating oncogenes and tumor suppressor genes and by recovering the causal of the BRCA1 gene. In addition, the method recovers the synthetic lethality phenomenon between PARP1 and BRCA1 that constitutes an optimal anti-cancer treatment because it specifically targets tumor cells.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France