# Contents

Analyses des motifs accessibles dans les modèles KappaJérôme Feret
[1]Département d'informatique de l'ÉNS, ÉNS, CNRS, Université PSL, Paris, France
[2]INRIA, Paris, France

# Chapter 1

# Introduction

## 1.1 Contexte et motivations

Décrire et analyser les systèmes à grande échelle et fortement combinatoires qui sont issus de certains modèles mécanistiques de biologie des systèmes est encore hors de portée de l'état de l'art. Dans de tels modèles, le comportement individuel des occurrences de protéines, qui peuvent établir des liaisons et modifier leur capacité d'interaction, est influencé par des compétitions pour des ressources communes. De plus, les occurrences de protéines peuvent former une grande diversité de complexes biochimiques différents. La concurrence entre des interactions à différentes échelles de temps génère des boucles de rétro-actions non linéaires qui contrôlent l'abondance de ces complexes biochimiques. Enfin, ces systèmes font intervenir des interactions entre de très petites molécules, comme des ions ou des ligands et des complexes biochimiques gigantesques comme les brins d'acide désoxyribonucléique, le ribosome, ou le signalosome. Comprendre comment le comportement collectif des populations de protéines qui définit le phénotype, est engendré par le comportement individuel des occurrences de ces protéines reste un problème largement ouvert et un enjeu crucial.

Alors que les progrès technologiques permettent d'obtenir rapidement une quantité toujours plus importante de détails à propos des interactions mécanistiques potentielles entre les occurrences de protéines, et ce, à un prix très accessible, la communauté scientifique est encore bien loin de comprendre globalement comment le comportement macroscopique des systèmes dans leur ensemble émerge de ces interactions. C'est l'objectif annoncé de la biologie des systèmes. Mais ce but est sans espoir à moins que des méthodes spécifiques et innovantes pour décrire ces systèmes complexes et analyser leur propriété ne soient conçues. Bien entendu, ces méthodes devront passer à l'échelle de la très grande quantité d'informations qui est publiée dans la littérature à un rythme qui augmente de manière exponentielle.

## 1.2 Les langages de modélisation de systèmes d'interactions moléculaires

Les langages formels ont été beaucoup utilisés pour décrire des modèles d'interactions mécanistiques entre occurrences de protéines. Ils procurent des outils mathématiques pour traduire ces interactions et définir rigoureusement le comportement des systèmes ainsi représentés grâce à un choix de sémantiques qualitatives, stochastiques ou différentielles.

Les langages tels que les réseaux réactionnels [58] ou les réseaux de Petri classiques [82], se basent sur le paradigme de la réécriture multi-ensemble. Les interactions consistent à consommer des réactifs en échange de produits. Des constantes cinétiques permettent de préciser soit la vitesse, soit la fréquence moyenne – selon le choix de la sémantique – d'application des différentes réactions. Ceci les rend très utiles pour décrire et formaliser le comportement de systèmes d'interactions de petite ou moyenne taille. Cependant, ces langages peinent à représenter de grands modèles car ils ont besoin d'un nom (ou d'un emplacement dans le cas des réseaux de Petri) par type de complexes biomoléculaires.

Des langages de plus haut niveau, inspirés des différents paradigmes de programmation, tels que les tableaux d'états à messages [43], les automates communicants [101], les algèbres de processus [105, 31], les langages orientés objet [55], les réseaux de Petri colorés [70] et la réécriture de graphes à sites [52, 56, 5, 86], exploitent le fait que les interactions dépendent généralement de conditions locales sur les configurations des occurrences

de protéines au sein des complexes biochimiques. Ces langages permettent ainsi de traduire les systèmes d'interactions entre les occurrences de protéines de manière plus parcimonieuse : seuls les détails qui importent pour une interaction donnée sont mentionnés pour décrire cette interaction.

Il est important de distinguer les approches basées sur les agents de celles basées sur les règles de réécriture. Dans les approches basées sur les agents, chaque entité, que ce soit un processus [31] ou un objet [55], doit contenir la description de tous ses comportements possibles. Les changements entre les configurations des différentes entités se synchronisent par le biais de règles de communication. Ces règles, généralement en très petit nombre, définissent la sémantique opérationnelle des langages. Il est possible de conditionner le comportement d'un agent à des propriétés de l'état d'un autre agent auquel cet agent serait lié, mais cela nécessite de recourir à des processus fictifs pour aller chercher cette information. Cette astuce était en fait déjà utilisée dans les premiers modèles décrits en $\pi$-calcul [105]. Cependant, en général, les approches basées sur les agents donnent lieu à des systèmes de processus à états finis [89]. Ceci permet d'étudier leur comportement à l'aide d'outils de vérification symbolique de modèles comme PRISM [93].

Lorsque les occurrences des protéines admettent trop de configurations différentes ou lorsque leurs capacités d'interaction dépendent trop des occurrences des protéines auxquelles elles sont liées, les approches fondées sur les agents ne passent pas à l'échelle, tant au niveau de la description des modèles que pour le calcul de leurs propriétés.

Dans les approches fondées sur les règles, les modèles sont définis par des règles d'interaction. Chaque règle définit sous quelles conditions sur les configurations des agents une interaction peut avoir lieu et quels sont les effets de cette interaction. Ainsi l'état des agents ne définit pas une fois pour toute les capacités d'interaction de cet agent. Ce sont les règles du modèle qui le font. Il n'est pas non plus nécessaire de donner la liste exhaustive de toutes les configurations des agents. Les règles peuvent se contenter de ne mentionner que les parties importantes des agents pour l'interaction qu'elles décrivent. Les approches fondées sur les règles passent mieux à l'échelle et facilitent la mise à jour des modèles. De plus, comme il n'est pas nécessaire de spécifier explicitement toutes les capacités d'interaction des occurrences des protéines, elles encouragent à une modélisation sans *a priori* où les interactions émergent des règles lors de la conception du modèle.

Le calcul des ambients [27, 28], des bioambients [104] et celui des membranes [26] sont un peu particuliers. Ils permettent de décrire des boîtes ou des compartiments, qui peuvent être arbitrairement imbriqués au sein d'une arborescence, alors que des agents, contenus dans les boites dans le cas des ambients, ou dans leurs parois dans le cas des membranes, permettent à ces compartiments de se déplacer ou de se fusionner. Les capacités d'interaction des agents peuvent alors dépendre de leur localisation dans la hiérarchie des compartiments. La calcul projectif des membranes [53] représente plus fidèlement la disposition des compartiments au sein d'une cellule, en rendant la description de l'état du système indépendante du choix de la racine de l'arborescence des compartiments.

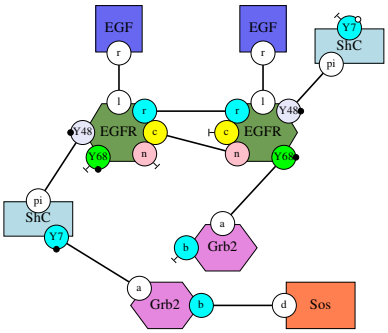
## 1.3 Le langage Kappa

Les langages de réécriture de graphes à sites [52, 56, 5, 86] permettent de représenter de manière transparente les réseaux d'interactions entre des occurrences de protéines grâce à leur syntaxe qui est fortement inspirée de la chimie.

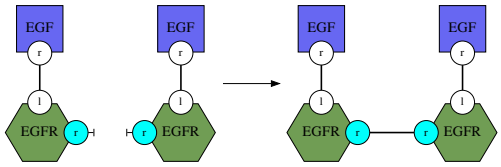
Dans Kappa, chaque complexe biochimique est représenté par un graphe à sites. Un exemple de graphe à sites est donné en Fig. 1.1(a). Dans un graphe à sites, des nœuds qui représentent des occurrences de protéines, sont associés à une liste de sites d'interaction. Ces sites peuvent être libres ou liés deux à deux. En outre, certains sites portent une propriété, qui peut servir à représenter un niveau d'activation. Les interactions entre occurrences de protéines peuvent modifier leurs conformations en dépliant ou en repliant leurs chaînes de nucléotides, ce qui peut révéler ou cacher des sites d'interaction. Dans Kappa, la structure tri-dimensionnelle des occurrences de protéines n'est pas représentée explicitement. En revanche, les conditions pour qu'un site d'interaction soit visible sont spécifiées dans la description des interactions elles-mêmes.

L'évolution d'un système Kappa se décrit grâce à des règles de réécriture hors-contexte. En Fig. 1.1(b) est dessinée une règle pour la formation de dimers. Deux récepteurs ($EGFR$) qui sont tous deux liés à des ligands ($EGF$) peuvent se lier entre eux pour former un dimer. En Fig. 1.1(c) est donnée une autre règle issue d'un modèle de réparation de l'ADN, dans laquelle une enzyme, la Glycolase ($DG$), peut glisser aléatoirement dans les deux sens, le long d'un brin d'ADN [91].

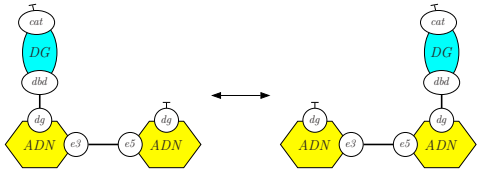
Une règle peut être comprise de manière intentionnelle comme une transformation locale de l'état du système ou de manière extensionnelle comme l'ensemble, qu'il soit fini ou non, des réactions biochimiques qui peuvent




(a) Un graphe à sites.




(b) Une règle pour lier des occurrences de protéines.




(c) Une règle pour faire glisser des occurrences de protéines.

Figure 1.1: En 1.1(a) est dessiné un graphe à site. Il s'agit d'un complexe biochimique composé de deux occurrences du ligand (*EGF*), de deux occurrences du récepteur membranaire (*EGFR*), d'une occurrence de la protéine d'échafaudage (*Shc*), de deux occurrences de la protéine de transport (*Grb2*) et d'une occurrence de la protéine *Sos*. En 1.1(b) est donné un exemple de règle de liaison. Deux occurrences du récepteurs membranaires (*EGFR*), lorsqu'ils sont tous deux activés par une liaison avec des occurrences du ligand (*EGF*), peuvent se lier. Les autres sites sont omis car ils ne jouent aucun rôle dans cette interaction. En 1.1(c) est donnée une règle de déplacement. Une occurrence de l'enzyme Glycolase (*DG*) peut glisser dans les deux directions (selon une marche aléatoire) le long d'un brin d'ADN.

être obtenues en spécifiant entièrement les différents contextes d'application de ces règles. De cet ensemble de réactions, diverses sémantiques peuvent être définies pour décrire le comportement des systèmes. Ces sémantiques peuvent être qualitatives, stochastiques ou différentielles comme pour le cas des réseaux réactionnels et des réseaux de Pétri (les sémantiques quantitatives — stochastiques ou différentielles — nécessitent d'associer une constante de temps à chaque règle). Il est toutefois possible de simuler un modèle Kappa directement, sans passer par le réseau réactionnel sous-jacent. La simulation consiste alors à itérer la boucle événementielle suivante (celle-ci correspond à l'algorithme de Gillespie [73]). Étant donné l'état du système, représenté par un graphe à sites, l'ensemble de tous les événements possibles est calculé. Un événement consiste à appliquer une règle dans le graphe à une occurrence du motif qui constitue le membre gauche de cette règle. Chaque événement a une propension qui correspond à la constante de la règle correspondante. Le prochain événement est tiré au hasard selon une probabilité proportionnelle à sa propension, alors que le délai entre deux événements est tiré aléatoirement selon une loi exponentielle dont le paramètre est la somme des propensions de tous les événements potentiels du système. Il n'est pas raisonnable de recalculer la liste des événements potentiels à chaque fois après l'application d'une règle. Cet ensemble peut être mis à jour dynamiquement en tenant compte uniquement des nouveaux événements potentiels et des événements qui ne sont plus possibles du fait de l'application du dernier événement choisi [48]. Le simulateur actuel tire profit au maximum des sous-motifs communs dans les motifs qui apparaissent dans le membre gauche des règles pour découvrir les nouveaux événements et retirer les événements devenus obsolètes plus rapidement [18].

Le langage Kappa souffre de plusieurs limites. Par exemple, dans Kappa, les sites d'interaction d'une même occurrence d'une protéine doivent porter des noms différents ; par ailleurs, en ce qui concerne les propriétés géométriques, Kappa ne permet ni de représenter la structure tridimensionnelle des occurrences de protéines, ni leur répartition dans l'espace. Avoir des sites deux à deux différents dans chaque occurrence de protéines facilite grandement la recherche des occurrences des motifs dans les graphes, ce qui est non seulement crucial pour simuler les modèles de manière efficace, mais est aussi à la base de plusieurs constructions utilisées pour l'analyse statique et la réduction de modèles. Certains langages lèvent cette contrainte soit directement comme dans les langages BNGL [56] et mød [4], soit indirectement en utilisant un codage sous forme d'hyperliens comme c'est possible dans le langage React(C) [86]. Toutefois, l'efficacité des moteurs de simulation est fortement réduite quand de telles constructions sont utilisées. Pour ce qui est de la géométrie des protéines, les conditions liées aux conformations spatiales des protéines peuvent être encodées dans les règles de réécriture. Certaines extensions du langage permettent de représenter des contraintes sur la position relative des occurrences de protéines et des sites d'interaction dans les complexes biochimiques afin de restreindre l'ensemble des événements possibles à ceux

qui satisfont ces contraintes [44]. Enfin, dans Kappa, la distribution des occurrences de protéines dans l'espace est passée sous silence. Il est fait l'hypothèse que les occurrences de protéines sont parfaitement mélangées. Il est donc impossible de retrouver les phénomènes d'encombrement qui peuvent être dus à des accumulations d'occurrences de protéines dans certaines régions de la cellule. De même, les gradients de concentration locaux qui pourrait être dus à la présence d'une occurrence d'une protéine d'échafaudage ne peuvent pas être représentés (en Kappa, chaque occurrence d'une protéine d'échafaudage n'agit qu'en maintenant des occurrences de protéines dans le même complexe biochimique, une fois libérée, ces occurrences de protéines ne sont pas supposées rester, même pour un court instant dans le même voisinage). Une solution partielle consiste à encoder en Kappa une grille pour représenter de manière discrète les positions potentielles des occurrences de protéines. Ensuite, celles-ci peuvent glisser le long de cette grille grâce à des règles implémentant la diffusion des occurrences de protéines. Le langage SpatialKappa [108] permet d'utiliser ce procédé de manière transparente. Par ailleurs, le langage ML [83] permet de représenter des modèles d'interactions entre occurrences de protéines qui peuvent se déplacer de manière continue dans un milieu. Il est possible de munir un modèle Kappa d'un ensemble de compartiments statiques. Toutefois, ceci ne permet pas de modéliser le transport d'occurrences de protéines par le biais de vésicules. La machine formelle cellulaire [42] répond à cet enjeu, sans toutefois fournir de moteurs de simulation efficaces.

Les langages de réécriture de graphes à sites permettent de représenter les réseaux d'interactions entre occurrences de protéines, et ce, malgré leur forte combinatoire. Si le comportement de ces réseaux peut être formellement défini et simulé, des abstractions sont toutefois nécessaires pour calculer les propriétés du comportement collectif des populations de protéines.

## 1.4 Interprétation abstraite

L'interprétation abstraite a été introduite il y a maintenant un peu plus de quarante ans comme un cadre mathématique pour établir des liens formels entre le comportement de programmes, vu à différents niveaux d'abstraction. Depuis, l'interprétation abstraite a été utilisée non seulement pour comparer différentes méthodes et algorithmes d'analyse statique [39], mais aussi pour développer des analyseurs statiques qui peuvent calculer automatiquement les propriétés sur le comportement des programmes [8, 57]. L'interprétation abstraite s'est désormais développée dans l'industrie (entre autres, Amazon, Facebook, IBM, Google, MicroSoft et MathWorks ont chacune leurs propres analyseurs statiques basés sur l'interprétation abstraite).

L'interprétation abstraite repose sur la démarche suivante. Le comportement d'un programme (ou d'un modèle) peut en général être décrit comme le plus petit point fixe $lfp\,\mathbb{F}$ d'un opérateur $\mathbb{F}$ agissant sur les éléments d'un ensemble appelé le domaine concret $D$. Le domaine concret est habituellement l'ensemble des parties $\wp(S)$ d'un ensemble d'éléments $S$, qui peuvent être des états, des traces de calcul, *et cetera*. Une abstraction est alors vue comme un changement de granularité dans la description du comportement des programmes (ou des modèles) et ce changement de granularité peut être représenté en langage mathématique sous diverses formes telles qu'un opérateur de clôture supérieure, une famille d'idéaux, une famille de Moore ou une correspondance de Galois. Les correspondances de Galois se sont vite imposées comme l'outil le plus populaire pour décrire une interprétation abstraite. Un changement du niveau d'observation du comportement d'un programme (ou d'un modèle) peut ainsi être décrit en choisissant un ensemble $D^\sharp$ de propriétés d'intérêt. C'est le domaine abstrait. Cet ensemble est ordonné par un ordre partiel $\sqsubseteq$. Chaque élément $a^\sharp$ de ce domaine abstrait représente intentionnellement l'ensemble des éléments concrets qui satisfont cette propriété. Cet ensemble est noté $\gamma(a^\sharp)$. La fonction $\gamma$, ainsi définie, est croissante (si $a^\sharp \sqsubseteq b^\sharp$, alors $\gamma(a^\sharp) \subseteq \gamma(b^\sharp)$). Ainsi, l'ordre $\sqsubseteq$ représente le niveau d'information.

Un élément abstrait $a^\sharp$ est dit être une abstraction d'un ensemble $a$ d'éléments concrets, si et seulement si $a$ est un sous-ensemble de l'ensemble $\gamma(a^\sharp)$. Une correspondance de Galois est obtenue quand chaque sous-ensemble $a$ de l'ensemble $S$ admet une meilleure abstraction, c'est à dire, que pour chaque partie $a$ de l'ensemble $S$, il existe un élément abstrait, noté $\alpha(a)$ qui est d'une part une abstraction de l'ensemble $a$ et d'autre part, qui est plus petit (pour l'ordre $\sqsubseteq$) que n'importe quelle abstraction de l'ensemble $a$. Dans un tel cas, n'importe quelle fonction croissante $\mathbb{F}^\sharp$ opérant sur le domaine abstrait $D^\sharp$ et telle que $[\alpha \circ \mathbb{F} \circ \gamma](a^\sharp) \sqsubseteq \mathbb{F}^\sharp(a^\sharp)$ pour chaque élément abstrait $a^\sharp \in D^\sharp$, admet un plus petit point fixe (pour l'ordre $\sqsubseteq$) noté $lfp\,\mathbb{F}^\sharp$. De plus, la concrétisation de ce plus petit point fixe est un sur-ensemble du plus petit point fixe de la fonction $\mathbb{F}$ ; ainsi le comportement du programme ou du modèle peut être calculé dans le domaine abstrait au prix d'une perte potentielle d'information puisque le résultat final est un sur-ensemble de l'ensemble de tous les comportements possibles. Par construction, l'approche est correcte : aucun comportement de la sémantique concrète n'est

oublié. Par contre, quand le sur-ensemble ainsi calculé est un sur-ensemble strict, des comportements fictifs ont été introduits par l'analyse.

Le choix du domaine abstrait est crucial. Du point de vue de l'expressivité, le domaine abstrait doit permettre de décrire les propriétés d'intérêt des programmes (ou des modèles) ainsi que les propriétés intermédiaires qui sont nécessaires pour en établir la preuve de manière inductive. D'un point de vue algorithmique, ils doivent correspondre à des propriétés qui sont relativement simples à manipuler en machine. Enfin, la structure des chaînes croissantes d'éléments abstraits (pour l'ordre $\sqsubseteq$) est également importante pour que puissent être définis des opérateurs d'extrapolation précis, dans le cas où le domaine admettrait des chaînes croissantes infinies.

Plusieurs interprétations abstraites ont été proposées pour calculer automatiquement les propriétés des modèles en biologie des systèmes. Les premières ont naturellement été inspirées par les analyses de flot d'information [14, 60] et de dénombrement [98, 61] dans le $\pi$-calcul et le calcul des ambients. Ces analyses permettent de détecter avec précision dans quels compartiments des entités peuvent entrer dans des modèles-jouet de virus infectant des cellules. Elles trouvent également des exclusions mutuelles [74, 13]. Les analyses de dénombrement permettent aussi souvent de retrouver les invariants correspondant à la conservation du nombre de chaque sorte de protéines dans les réseaux réactionnels lorsque la composition des complexes biochimiques n'est pas représentée explicitement [1, 2]. Ces invariants sont aussi appelés invariants de places dans les réseaux de Petri.

Les modèles biologiques sont fortement concurrents et souffrent de l'explosion combinatoire dans le nombre d'entrelacements potentiels des différents événements possibles. L'interprétation abstraite a été utilisée pour oublier la séquentialité dans les traces d'exécution dans les processus de frappes [99], puis plus généralement pour les réseaux asynchrones discrets booléens ou multivalués [68]. Dans les modèles réseaux booléens ou multivalués, l'interprétation abstraite a également été utilisée pour calculer une approximation des ensembles constituant des trappes [35, 90], dans lesquels les systèmes ne peuvent plus sortir une fois entrés. Ces ensembles facilitent le calcul des trajectoires périodiques des modèles. Dans les modèles de réseaux métaboliques, l'interprétation abstraite a été utilisée pour décrire une analyse de dépendances, qui calcule l'impact potentiel de l'inhibition éventuelle d'une règle sur la concentration à l'équilibre des composants du système [88, 3].

L'interprétation abstraite peut servir à la calibration d'un modèle [92], en réalisant une partition de l'espace des paramètres en trois régions : une première région dans laquelle le modèle satisfait une propriété temporelle donnée par l'utilisateur, une seconde qui ne la satisfait pas et une troisième pour laquelle l'analyse n'a pu conclure si la propriété était satisfaite ou non.

L'interprétation abstraite est également très utilisée pour le calcul des trajectoires des systèmes hybrides [75].

## 1.5 L'écosystème Kappa

Plusieurs outils pour analyser et manipuler des modèles Kappa sont présentés ici.

### 1.5.1 Analyses statiques

Un outil d'analyse statique [17], basé sur le cadre de l'interprétation abstraite, permet de calculer automatiquement certaines propriétés des modèles. Le but est d'améliorer la confiance dans les règles qui constituent le modèle. Il s'agit de retrouver des propriétés d'intérêt que le modélisateur pouvait, ou non, avoir en tête lors de la conception de son modèle ou bien de trouver des erreurs dans la modélisation.

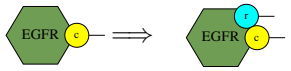
Cette analyse utilise un ensemble de motifs d'intérêt. Parmi ces motifs, l'analyse prouve que certains ne peuvent apparaître dans aucuns états potentiels du modèle. Les autres sont déclarés potentiellement accessibles : soit ils le sont effectivement, soit c'est une conséquence de la sur-approximation de l'analyse.

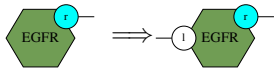
Les motifs d'intérêt permettent de poser les questions intéressantes sur la structure biochimique des occurrences des complexes lors de l'exécution du modèle. Actuellement, l'analyse pose trois types de questions : Existe-t-il une relation entre l'état de plusieurs sites dans les occurrences d'une protéine ? Lorsque deux occurrences de protéines sont liées entre-elles, existe-t-il une relation entre l'état de leurs sites respectifs ? Est-ce qu'une occurrence d'une protéine peut être doublement liée à une autre occurrence d'une protéine, est-ce qu'une occurrence de protéines peut être liée à des occurrences différentes d'un même type de protéines ? La première catégorie est une analyse relationnelle classique. Elle permet, par exemple, de détecter si un site ne peut être lié sans qu'un autre ne le soit ou de détecter si un site ne peut être lié sans être phosphorylé. La seconde est utile quand des sites fictifs permettent d'encoder la localisation des occurrences de protéines, il est alors possible de vérifier, chaque fois que deux occurrences de protéines sont liées, si elles se situent nécessairement dans un même

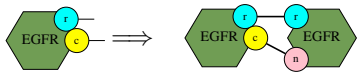
compartiment. Enfin, la troisième analyse la formation de doubles liaisons entre les occurrences de protéines. Le choix exact des questions posées par l'analyseur est fixé automatiquement suite à une inspection statique des règles du modèle.

Le résultat final de l'analyse d'accessibilité est présenté à l'utilisateur sous deux formes. D'une part, les règles dont le membre gauche est en contradiction avec les motifs qui ont été prouvés inaccessibles par l'analyse sont mentionnées à l'utilisateur. D'autre part, les propriétés intéressantes sur la structure des complexes biochimiques sont listées sous la forme de lemmes de raffinement.

Par exemple, les trois lemmes suivants :

- 

- 

- 


informent l'utilisateur que (pour le premier) dans une occurrence du récepteur membranaire, le site $c$ ne peut être lié sans que le site $r$ ne le soit également, que (pour le second) le site $r$ ne peut être lié sans que le site $l$ ne le soit aussi, et que (pour le troisième) quand une occurrence du récepteur membranaire a ses sites $r$ et $c$ tous deux liés, ils sont nécessairement liés tous deux à une même occurrence du récepteur membranaire.

Un lemme de raffinement est ainsi présenté comme une implication entre un motif et une liste de motifs. Ici, les listes de motifs sont toutes réduites à un élément. Il faut interpréter une telle implication par le fait que toute occurrence du membre gauche de l'implication dans un état accessible peut se raffiner dans au moins un des motifs du membre droit.

Lorsque l'utilisateur obtient des propriétés auxquelles il ne s'attend pas, il doit retourner à son modèle pour comprendre l'origine du problème. Les erreurs typographiques sont assez courantes. Il arrive aussi souvent que certaines parties du modèle manquent, il faut aller les compléter ou les remplacer par des règles fictives si l'information n'est pas disponible dans la littérature. Il se peut aussi que l'état initial du modèle ait été mal choisi. Enfin, les erreurs peuvent aussi être dues à des relations causales complexes. L'analyse statique peut alors être complétée par l'analyse causale pour comprendre comment les configurations inattendues se produisent.

### 1.5.2 Analyses causales

La causalité est un outil très utile pour comprendre le comportement individuel des occurrences de protéines dans un modèle Kappa. Son but est d'étudier en quoi certains événements ont été nécessaires pour que d'autres événements aient pu avoir lieu.

Une trace causale est alors un ensemble d'événements dont certaines paires sont ordonnées par une relation de causalité. Celle-ci indique si l'application d'un événement a rendu possible l'application d'un autre. Un exemple de trace causale est donné en Fig. 1.2. Il s'agit de l'ensemble des événements pour qu'une occurrence du récepteur membranaire recrute une occurrence de la protéine *Sos* par le biais de son site *Y68*. Il faut tout d'abord activer deux occurrences du récepteur membranaire *EGFR* en les liant à des occurrences du ligand *EGF*. Les deux occurrences du récepteur peuvent alors établir une liaison symétrique, puis une liaison asymétrique ce qui permet de différencier une des deux occurrences du récepteur membranaire. Le site *Y68* de cette occurrence peut alors être phosphorylé pour qu'il puisse se lier à une occurrence de la protéine de transport *Grb2*. Indépendamment, cette occurrence de la protéine de transport peut s'être liée à une occurrence de la protéine *Sos*.

Dans cette trace, tous les événements sont nécessaires, mais d'autres *scenarii* peuvent exister. Par exemple, les occurrences du récepteur membranaire peuvent recruter une occurrence de la protéine *Sos* par le biais du site *Y48*, ce qui donne lieu à une autre trace causale. Une trace causale décrit, en fait, un ensemble d'événements qui sont nécessaires dans un scénario potentiel.

Les traces causales sont obtenues en partant des résultats de la simulation, en relevant toutes les occurrences d'un événement d'intérêt. Pour chaque occurrence, une trace est extraite en collectant les événements nécessaires à cette occurrence ou récursivement à tout autre événement lui-même nécessaire [46]. Les événements sont ensuite organisés sous la forme d'un graphe acyclique orienté grâce à la transformation de Mazurkiewicz

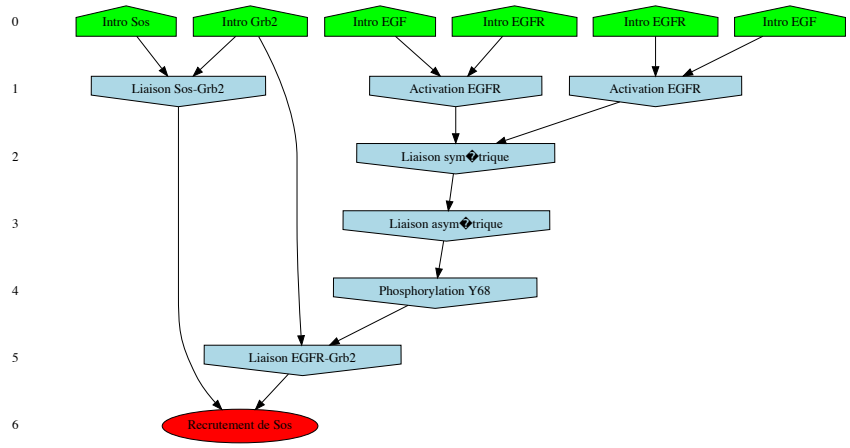



Figure 1.2: Une des deux traces causales pour le recrutement d'une occurrence de la protéine *Sos* par une occurrence du récepteur membranaire. Les nœuds verts représentent l'introduction des occurrences de protéines, les nœuds bleus représentent l'application des règles, le nœud rouge représente le but à observer. Les arcs décrivent les relations causales entre les événements.

[96]. Cette transformation exploite le fait que certains événements, causalement indépendants, commutent. Un moteur de recherche opérationnelle est ensuite utilisé pour retirer de cette trace causale les événements qui peuvent l'être. Une description de cette approche dans un formalisme catégorique est décrit dans cette publication [45].

Les traces causales donnent une vision des voies de signalisation qui privilégie l'acquisition du signal. Dans un modèle, toutes les interactions sont en général réversibles, ce qui est nécessaire pour que l'occurrence d'une kinase, par exemple, puisse agir sur plusieurs occurrences de sa protéine cible à tour de rôle. Cet aspect, gestion de ressources, n'est pas du tout décrit dans les traces causales. Les traces causales ne peuvent dont pas remplacer les règles d'un modèle. Il s'agit juste d'un outil pour comprendre comment un objectif peut être atteint, mais qui ne permet pas à lui seul de définir le comportement collectif du modèle.

Les traces causales dépendent fortement de la syntaxe du langage. En effet, la syntaxe définie quelles préconditions peuvent être utilisées dans les règles, ce qui a une incidence sur le fait que deux événements puissent être vus ou non, comme indépendants causalement. Aussi, le fait que Kappa utilise de la réécriture hors contexte où seuls les sites qui ont une importance dans une interaction ont besoin d'être mentionnés, permet d'avoir plus d'événements qui commutent. Chaque trace causale peut alors résumer un plus grand nombre de traces classiques.

En Fig. 1.3 est considéré l'exemple d'une sorte de protéines avec deux sites de phosphorylation. Chaque site peut être phosphorylé indépendamment de l'état de l'autre site, ce qui se traduit en Kappa par les deux règles données en Fig. 1.3(a). Ces règles peuvent être appliquées dans n'importe quel ordre. Il y a donc une seule trace causale pour obtenir une occurrence de protéines doublement phosphorylée. Cette trace est dessinée en Fig. 1.3(e). Dans un réseau réactionnel, les complexes sont nommés et leur structure biochimique ne peut pas être utilisée. Il faut donc quatre réactions pour simuler ces deux règles Kappa. Or, chacune de ces réactions spécifie exactement quel réactif elle utilise, ce qui empêche les réactions de commuter. Il y alors deux traces causales différentes selon que le site de droit ou de gauche ait été phosphorylé en premier.

Pour conclure sur la causalité, il est important de remarquer que les traces causales s'appuient sur une vision positive de la causalité. Ce n'est en général pas suffisant pour comprendre le comportement des voies de signalisation intracellulaires. En effet, il y a souvent dans ces voies des événements qui ne sont certes pas nécessaires mais qui rendent d'autres événements plus probables. C'est le cas d'une interaction qui stabiliserait une structure instable pour lui laisser le temps de réaliser une certaine interaction. D'un point de vue logique, la stabilisation de la structure n'est pas requise. Mais il est improbable que sans elle, l'autre interaction puisse avoir lieu. Ces effets cinétiques sont capturés par les notions de causalité contre-factuelles [77], dont l'adaptation à Kappa [94] ouvre des pistes de recherches pleines de promesses.


(a) Règles d'interaction.

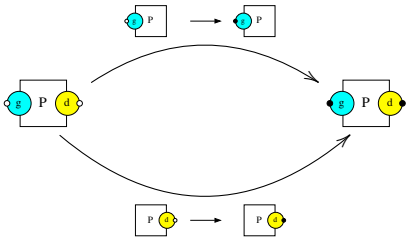

(b) Dictionnaire des complexes biochimiques.

$$A \to C \qquad A \to B$$
$$B \to D \qquad C \to D$$

(c) Réactions.

$$A \to B \to D$$
$$A \to C \to D$$

(d) Les traces causales de ce réseau réactionnel.



(e) L'unique trace causale de cet ensemble de règle.

Figure 1.3: En 1.3(a), un modèle formé de deux règles d'interaction. En 1.3(b) l'ensemble des toutes les sortes de complexes biochimiques accessibles à partir d'une occurrence de la protéine entièrement non phosphorylée, un nom est donné à chaque sorte de complexe biochimique. En Fig. 1.3(c), le réseau réactionnel sous-jacent. Contrairement aux règles d'interaction, les réactions testent l'intégralité de l'état de l'occurrence de la protéine. Ainsi, les réactions qui phosphorylent les deux sites ne commutent pas. Il y a donc deux traces causales, selon que le site droit ou gauche ait été phosphorylé en premier avec les réactions (Fig. 1.3(d)). En Kappa, les règles de phosphorylation d'un site s'appliquent quelque soit l'état de l'autre site. Ainsi les traces causales ne distinguent pas quel site est phosphorylé en premier. Il n'y a alors qu'un seul type de trace causale (Fig. 1.3(e))).

### 1.5.3 Réduction de modèles

La réduction de modèles consiste à simplifier un modèle en ajustant le grain d'observation. Les réductions de modèles peuvent se formaliser comme des transformations de graphes [71], des transformations tropicales [103], des bisimulations [24, 29], ou, tout simplement, des changements de variables [64]. Elles peuvent être classées selon la classe de propriétés qu'elles préservent.

Des outils de réduction exacte permettent de simplifier à la fois les systèmes d'équations différentielles [64, 47] et les systèmes stochastiques [65] qui sont décrit en Kappa. Ces algorithmes trouvent automatiquement des changements de variables par inspection statique des règles initiales des modèles et dérivent des modèles réduits en conséquence. La preuve de correction de ces algorithmes est faite par interprétation abstraite : le modèle réduit définit la projection exacte, par le changement de variables découvert par l'analyse, du comportement transitoire du modèle avant réduction. L'ensemble des complexes biochimiques, le changement de variables et la description extensionnelle du modèle avant réduction ne sont jamais représentés explicitement, ce qui permet à la méthode de passer à l'échelle.

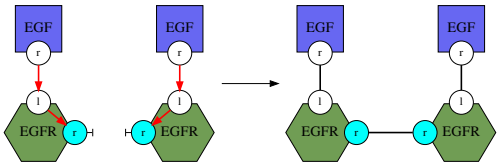
Les outils de réduction de modèles pour Kappa combinent deux types d'abstraction : le premier exploite les symétries potentielles au sein des sites d'interaction des occurrences des protéines du modèle, alors que le second identifie parmi les corrélations éventuelles entre les états des sites des occurrences des protéines, celles qui n'ont aucun impact sur leur comportement collectif. Les symétries sont décrites comme des actions de groupes qui préservent l'ensemble des règles de réécriture qui constituent un modèle [24, 63]. Elles induisent une relation d'équivalence entre les complexes biochimiques qui, elle-même, défini une relation de bisimulation sur les différents états du modèle. Les états en relation seront regroupés en un seul dans le modèle réduit. Intuitivement, cette analyse détecte quels sites ont exactement les mêmes capacités d'interaction et ignore la différence entre ces sites : dans le modèle réduit, la configuration d'une occurrence d'une protéine est définie par le nombre de sites dans un certain état en faisant abstraction de quels sites précis sont dans cet état. Ceci engendre une réduction d'un facteur exponentiel : par exemple, pour un type de protéines avec $n$ sites symétriques pouvant chacun prendre deux états différents, la réduction permet de passer de $2^n$ configurations potentielles à seulement $(n+1)$.

La deuxième approche se base sur l'analyse du flot d'information entre les différents sites d'interaction des complexes biochimiques. Cela permet de comprendre quelles corrélations entre l'état des différents sites peuvent avoir une influence sur le comportement global du système et de passer les autres sous silence. Une approximation qualitative du flot d'information est calculée en répertoriant, au sein des règles de réécriture, tous les chemins entre les sites dont l'état est testé (ceux qui apparaissent dans le membre gauche d'une règle) et les sites dont l'état est modifié (ceux qui apparaissent dans le membre droit de cette règle avec un état différent de celui du membre gauche) (voir en Fig. 1.4(a)). Chaque motif est alors annoté en regroupant le flot d'information présent dans chacune des règles qui peut s'y appliquer. Les motifs intéressants sont ceux pour lesquels il existe un site d'interaction qui est accessible par tous les autres en suivant cette annotation. Par exemple, le complexe biochimique dessiné en Fig. 1.1(a) contient les quatre motifs d'intérêt donnés en Fig. 1.4(b) avec leur annotation. Dans ce modèle, les motifs d'intérêt sont exactement ceux qui décrivent l'état d'un seul site *Y48* ou *Y68*. Ainsi la corrélation entre l'état des différents sites *Y48* et *Y68* n'est plus représentée dans le modèle réduit. D'un point de vue combinatoire, ceci permet de passer de $m^2 \cdot n^2$ complexes biochimiques à $m+n$ motifs d'intérêt (où $m$ et $n$ représentent respectivement le nombre de configurations différentes pour la partie du complexe liée aux sites *Y48* et *Y68*).

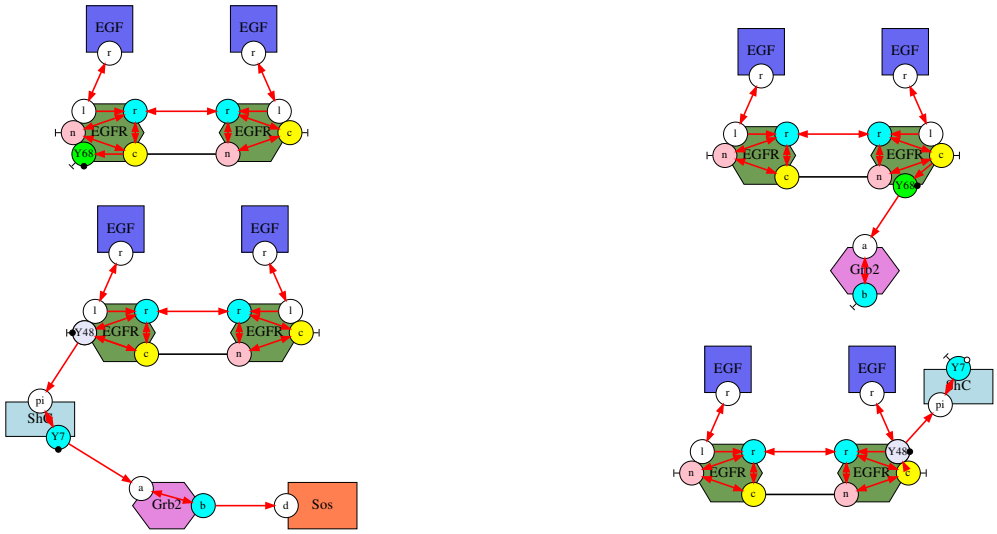
Sur un modèle plus complet [11, 107, 21, 46], cet outil permet de passer de $10^{20}$ complexes biochimiques à $175,000$ motifs d'intérêt, en moins de 10 minutes.

Des méthodes approchées utilisent des formes tronquées de développement formels de la sémantique stochastique [72], alors que les méthodes de tropicalisation exploitent la séparation entre les échelles de temps et de concentration [102]. Ces méthodes ne procurent pas de bornes d'erreur explicites. Par ailleurs, elles nécessitent une description extensionnelle des réseaux réactionnels sous-jacents.

Des méthodes exactes opèrent de manière analytique pour extraire des relations d'équivalence entre les complexes biochimiques de la description explicite des réseaux réactionnels [29] ou même directement sur des systèmes d'équations différentielles [30]. Elles permettent de calculer la meilleure bisimulation en avant, parmi celles qui sont basées sur un partitionnement des variables, et quelles variables prennent toujours la même valeur. La notion de symétries développée pour Kappa est plus restrictive car elle se concentre sur les bisimulations qui correspondent à un certain groupe de transformations. En revanche, elle permet de détecter des relations de proportionnalité entre variables. Par ailleurs, elle ne nécessite de représenter, ni les réseaux réactionnels, ni les systèmes différentiels sous-jacents, évitant ainsi un calcul dont la durée est souvent prohibitive [25].




(a) La règle de formation de dimers annotée par le flot d'information qu'elle induit.




(b) Les quatre motifs d'intérêt qui apparaissent dans le complexe biochimique de la Fig. 1.1(a).

Figure 1.4: En 1.4(a), chaque chemin entre un site dont l'état est testé et un site dont l'état est modifié dans une composante connexe du membre gauche d'une règle induit un flot d'information. Ici, la capacité de lier le site $r$ d'une occurrence du récepteur dépend du fait que cette occurrence soit liée à une occurrence du ligand. En 1.4(b) sont représentés les quatre motifs d'intérêt qui apparaissent dans le complexe biochimique dessiné en Fig. 1.1(a). Ils sont tous quatre annotés par une relation qui spécifie comment l'information se propage – ou s'est propagée – à travers leurs différents sites d'interaction (cette relation est obtenue en recopiant le flot d'information des règles compatibles avec ces motifs). Ils contiennent chacun un site accessible par tous les autres en suivant cette relation.

La réduction de modèles basée sur l'étude du flot d'information est à la fois une généralisation et une formalisation d'approches systématiques existantes [15, 34]. L'utilisation d'un langage formel et l'interprétation abstraite de sa sémantique a permis d'établir formellement la correction de ces approches.

## 1.6 Contributions

Le reste de ce document décrit le langage Kappa [51, 52] sous forme graphique, ainsi que l'analyse statique qui permet de détecter quels motifs peuvent se former lors de l'exécution des modèles [50, 67].

En particulier, la notion de graphe à sites, qui représente l'état des systèmes modélisés, est introduite Chap. 2, alors que celle de règle de réécriture est décrite Chap. 3. Par soucis de simplicité, seul un fragment du langage est considéré. En effet, certaines constructions du langage complet font intervenir des effets de bord (qui peuvent provoquer des transformations de l'état des occurrences de protéines, en dehors des occurrences des motifs de réécriture). S'il est possible d'adapter les différentes définitions pour traiter ces effets de bords, cela n'apporte pas grand chose conceptuellement. Par ailleurs, ce chapitre traite uniquement d'une analyse du comportement qualitatif des modèles, l'aspect quantitatif, les constantes cinétiques, ne sont pas abordées.

L'analyse statique, qui est introduite Chap. 4 permet de détecter, au sein d'un ensemble de motifs d'intérêt paramètre de l'analyse, lesquels ne peuvent jamais se former quelle que soit l'exécution du système. C'est une analyse approchée. Les motifs déclarés inaccessibles sont bien inaccessibles. Par contre, l'analyse n'apporte aucune information à propos des autres motifs. Par soucis d'efficacité, les ensembles de motifs sont organisés (Sect. 4.4.1) sous la forme d'une collection d'arbres de décision dans lesquels des motifs initiaux sont raffinés peu

à peu en ajoutant de l'information contextuelle [67]. Cette analyse est implantée dans l'analyseur statique KaSa [17] et le choix des arbres de décisions, qui paramétrise l'analyse, est fait automatiquement par une pré-analyse. Le chapitre se conclut Sect. 7 et quelques perspectives sont données. La description du langage et de l'analyse reste volontairement assez haut niveau. Une formalisation complète et rigoureuse pour le langage complet est disponible dans les différents articles scientifiques qui sont cités dans le corps du texte.

# Chapter 2

# Graphes à sites

La section présente décrit la notion de graphe à sites, qui permettra de représenter à la fois les différents états possibles pour les systèmes modélisés, mais aussi, les motifs qui seront utilisés dans la section 3 pour décrire, grâce à des règles de réécriture, l'évolution de l'état de ces systèmes.

## 2.1 Signature

En Kappa, il faut tout d'abord définir la signature des modèles. La signature d'un modèle décrit tous les ingrédients qui peuvent intervenir dans celui-ci. Elle peut être représentée graphiquement par une *carte de contacts*, comme celle dessinée en Fig. 2.1. Une carte de contacts comprend des nœuds pour représenter les différentes *sortes de protéines*. Ces nœuds sont nommés et adoptent des formes et des couleurs variées pour les distinguer plus facilement. Chaque sorte de protéines est associée à un ensemble de *sites d'interaction*. Ces sites sont représentés en périphérie de chaque sorte de protéines par des cercles colorés et nommés, eux-aussi. En Kappa, une sorte de protéines donnée ne peut avoir deux sites portant le même nom. Chaque site d'interaction est associé à un ensemble de pastilles colorées qui peuvent servir à représenter son *état d'activation*, comme par exemple le fait d'être – ou non – phosphorylé ou comme le fait d'être méthylé – ou non. Un état d'activation peut aussi éventuellement servir à représenter la localisation d'une occurrence d'une protéine au sein d'un ensemble fini et fixe de compartiments cellulaires. Les sites d'interaction peuvent également porter un *état de liaison* : les sites qui portent le symbole ⊣ peuvent potentiellement rester libre ; la carte de contacts contient aussi des arcs non-orientés entre les sites qui peuvent potentiellement être liés deux à deux. En particulier, un site peut être lié à plusieurs sites dans la carte de contacts (il sera expliqué plus tard que de telles liaisons sont en compétition). Par ailleurs, un site peut être lié à lui-même dans une carte de contacts (il sera expliqué plus tard que ceci signifie que deux sites de deux occurrences différentes d'une même sorte de protéines peuvent être liés entre-eux).

**Exemple 2.1.1** *En Fig. 2.1 est donné un exemple de carte de contacts qui correspond aux premières interactions qui interviennent dans l'activation du facteur de croissance de l'épiderme. Cet exemple est inspiré d'un modèle BNGL disponible dans la littérature [10]. Ce modèle a été étendu pour décrire la liaison asymétrique entre les récepteurs EGFR et traduit en Kappa. Cette carte introduit cinq sortes de protéines : des ligands EGF,*

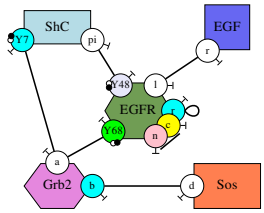



Figure 2.1: Une carte de contacts. Elle définit la signature d'un modèle en donnant la liste de toutes les sortes de protéines, leurs différents sites d'interaction, les différents états d'activation que peuvent prendre ces sites et les différentes liaisons potentielles entre ces sites.

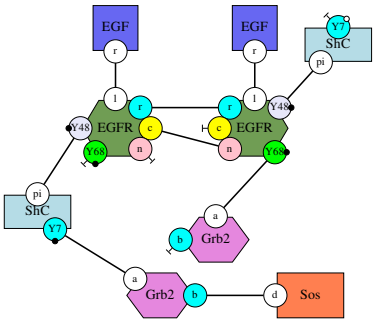



Figure 2.2: Un complexe biochimique. Il contient plusieurs occurrences de protéines. Chaque occurrence documente l'ensemble de ses sites d'interaction. Les sites qui peuvent porter un état d'activation en ont un. Par ailleurs, les sites sont soit libres, soit liés deux à deux.

*des récepteurs membranaires EGFR, des protéines d'échafaudage ShC, des protéines de transport Grb2 et des protéines cibles Sos (cette dernière sera ensuite phosphorylée ce qui initiera les étapes suivantes de la cascade d'interactions). Chaque occurrence du ligand EGF a un seul site qui est nommé r ; chaque occurrence du récepteur membranaire EGFR a six sites qui sont nommés respectivement l, r, c, n, Y48 et Y68 ; chaque occurrence de la protéine d'échafaudage ShC dispose de deux sites qui sont nommés respectivement Y7 et pi ; chaque occurrence de la protéine de transport Grb2 a deux sites qui sont respectivement nommés a et b ; enfin chaque occurrence de la protéine cible Sos a un seul site qui est nommé d. Seuls les sites Y48 et Y68 des occurrences de la protéine EGFR et le site Y7 des occurrences de la protéine ShC portent un état d'activation. Ces sites sont annotés par deux pastilles colorées, une blanche et une noire. La pastille blanche indique que ces sites peuvent être dans l'état non-phosphorylé, alors que la noire indique que ces sites peuvent être dans l'état phosphorylé. De plus, chaque site peut être libre (symbole ⊣) ou lié. Les liaisons possibles entre sites sont entre le site r d'une occurrence de la protéine EGF et le site l d'une occurrence de la protéine EGFR ; entre les sites r de deux occurrences différentes de la protéine EGFR ; entre le site c et le n des occurrences de la protéine EGFR (il sera bientôt expliqué que la carte de contacts ne précise pas si ce doit être entre deux occurrences différentes de la protéine EGFR) ; entre le site Y48 d'une occurrence de la protéine EGFR et le site pi d'une occurrence de la protéine ShC ; entre le site a d'une occurrence de la protéine Grb2 et le site Y68 d'une occurrence de la protéine EGFR ; entre le site a d'une occurrence de la protéine Grb2 et le site Y7 d'une occurrence de la protéine ShC (il y a donc conflit entre ces deux liaisons potentielles) ; enfin entre le site b d'une occurrence de la protéine Grb2 et le site d d'une occurrence de la protéine Sos.*

## 2.2 Complexes biochimiques

Les modèles Kappa décrivent l'évolution d'une soupe de *complexes biochimiques*. Un complexe biochimique est formé de plusieurs occurrences de protéines. Chaque occurrence d'une protéine est associée à un ensemble de sites d'interaction. Chaque site peut éventuellement porter un état d'activation, mais un seul. De ce fait, si un site peut être activé de deux manières différentes, avec un état de phosphorylation et un état de méthylation par exemple, ou si un site peut être doublement activé, doublement phosphorylé par exemple, il est important de définir une pastille différente pour toutes les combinaisons potentielles d'états de ce site. Enfin, chaque site doit être soit libre, soit lié à exactement un autre site. Contrairement à la carte de contacts, un site ne peut pas être lié à lui-même dans un complexe biochimique. De plus, un site ne peut pas être lié simultanément à deux sites. Un complexe biochimique forme un graphe connexe, ce qui signifie qu'il est possible de passer de n'importe quelle occurrence de protéines à n'importe quelle autre, en suivant zéro, un ou plusieurs liens.

**Exemple 2.2.1** *En Fig. 2.2 est donné un exemple de complexe biochimique. Ce complexe est formé de deux occurrences du ligand EGF, de deux occurrences du récepteur membranaire EGFR, de deux occurrences de la protéine d'échafaudage ShC, de deux occurrences de la protéine de transport Grb2 et d'une occurrence de la protéine Sos. Chaque occurrence du récepteur membranaire est liée au site r d'une occurrence du ligand par son site l. Les occurrences de récepteur forment un dimer grâce à une double liaison, une liaison symétrique par leurs sites r respectifs et une liaison asymétrique entre le site c de l'un et le site n de l'autre. L'occurrence du récepteur membranaire dont le site c est lié a son site Y68 phosphorylé et libre, alors que son site Y48 est*

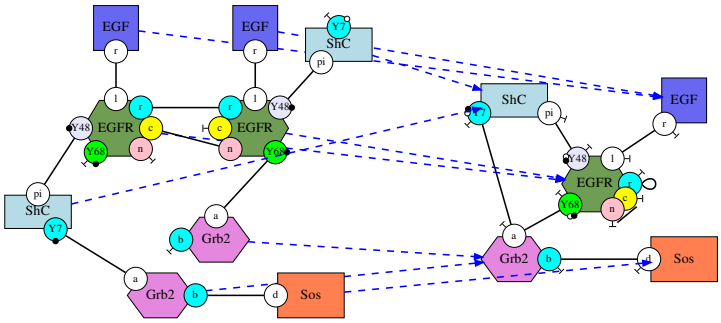



Figure 2.3: L'unique projection entre le complexe biochimique de la Fig. 2.2 et la carte de contacts de la Fig. 2.1. Cette projection est obtenue en associant chaque occurrence de protéines de l'espèce biochimique à l'unique sorte de protéines correspondante dans la carte de contacts.

*phosphorylé et lié au site pi d'une occurrence de la protéine d'échafaudage. Le site Y7 de cette occurrence de la protéine d'échafaudage est phosphorylé et lié au site a d'une occurrence de la protéine de transport dont le site b est lié au site d d'une occurrence de la protéine Sos. L'autre occurrence du récepteur a son site Y48 phosphorylé et lié au site pi de l'autre occurrence de la protéine d'échafaudage. Le site Y7 de cette occurrence de la protéine d'échafaudage n'est ni phosphorylé, ni lié à un autre site. Enfin, le site Y68 de cette seconde occurrence du récepteur membranaire est lié au site a de l'autre occurrence de la protéine de transport. Celle-ci a son site b libre.*

La signature d'un modèle restreint l'ensemble des complexes biochimiques de ce modèle. Tous les complexes biochimiques qui sont corrects du point de vue de la syntaxe ne sont ainsi pas adéquats. Ce rôle est assuré par la carte de contacts, qui d'une part, donne la liste de tous les sites d'interaction de chaque sorte de protéines en indiquant lesquels peuvent porter un état de liaison et un état d'activation et d'autre part, résume l'ensemble des états potentiels de ces sites. Plus précisément, toute occurrence de protéines dans un complexe biochimique doit mentionner les mêmes sites que le nœud correspondant dans la carte de contacts. De plus, un site dont le site correspondant dans la carte de contacts admet au moins un état d'activation doit nécessairement avoir un état d'activation. Il en est de même pour l'état de liaison. Ces contraintes assurent que l'état de chaque occurrence de protéines d'un complexe biochimique est entièrement défini. Trois contraintes supplémentaires assurent que l'état des sites est conforme à la carte de contacts : premièrement, un site ne peut porter un état d'activation que si le site correspondant dans la carte de contacts porte également cet état d'activation ; deuxièmement, un site ne peut être libre que si le site correspondant dans la carte de contacts peut l'être lui-aussi ; troisièmement, deux sites ne peuvent être liés que si les deux sites correspondants le sont également dans la carte de contacts. Ces trois dernières contraintes peuvent se formaliser par le fait que chaque complexe biochimique se projette sur la carte de contacts : ainsi la fonction qui associe à chaque nœud d'un complexe biochimique l'unique nœud de la même sorte dans la carte de contacts doit être un *homomorphisme*. En d'autres termes, la carte de contacts peut être vue comme un repliage de tous les complexes biochimiques du modèle et chaque nœud de la carte de contacts résume toutes les configurations possibles des protéines du type correspondant.

**Exemple 2.2.2** *En Fig. 2.3 est représentée la projection entre le complexe biochimique dessiné dans la Fig. 2.2 et la carte de contacts donnée en Fig. 2.1. Cette projection montre que ce complexe biochimique est compatible avec cette carte de contacts.*

## 2.3 Motifs

L'évolution des complexes biochimiques est décrite par des règles de réécriture. Celles-ci définissent à la fois les conditions qui doivent être réalisées pour qu'une interaction donnée puisse avoir lieu et les effets potentiels de cette interaction. Avant d'expliquer ce que sont ces règles de réécriture, il est nécessaire d'expliquer la notion de motifs qui permet donc de spécifier sous quelles conditions une interaction peut avoir lieu.

Nous nous concentrons sur les motifs connexes. Des motifs plus élaborés peuvent être obtenus en juxtaposant plusieurs motifs connexes. Un *motif connexe* est une portion contigüe de complexe biochimique. De ce fait, il peut comporter zéro, une ou plusieurs occurrences de chaque sorte de protéines. Chaque occurrence de protéines

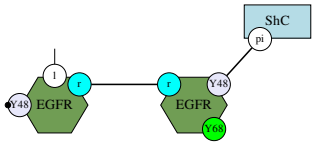



Figure 2.4: Un motif connexe. Il contient plusieurs occurrences de protéines. Chaque occurrence de protéines documente un sous ensemble de ses sites d'interaction. Chaque site peut éventuellement porter un état d'activation et éventuellement un état de liaison (en conformité avec la signature du modèle, donnée en Fig. 2.1). Comme état de liaison, un site peut être libre, lié sans que le site partenaire ne soit précisé ou être lié à un autre site.

est associée à un ensemble de sites d'interaction. Chaque site peut éventuellement porter un état d'activation. Enfin chaque site peut être libre, lié sans que le site auquel il est lié ne soit précisé ou lié exactement à un autre site (différent de lui-même donc). L'état de liaison d'un site peut également ne pas être spécifié.

**Exemple 2.3.1** *En Fig. 2.4 est donné un exemple de motif connexe. Ce motif est formé de deux occurrences du récepteur membranaire EGFR et d'une occurrence de la protéine d'échafaudage ShC. L'occurrence de la protéine d'échafaudage mentionne uniquement son site pi. Celui-ci est lié au site Y48 d'une des deux occurrences du récepteur membranaire. L'état d'activation de ce dernier site n'est pas précisé. Cette occurrence du récepteur membranaire mentionne également son site Y68, sans en préciser ni l'état d'activation ni l'état de liaison, et son site r, lui-même lié au site r de l'autre occurrence du récepteur membranaire. Cet autre occurrence mentionne également son site Y48 qui est phosphorylé mais dont l'état de liaison n'est pas spécifié et son site l qui est lié à un site qui n'est pas précisé.*

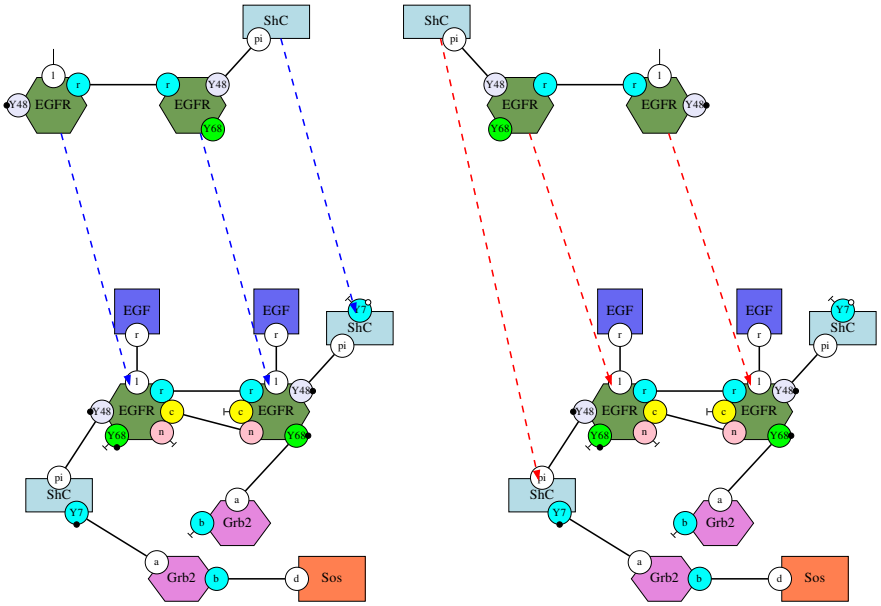
Comme c'était le cas pour les complexes biochimiques, la carte de contacts contraint les motifs que l'on peut écrire dans un modèle. Ainsi, une occurrence de protéines dans un motif ne peut comporter que des sites d'interaction qui sont associés à cette sorte de protéines dans la carte de contacts. Un site ne peut porter un état d'activation que si le site correspondant dans la carte de contacts admet cet état d'activation. Un site ne peut être libre que si le site correspondant peut être libre dans la carte de contacts. Un site ne peut être lié sans préciser à quel site que si le site correspondant est lié à au moins un site dans la carte de contacts. Enfin, deux sites ne peuvent être liés ensemble que si les deux sites correspondants sont liés ensemble dans la carte de contact. En d'autres termes, comme c'était le cas pour les complexes biochimiques, il doit être possible de projeter le motif sur la carte de contacts. Cela veut dire que la fonction qui associe à chaque nœud d'un motif l'unique nœud de la même sorte dans la carte de contacts est un homomorphisme.

## 2.4 Plongements entre motifs

Un motif peut contenir plus ou moins d'information. En effet, il est possible d'ajouter des sites dans une occurrence de protéines qui ne mentionne pas tous ses sites. Par ailleurs, il est possible d'ajouter un état de liaison et/ou un état d'activation à un site qui en manque. Il est possible de préciser à quel site un site est lié quand le partenaire de celui-ci n'est pas précisé. Il est même possible de lier un site au site d'une nouvelle occurrence de protéines. Nous dirons alors que le premier motif apparait dans le second ou encore que le second motif contient une occurrence du premier. Dans ce cas, la relation entre les occurrences de protéines du motif initial et celles du motif ainsi obtenu est formalisée par un plongement. Un *plongement* d'un motif vers un autre motif est une fonction qui envoie chaque occurrence de protéines du premier motif vers une occurrence de protéines du second tout en préservant la structure des graphes à sites, c'est à dire les sortes de protéines, les sites qui sont mentionnés, les états d'activation et les états de liaisons qui sont documentés.

Il est intéressant de remarquer que les complexes biochimiques sont des motifs connexes particuliers. Dans ces derniers, chaque occurrence de protéines décrit tous ses sites, avec un état d'activation et un état de liaison quand ils en ont un. Il n'est donc pas possible d'ajouter d'information dans les complexes biochimiques. Un complexe biochimique ne peut se plonger dans aucun autre motif connexe.

**Exemple 2.4.1** *Deux exemples de plongements sont donnés en Fig. 2.5. Ce sont les seuls plongements entre ce motif et ce complexe biochimique. Dans le premier (voir en Fig. 2.5(a)) l'unique occurrence de la protéine d'échafaudage du motif est associée à l'occurrence de la protéine d'échafaudage du complexe biochimique*




(a) Premier plongement. (b) Second plongement.

Figure 2.5: Deux plongements entre le motif donné dans la Fig. 2.4 et le complexe biochimique donné dans la Fig. 2.2. En 2.5(a) l'occurrence de la protéine d'échafaudage est associée à l'occurrence de la protéine d'échafaudage dont le site *Y7* est libre. En 2.5(b) l'occurrence de la protéine d'échafaudage est associée à l'occurrence de la protéine d'échafaudage dont le site *Y7* est lié.

*dont le site Y7 est libre. L'occurrence du récepteur membranaire qui est liée à l'occurrence de la protéine d'échafaudage du motif est associée à l'occurrence du récepteur membranaire qui est liée à l'occurrence de la protéine d'échafaudage dont le site Y7 est libre. Enfin, l'autre occurrence du récepteur membranaire du motif est associé à l'autre occurrence du récepteur membranaire du complexe biochimique. Il est possible de remarquer que le site l de cette dernière occurrence du récepteur est lié dans le motif, sans que le site partenaire ne soit précisé, alors qu'il est explicitement lié au site r d'une occurrence du ligand dans le complexe biochimique. Dans le second plongement (voir en Fig. 2.5(b)) l'occurrence de la protéine d'échafaudage du motif est associée à l'occurrence de la protéine d'échafaudage du complexe biochimique dont le site Y7 est lié. L'occurrence du récepteur membranaire qui est liée à l'occurrence de la protéine d'échafaudage du motif est associée à l'occurrence du récepteur membranaire qui est liée à l'occurrence de la protéine d'échafaudage dont le site Y7 est lié. Enfin, l'autre occurrence du récepteur membranaire du motif est associée à l'autre occurrence du récepteur membranaire du complexe biochimique.*

Il est important de remarquer qu'un plongement d'un motif connexe vers un autre motif est entièrement caractérisé par l'image d'une occurrence de protéines. Pour avoir les autres associations, il suffit de suivre les liens et d'utiliser le fait qu'ils sont nécessairement préservés par le plongement. Cette propriété facilite la recherche d'occurrences de motifs dans les autres. Les graphes Kappa sont dits *rigides* [47, 100].

# Chapter 3

# Réécriture de graphes à sites

Les motifs vont permettre de spécifier l'évolution potentielle de l'état des systèmes modélisés en Kappa, grâce à des règles de réécriture. C'est l'objet de cette section.

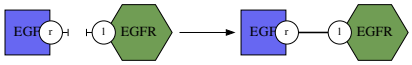
Afin de simplifier la présentation, seul un fragment du langage Kappa est présenté. En particulier, les règles de réécriture qui sont introduites dans cette section n'engendrent pas d'effets de bord. Un effet de bord est une transformation à l'extérieur du membre gauche des règles. Les effets de bords peuvent être dus à des sites libérés sans préciser à quels sites ils sont liés ou à des occurrences de protéines dégradées. Ces constructions n'ont pas été considérées afin de simplifier la présentation. Cela a permis de présenter tous les différents concepts de la syntaxe et de la sémantique de Kappa sous forme graphique.

## 3.1 Règles d'interaction

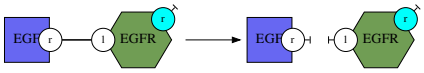
Les complexes biochimiques peuvent se transformer en appliquant des règles d'interaction. Une *règle d'interaction* est définie par une paire de motifs, qui contiennent exactement les mêmes sortes de protéines. Le premier motif spécifie quelles conditions locales doivent être réalisées pour permettre à l'interaction de se produire. La différence entre ces deux motifs décrit quelle transformation résulte de cette interaction. Aussi le second motif d'une règle doit pouvoir être obtenu à partir du premier en changeant uniquement l'état d'activation et/ou de liaison de certains sites d'interaction.

**Exemple 3.1.1** *Des exemples de règles d'interaction sont données en Fig. 3.1. Celles-ci décrivent les interactions qui sont impliquées dans le recrutement des occurrences de la protéine cible par les occurrences du récepteur membranaire par leur site Y68, dans le modèle des premières étapes de l'acquisition du facteur de croissance de l'épiderme. Le recrutement par le site Y48 implique des règles d'interaction similaires, qui ne seront donc pas détaillées. La colonne de gauche décrit les interactions qui font progresser le recrutement d'une occurrence de la protéine cible. La première étape est l'activation d'une occurrence du récepteur membranaire par une occurrence du ligand (voir en Fig. 3.1(a)). En se liant à une occurrence du ligand, une occurrence du récepteur change de conformation et peut alors établir une liaison symétrique avec une autre occurrence du récepteur qui doit pour cela être elle-même activée (voir en Fig. 3.1(c)). Comme seules les occurrences du ligand peuvent se lier aux sites l des occurrences du récepteur, il n'est pas nécessaire de mentionner les occurrences du ligand dans la règle. Il suffit d'écrire que les sites l des deux occurrences du récepteur doivent être liés sans préciser à quels sites. Après cette étape, les deux occurrences du récepteur tiennent le même rôle. Pour les distinguer, une liaison asymétrique peut alors s'établir (voir en Fig. 3.1(e)) entre le site c d'une des deux occurrences et le site n de l'autre occurrence. Le site Y68 de l'occurrence du récepteur qui est liée par son site c peut alors se faire phosphoryler par l'autre occurrence du récepteur membranaire (voir en Fig. 3.1(g)). Cela change la conformation de cette occurrence du récepteur membranaire et lui permet de se lier à une occurrence de la protéine de transport (voir en Fig. 3.1(i)). Indépendamment, les occurrences de la protéine de transport peuvent se lier aux occurrences de la protéine cible (voir en Fig. 3.1(k)).*

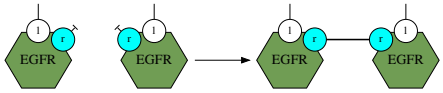
*Chacune de ces interactions est réversible. Cependant les interactions inverses ne peuvent s'effectuer que sous certaines conditions. Ces interactions sont décrites dans la colonne de droite. Les liaisons symétriques entre les occurrences du récepteur membranaire capturent les occurrences du ligand qui ne peuvent alors pas se libérer (voir en Fig. 3.1(b)). Les liaisons asymétriques empêchent les liaisons symétriques de se briser (voir en Fig. 3.1(d)). Les liaisons asymétriques peuvent se briser sans condition (voir en Fig. 3.1(f)). La phosphorylation*



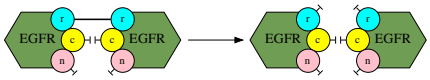

(a) Activation d'une occurrence du récepteur.



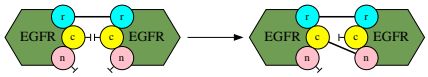

(b) Désactivation d'une occurrence du récepteur.



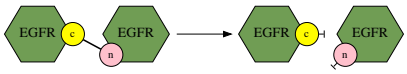

(c) Liaison symétrique



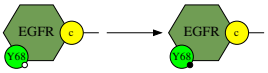

(d) Déliaison du lien symétrique.




(e) Liaison asymétrique.




(f) Déliaison du lien asymétrique.

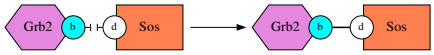



(g) Phosphorylation d'une occurrence du récepteur.


(h) Déphosphorylation d'une occurrence du récepteur.


(i) Recrutement d'une occurrence de la protéine transporteur.


(j) Déliaison d'une occurrence de la protéine transporteur.




(k) Liaison d'une occurrence de la protéine transporteur à une occurrence de la protéine cible


(l) Déliaison d'une occurrence de la protéine transporteur d'une occurrence de la protéine cible

Figure 3.1: Règles d'interaction impliquées dans le recrutement d'une occurrence de la protéine cible par la voie de signalisation courte (sans passer par la protéine d'échafaudage).

*du site Y68 d'une occurrence du récepteur est bloquée quand ce site est lié (voir en Fig. 3.1(h)). Les liaisons entre les occurrences du récepteur et les occurrences de la protéine de transport d'une part, et celles entre les occurrences de la protéine de transport et celles de la protéine cible d'autre part, peuvent se défaire sans condition (voir en Figs. 3.1(j) et 3.1(l)).*

Dans le langage complet, il est possible de détruire un lien entre deux occurrences de protéines en ne spécifiant qu'un seul des deux sites de liaisons. De plus, une règle peut également détruire des occurrences de protéines. Ces constructions peuvent induire des effets de bord, puisqu'appliquer de telles interactions est susceptible de libérer des sites qui ne sont pas décrits dans les membres gauches des règles correspondantes. Par ailleurs, le langage complet permet aussi de synthétiser de nouvelles occurrences de protéines.

## 3.2 Réactions induites par une règle d'interaction

Comme signalé précédemment, le membre gauche d'une règle d'interaction spécifie dans quel contexte cette interaction peut avoir lieu. Il est alors possible d'ajouter des contraintes sur les conditions d'application d'une règle en raffinant les motifs qui apparaissent dans les membres gauches et droits des règles exactement de la même manière. Une règle d'interaction qui ne peut plus être raffinée (sans ajouter de nouvelles composantes connexes) est alors appelée une *règle-réaction* [80].

**Exemple 3.2.1** *En Fig. 3.2 est montré un exemple de deux raffinements d'une même règle d'interaction en deux règles-réactions. La règle d'interaction est celle qui permet de casser, en l'absence de lien asymétrique, le lien symétrique entre deux occurrences du récepteur membranaire (voir en Fig. 3.1(d)).*

1. *Dans le premier raffinement (voir en Fig. 3.2(a)), la règle est appliquée à un dimer dont la première occurrence du récepteur est liée par son site Y48 à une occurrence de la protéine d'échafaudage dont le site Y7 est libre et phosphorylé et par son site Y68 à une occurrence de la protéine de transport elle-même liée à une occurrence de la protéine cible. La deuxième occurence du recepteur de ce dimer est liée par son site Y48 à une occurrence de la protéine d'échafaudage dont le site Y7 est libre et non-phosphorylé et par son site Y68 à une occurrence de la protéine de transport dont le site b est libre.*

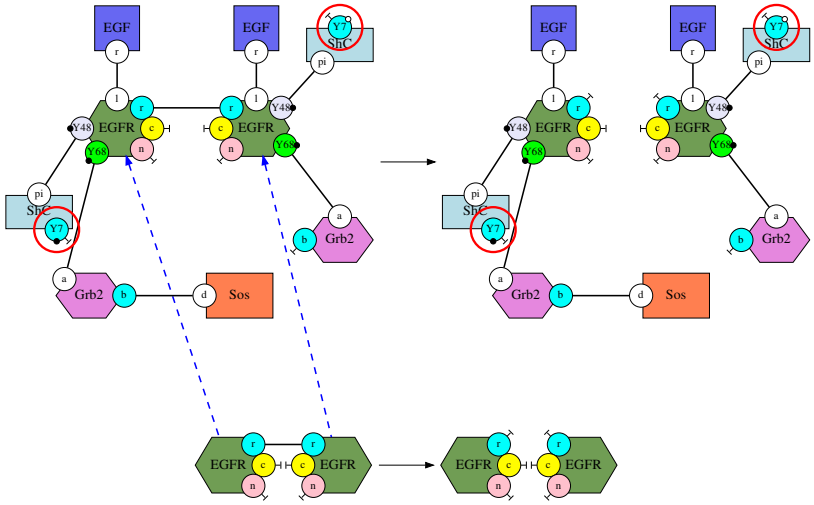
2. *Dans le second (voir en Fig. 3.2(b)), les deux occurrences de la protéine d'échafaudage ont été interverties. Ainsi, la règle est appliquée à un dimer dont la première occurrence du récepteur est liée par son site Y48 à une occurrence de la protéine d'échafaudage dont le site Y7 est libre et non-phosphorylé et par son site Y68 à une occurrence de la protéine de transport elle-même liée à une occurrence de la protéine cible. La seconde occurrence du récepteur de ce dimer est liée par son site Y48 à une occurrence de la protéine d'échafaudage dont le site Y7 est libre et phosphorylé et par son site Y68 à une occurrence de la protéine de transport dont le site b est libre.*

*Bien que les deux complexes biochimiques qui apparaissent dans les membres gauches de ces deux règles-réactions soient formés exactement des mêmes occurrences de protéines et dans les mêmes configurations, puisque seul l'agencement entre ces occurrences change, il apparait que les deux règles-réactions obtenues ne produisent pas les mêmes complexes biochimiques. Ceci justifie pleinement le choix, dans Kappa, de représenter la topologie des liens entre les occurrences de protéines. Sans celle-ci il est impossible de décrire fidèlement la séparation des occurrences de récepteurs, tout en respectant la distribution des différentes occurrences de protéines et de leurs configurations dans chacun des complexes biochimiques résultant de cette séparation. Par exemple, dans le langage BCS [54], les complexes biochimiques sont représentés par l'ensemble des occurrences de protéines qui les constituent, ainsi que leurs configurations, mais sans préciser la topologie des liens entre ces occurrences de protéines. Aussi il est impossible de représenter fidèlement la règle de déliaison qui est déssinée en Fig. 3.1(d) dans ce langage.*

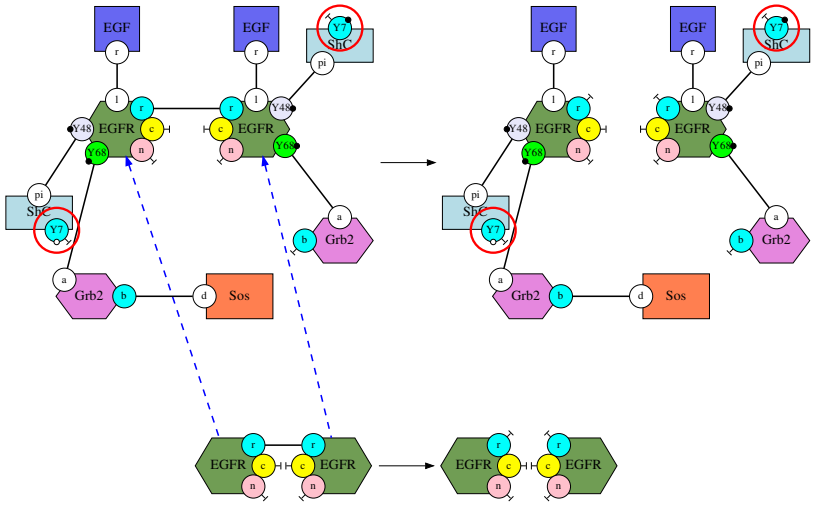
## 3.3 Réseaux de réactions sous-jacents

Un ensemble de règles peut alors être traduit en un ensemble – éventuellement infini – de règles-réactions en remplaçant chaque règle d'interaction par l'ensemble des règles-réactions qui peuvent être obtenues comme raffinement de ces règles.

Ensuite, quitte à nommer les différents complexes biochimiques qui peuvent intervenir dans les règles-réactions ainsi obtenues, nous pouvons assimiler ces règles-réactions à un réseau de réactions (éventuellement




(a) Premier raffinement.



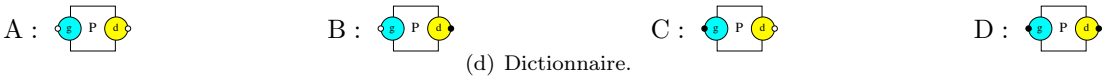

(b) Second raffinement.

Figure 3.2: Deux exemples de raffinements d'une même règle d'interaction en deux règles-réactions. Les différences entre ces deux raffinements sont mises en valeur par des cercles rouges (les deux occurrences de la protéine *ShC* ont été échangées). Dans les deux cas, la règle-réaction est obtenue en ajoutant dans le membre gauche et dans le membre droit de la règle d'interaction exactement la même information sur le contexte d'application de la règle.


(a) Carte de contacts.


(b) Règles d'interaction.


(c) Règles-réactions.




(d) Dictionnaire.

$A \to C$
$A \to B$
$B \to D$
$C \to D$
$C \to A$
$B \to A$
$D \to B$
$D \to C$

(e) Réactions.

Figure 3.3: Un modèle formé d'une carte de contacts et de quatres règles d'interaction et sa traduction sous forme réseau réactionnel.

infini), dans lequel chaque réaction est spécifiée par une liste de réactifs et une liste de produits parmi un ensemble d'espèces biochimiques représentées uniquement par des noms (en passant sous silence leurs structures biochimiques). Ce réseau de réactions est défini de manière unique modulo le choix des noms associés aux espèces biochimiques.

**Exemple 3.3.1** *Pour conclure cette section, nous détaillons la génération d'un réseau de réactions à partir d'un ensemble jouet de règles Kappa. Nous considérons un modèle avec une seule sorte de protéines qui admet deux sites, g et d, chacun pouvant être phosphorylé ou non. La signature du modèle est donnée par la carte de contacts qui est dessinée en Fig. 3.3(a). Le phosphorylation et la déphosphorylation de chaque site dans une occurrence de protéines peut se faire indépendamment de l'état de l'autre site, ce qui est formalisé dans les quatre règles données en Fig. 3.3(b). Ainsi ni les règles de phosphorylation, ni celles de déphosphorylation d'un site, ne mentionnent l'état de phosphorylation de l'autre site.*

*Pour obtenir les règles-réactions associées à ce modèle jouet, il suffit d'expliciter dans quel contexte local les interactions peuvent se produire. Ainsi chaque règle Kappa donne ici lieu à deux règles-réactions selon que le site qui n'est pas mentionné dans la règle initiale est phosphorylé ou non. Ces règles-réactions sont données en Fig. 3.3(c).*

*La prochaine étape est de nommer les différents complexes biochimiques qui interviennent dans les règles-réactions ainsi obtenues. Une occurrence de protéines dont aucun site n'est phosphorylé, est appelée A, une occurrence de protéines dont seul le site d est phosphorylé, est appelée B, une occurrence de protéines dont seul le site g est phosphorylé, est appelée C et une occurrence de protéines dont les deux sites sont phosphorylés, est appelée D. Les réactions données en Fig. 3.3(e) sont obtenues en remplaçant chaque occurrence de complexe biochimique par son nom dans les règles-réactions.*

Le choix d'une sémantique en terme de réseaux de réactions a été fait pour simplifier la présentation. C'était ainsi que le langage BNGL avait été implanté initialement [10]. Une telle sémantique est toutefois assez peu utile en pratique, car un modèle Kappa engendre en général un trop grand nombre de réactions. Par contre, la

sémantique de Kappa peut être formalisée directement, soit sous forme d'une algèbre de processus [50, 65], soit dans un cadre catégorique [45, 63]. La première méthode est plus opérationnelle alors que la seconde abstrait au contraire beaucoup de détails. Il faut cependant noter que les cadres catégoriques usuels de la réécriture de graphes, que ce soit par push-out simples [95], push-out doubles [37] ou sesqui-pushout [36]) ne représentent pas fidèlement les effets de bord avec la définition usuelle des plongements entre graphes à sites. Deux approches connues permettent d'y remédier. Il est possible soit de changer la définition des plongements [45, 63], soit d'enrichir les objets de la catégorie par des contraintes [6].

La simulation d'un modèle Kappa opère directement par réécriture du graphe qui représente l'état du système, sans avoir à considérer le réseau de réactions sous-jacent [48, 18].

# Chapter 4

# Analyse des motifs accessibles

Si la carte de contacts (e.g. voir en Fig. 2.1 à la page 13) donne un aperçu rapide de toutes les interactions potentielles entre les différents sites des occurrences des protéines dans un modèle, elle n'est en générale pas suffisante pour décrire précisément la structure de ses complexes biochimiques. En effet, l'état des différents sites d'interaction d'un complexe biochimique est souvent contraint par des invariants structurels. Par exemple, dans le modèle des premières étapes de l'acquisition du facteur de croissance de l'épiderme, les sites *Y48* et *Y68* des occurrences du récepteur membranaire, ainsi que le site *Y7* des occurrences de la protéine d'échafaudage, ne peuvent être liés à un autre site sans être phosphorylés (à moins que ce soit le cas dans l'état initial). Par ailleurs, lorsque les deux sites $r$ et $c$ d'une occurrence du récepteur sont liés simultanément, ils sont nécessairement liés respectivement au site $r$ et au site $n$ d'une même occurrence du récepteur (ce qui forme une double liaison). Un autre exemple concerne les modèles avec des compartiments, comme, par exemple, une cellule dont on distingue le noyau du cytoplasme. La localisation de chaque occurrence de protéines peut alors être spécifiée comme l'état d'activation d'un site fictif. Dans de tels modèles, toutes les occurrences de protéines d'un même complexe biochimique sont en général localisées dans un même compartiment, ce qui se traduit par la contrainte que le site fictif de deux occurrences de protéines liées entre elles doit toujours être dans le même état. Dans certains cas, il est toutefois possible d'avoir des complexes trans-membranaires avec des portions localisées dans des compartiments voisins, c'est à dire de part et d'autre d'une membrane.

Dans cette section est décrite une analyse statique qui permet de détecter automatiquement ces contraintes, afin de vérifier que les propriétés auxquelles peut s'attendre le modélisateur sont bien vérifiées ou bien de détecter certaines erreurs de modélisation. En particulier, cette analyse permet de trouver des *règles mortes*. Ce sont des règles qui ne peuvent jamais s'appliquer dans un modèle, car les contraintes qui sont exprimées dans leurs membres gauches ne sont pas réalisables. C'est souvent la conséquence d'erreurs typographiques (par exemple, quand une même sorte de protéines est désignée par deux noms différents dans l'encodage d'un modèle), d'un état initial incomplet, d'interactions manquantes dans le modèle (par exemple, quand l'activation d'un site n'est pas décrite, alors qu'elle est nécessaire pour la suite de la cascade d'interactions) ou de conditions causales plus complexes qu'il faut alors élucider.

Cette analyse est implantée dans l'analyseur statique KaSa [17] et intégrée dans la plate-forme de modélisation en ligne dédiée au langage Kappa [20]. Ceci permet d'assister le modélisateur pendant l'écriture du modèle en lui fournissant les contraintes structurelles qui sont vérifiées par les complexes biochimiques et en l'avertissant de la présence de règles mortes, après chaque ajout ou modification d'une règle d'interaction.

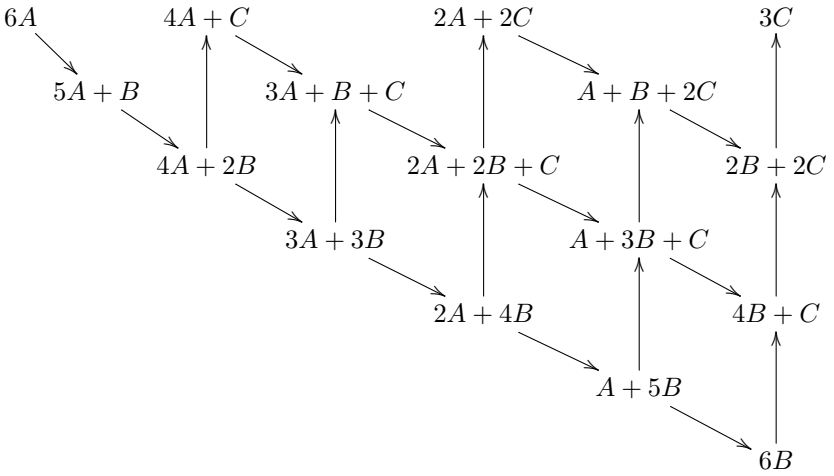
## 4.1 Accessibilité dans un réseau réactionnel

La première étape consiste à définir l'ensemble des états accessibles dans un modèle Kappa. Comme nous l'avons vu dans la section 3.3 page 21, un modèle Kappa induit un réseau réactionnel, ce qui permet de définir directement l'ensemble des états accessibles d'un modèle Kappa sans recourir à des constructions compliquées.

Soit un réseau réactionnel, c'est à dire un ensemble d'espèces biochimiques $\mathcal{S}$ et un ensemble de réactions $\mathcal{R}$. Chaque réaction est donnée par deux listes d'espèces biochimiques : ses réactifs et ses produits. Ce réseau induit un système de transitions dans lequel l'*état* du réseau est défini comme un certain nombre (éventuellement nul) d'occurrences de chacune des espèces biochimiques – c'est à dire une fonction de l'ensemble $\mathcal{S}$ vers l'ensemble $\mathbb{N}$ des entiers naturels — et les *transitions* permettent de sauter d'un état à un autre en consommant les réactifs

$$\begin{cases} A \to B \\ 2B \to C \end{cases}$$

(a) Un réseau réactionnel.




(b) Système de transitions sous-jacent.

Figure 4.1: Un réseau réactionnel et sa sémantique. En 4.1(a) un réseau réactionnel formé de deux réactions. La première permet de transformer une occurrence de l'espèce biochimique $A$ en une occurrence de l'espèce biochimique $B$, la seconde permet de transformer deux occurrences de l'espèce biochimique $B$ en une occurrence de l'espèce biochimique $C$. La restriction de l'ensemble de toutes les transitions possibles aux états qui sont atteignables à partir d'un état initial formé de six occurrences de la protéine $A$ est dessinée en 4.1(b) sous la forme d'un système de transitions.

d'une réaction et en ajoutant les produits de cette même réaction (en tenant compte de leur multiplicité). Une transition n'est possible que si l'état courant du système contient tous les réactifs qui sont nécessaires à la réaction (en tenant compte, une nouvelle fois, de leur multiplicité). Une transition d'un état $q$ vers un autre état $q'$ est alors notée $q \to q'$.

**Exemple 4.1.1** *Un système de transitions est donné comme exemple en Fig. 4.1(b). Il correspond à la restriction du système de transitions associé aux réactions qui sont données en Fig. 4.1(a) aux états accessibles à partir d'un état initial formé de six occurrences de l'espèce biochimique A. Dans ce réseau, la somme entre le nombre d'occurrences de A, de B et deux fois celui de C est toujours égal à la quantité initiale de A. En effet, cette quantité n'est modifiée par l'application d'aucune des réactions du réseau.*

Étant donné un ensemble d'états initiaux potentiels, $\mathcal{I} \subseteq \mathcal{S}^{\mathbb{N}}$, nous définissons l'ensemble des états accessibles comme étant ceux susceptibles d'être atteints à partir d'un état initial (de l'ensemble $\mathcal{I}$) en appliquant un nombre arbitraire (éventuellement nul) de transitions. Cet ensemble peut se définir comme le plus petit point-fixe de la fonction suivante :

$$\mathbb{F} : \begin{cases} \wp(\mathcal{S}^{\mathbb{N}}) & \to \wp(\mathcal{S}^{\mathbb{N}}) \\ X & \to \mathcal{I} \cup \{ q' \mid \exists q \in X,\ q \to q' \}\,. \end{cases}$$

Il faut noter que la fonction $\mathbb{F}$ est croissante, ce qui signifie que si $X_1$ et $X_2$ sont deux ensembles d'états tels que l'ensemble $X_1$ soit un sous-ensemble de l'ensemble $X_2$, alors l'ensemble $\mathbb{F}(X_1)$ est nécessairement un sous-ensemble de l'ensemble $\mathbb{F}(X_2)$ lui-aussi. Comme, de plus, cette fonction est définie sur l'ensemble des parties d'un ensemble, le *théorème de Tarski* [110] assure que la fonction $\mathbb{F}$ admet un point fixe, plus petit que tout autre point fixe de $\mathbb{F}$. Ce plus-petit point fixe, que l'on note $\mathit{lfp}\,\mathbb{F}$, est en fait l'ensemble des états accessibles.

Malheureusement, le calcul de ce plus petit point fixe peut être coûteux, voire ne pas terminer. Ceci motive la construction d'abstractions pour calculer un sur-ensemble des états accessibles en un temps de calcul acceptable.

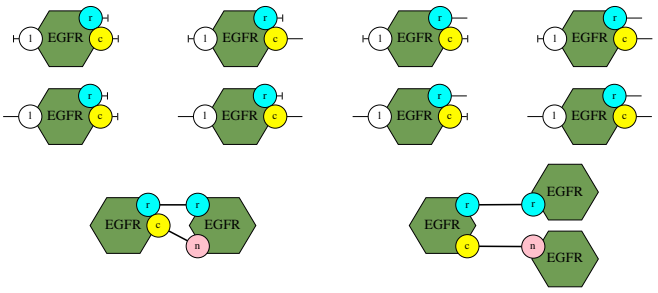



Figure 4.2: Un ensemble de motifs d'intérêt pour l'analyse des complexes biochimiques accessibles dans le modèle des premières interactions qui interviennent dans l'acquisition du facteur de croissance de l'épiderme. Les huit premiers motifs permettent de s'intéresser aux relations potentielles entre l'état de liaison des sites $l$, $r$ et $c$ dans les occurrences du récepteur membranaire. Ces 8 motifs correspondent exactement à chaque combinaison possible pour l'état de ces 3 sites, chacun de ces sites pouvant être libre ou lié. Les deux derniers motifs permettent de distinguer deux occurrences du récepteur liées par une double liaison d'une chaîne d'au moins trois occurrences du récepteur.

## 4.2 Abstraction d'un ensemble d'états

Lorsqu'un réseau est induit par un modèle Kappa, la structure biochimique associée aux espèces de ce réseau peut être utilisée pour construire une abstraction. Une possibilité consiste à choisir un ensemble de motifs connexes afin d'abstraire les ensembles d'états par le sous-ensemble parmi ces motifs de ceux qui apparaissent au moins une fois dans au moins un état de cet ensemble. Le choix des motifs connexes considérés est important : il définit le compromis entre l'expressivité de l'abstraction, c'est à dire son niveau d'approximation, et sa complexité, c'est à dire le coût pour effectuer des calculs à ce niveau d'abstraction.

**Exemple 4.2.1** *Un exemple de motifs d'intérêt pour le modèle des premières interactions de l'acquisition du facteur de croissance de l'épiderme est donné en Fig. 4.2. Les huit premiers motifs se concentrent sur l'analyse des relations potentielles entre l'état des sites l, r et c dans les occurrences du récepteur membranaire. Ils correspondent à toutes les combinaisons syntaxiquement possibles pour l'état de liaison de ces* 3 *sites. Ce sont des* vues locales *(ou plus précisément des sous-vues locales) [50]. Elles permettent d'abstraire un ensemble de complexes biochimiques par l'ensemble de toutes les configurations potentielles de toutes ses occurrences de protéines, vues indépendamment les unes des autres. Ceci revient à garder uniquement l'information à propos de l'état de liaison et l'état d'activation de chaque site dans chaque occurrence de protéines tout en passant sous silence à quel site chaque site lié l'est.*

*La formation de dimers dans ce modèle fait intervenir des doubles liaisons. Il est légitime de se demander s'il est possible de former des chaines comportant successivement au moins trois occurrences du récepteur membranaire. C'est le but des deux derniers motifs de l'ensemble. Ils permettent de distinguer le cas d'une double liaison entre deux occurrences du récepteur de celui de trois occurrences du récepteur liées consécutivement, en s'interrogeant pour chaque occurrence du récepteur membranaire dont les sites r et c sont liés, si elle peut être liée à une même occurrence du récepteur ou si elle peut être liée à deux occurrences différentes. En toute rigueur, pour s'assurer qu'une chaine d'au moins trois occurrences du récepteur ne peut pas se former, il faut également considérer des motifs d'intérêt similaires pour la paire de sites r et n et la paire de sites c et n.*

Plus précisément, l'abstraction est paramétrée par le choix d'un ensemble $\mathcal{P}$ de motifs connexes. L'ensemble $\mathcal{P}$ regroupe des motifs d'intérêt, ainsi que des motifs qui seront utilisés de manière intermédiaire dans la preuve que certains de ces motifs d'intérêt sont inaccessibles. Une sous-partie de l'ensemble $\mathcal{P}$ est appelée une propriété abstraite. Chaque propriété abstraite représente un ensemble d'états concrets : un état concret $q$ sera dît compatible avec une propriété abstraite $X^\sharp$ si et seulement si aucun motif qui est dans l'ensemble $\mathcal{P}$ sans être dans l'ensemble $X^\sharp$ n'apparait dans un complexe biochimique présent dans l'état $q$. L'ensemble de tous les états concrets compatibles avec la propriété abstraite $X^\sharp$ est alors noté $\gamma_{\mathcal{P}}(X^\sharp)$. Qui peut le plus, peut le moins : plus nombreux sont les motifs autorisés, plus nombreux sont les complexes biochimiques compatibles. La fonction $\gamma_{\mathcal{P}}$ est donc croissante. Elle permet de définir formellement la notion d'abstraction d'un ensemble d'état : une propriété abstraite $X^\sharp$ sera dite être une abstraction d'un ensemble d'état $X$ si et seulement si l'ensemble $X$ est

inclus dans l'ensemble $\gamma_{\mathcal{P}}(X^\sharp)$. La fonction $\gamma_{\mathcal{P}}$ est couramment appelée la *fonction de concrétisation*. De plus l'image d'une propriété abstraite par cette fonction, est appelée sa *concrétisation*.

Réciproquement, étant donné un ensemble d'états $X$, l'ensemble des éléments de l'ensemble $\mathcal{P}$ qui apparaissent dans au moins un complexe biochimique d'un état élément de l'ensemble $X$ sera noté $\alpha_{\mathcal{P}}(X)$. La fonction $\alpha_{\mathcal{P}}(X)$ est croissante également. La propriété abstraite $\alpha_{\mathcal{P}}(X)$ est en fait la *meilleure approximation* de l'ensemble d'états $X$, ce qui signifie que d'une part c'est une abstraction de l'ensemble $X$ (i.e. $X \subseteq \gamma_{\mathcal{P}}(\alpha_{\mathcal{P}}(X))$) et que d'autre part c'est un sous-ensemble de toute autre abstraction de $X$ (i.e. pour tout sous-ensemble $Y$ de l'ensemble $\mathcal{P}$ tel que $X \subseteq \gamma_{\mathcal{P}}(Y)$, l'inclusion $\alpha_{\mathcal{P}}(X) \subseteq Y$ est vérifiée). La paire de fonctions $(\alpha_{\mathcal{P}}, \gamma_{\mathcal{P}})$ est alors appelée une *correspondance de Galois* [40, 38].

**Exemple 4.2.2** *En Fig. 4.3 est introduit un exemple jouet pour mieux comprendre le comportement des fonctions d'abstraction et de concrétisation. La signature de ce modèle peut être consultée en Fig. 4.3(a). Il existe une seule sorte de protéines, qui est appelée A. Cette protéine est munie de deux sites g et d (pour gauche et droite). La carte de contacts spécifie que chaque site peut être libre et qu'un site g peut être lié à un site d d'une même ou d'une autre occurrence de la protéine A. L'abstraction induite par l'ensemble des motifs d'intérêt donné en Fig. 4.3(b) repose sur les vues locales des occurrences de cette protéine. Elle permet de se poser la question de l'existence ou non, d'une relation entre l'état de liaison des sites g et d dans chaque occurrence de la protéine A.*

**Exemple 4.2.3** *Des exemples de meilleures approximations sont donnés en Fig. 4.3(c). Par abus de langage, nous appelons la meilleure approximation d'un complexe biochimique, la meilleure approximation de l'ensemble formé d'un seul état lui même formé de ce seul complexe. Le modèle admet deux types de complexes biochimiques. Les occurrences de la protéine A peuvent former des chaines d'occurrences de protéines liées successivement par leur site d et g, laissant le site g de la première occurrence de la protéine A et le site d de la dernière occurrence de la protéine A libres. Les occurrences de la protéine A peuvent aussi former des anneaux en reliant le premier et le dernier sites d'une chaine d'occurrences de protéines. La meilleure approximation d'une chaine d'occurrences de la protéine A dépend de la taille de cette chaine. Ainsi, la meilleure approximation d'une chaine réduite à une occurrence de la protéine A est l'ensemble qui contient uniquement la vue locale dont les deux sites sont libres ; la meilleure approximation d'une chaine formée d'exactement deux occurrences de la protéine A est l'ensemble qui contient deux vues locales : l'une avec le site g libre et le site d lié, l'autre avec le site g lié et le site d libre ; enfin la meilleure approximation des chaines d'occurrences de la protéine A de longueur au moins égale à 3 contient également la vue locale dont les deux sites g et d sont liés. Par contre, la meilleure approximation d'un anneau d'occurrences de la protéine A est toujours l'ensemble formé uniquement de la vue locale dont les deux sites g et d sont liées, et ce, quelle que soit la longueur de cette anneau. La fonction qui associe à chaque ensemble de complexe biochimique sa meilleure approximation est distributive. Cela signifie que la meilleure approximation d'un ensemble de complexes biochimiques est l'union de la meilleure approximation des singletons correspondants.*

**Exemple 4.2.4** *Des exemples de concrétisations sont donnés en Fig. 4.3(d). Par définition, la concrétisation de l'ensemble formé uniquement de la vue locale dans laquelle les deux sites sont libres, est l'ensemble de tous les états qui ne contiennent pas d'autres vues locales. Il s'agit donc des états qui ne contiennent que le complexe biochimique composé d'exactement une occurrence de la protéine A. Par la même démarche, la concrétisation de l'ensemble formé uniquement de la vue locale dont les deux sites sont liés est l'ensemble de tous les états qui ne contiennent que des anneaux d'occurrences de le protéine A (quitte à utiliser cette vue locale plusieurs fois). Par contre, tout complexe contenant une vue locale avec exactement un site libre, doit contenir également une vue locale avec l'autre site, libre. De ce fait, la concrétisation d'un ensemble composé d'une seule vue locale avec un site libre et l'autre lié, est l'ensemble ne contenant que l'état vide (qui est noté $\emptyset$). Si seules les deux vues locales où un site est lié et l'autre libre sont autorisées, seules des chaines de deux occurrences de la protéine A peuvent être construites. Enfin, si seule la vue locale avec les deux sites libres est interdite, il est possible de former n'importe quelle chaine d'occurrences de la protéine A de taille au moins égale à 2 et n'importe quel anneau d'occurrences de protéines (sans restriction de taille). La fonction de concrétisation n'est pas distributive (l'image de l'union de deux ensembles de vues locales peut être un sur-ensemble strict de l'union de leurs images).*

Les fonctions $\alpha_{\mathcal{P}}$ et $\gamma_{\mathcal{P}}$ se composent dans les deux sens. Ces compositions sont révélatrices des traits principaux du choix de l'abstraction. La composée $\gamma_{\mathcal{P}} \circ \alpha_{\mathcal{P}}$ caractérise le niveau d'approximation. En effet, pour tout ensemble d'états $X$, $\gamma_{\mathcal{P}}(\alpha_{\mathcal{P}}(X))$ est le plus grand ensemble d'état qui a la même meilleure approximation


(a) Carte de contacts.


(b) Motifs d'intérêt.

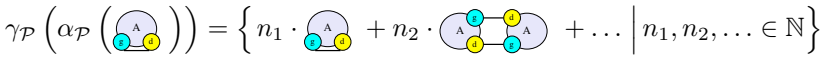
- $\alpha_{\mathcal{P}}\left(\left\{\right\}\right) = \left\{\right\}$
- $\alpha_{\mathcal{P}}\left(\left\{\right\}\right) = \left\{\right\}$
- $\alpha_{\mathcal{P}}\left(\left\{\right\}\right) = \left\{\right\}$
- $\alpha_{\mathcal{P}}\left(\left\{\right\}\right) = \left\{\right\}$
- $\alpha_{\mathcal{P}}\left(\left\{\right\}\right) = \left\{\right\}$

(c) Exemples de meilleures approximations.

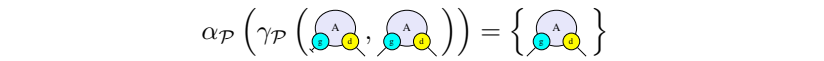
- $\gamma_{\mathcal{P}}\left(\left\{\right\}\right) = \left\{ n \cdot \;\middle|\; n \in \mathbb{N} \right\}$
- $\gamma_{\mathcal{P}}\left(\left\{\right\}\right) = \left\{ n_1 \cdot \; + n_2 \cdot \; + \ldots \;\middle|\; n_1, n_2, \ldots \in \mathbb{N} \right\}$
- $\gamma_{\mathcal{P}}\left(\left\{\right\}\right) = \{\emptyset\}.$
- $\gamma_{\mathcal{P}}\left(\left\{\right\}\right) = \{\emptyset\}.$
- $\gamma_{\mathcal{P}}\left(\left\{\right\}\right) = \left\{ n \cdot \;\middle|\; n \in \mathbb{N} \right\}.$
- $\gamma_{\mathcal{P}}\left(\left\{\right\}\right) = \left\{ \begin{array}{l} n_2 \cdot \\ +n_3 \cdot \; + \ldots \\ +n'_1 \cdot \; + n'_2 \cdot \; + \ldots \end{array} \;\middle|\; \begin{array}{l} n_2, n_3, \ldots \in \mathbb{N} \\ n'_1, n'_2, \ldots \in \mathbb{N} \end{array} \right\}.$

(d) Exemples de concrétisations

Figure 4.3: Un exemple jouet pour mieux comprendre le comportement des fonctions d'abstraction et de concrétisation. En 4.3(a), la signature du modèle : une seule sorte de protéines, $A$, avec deux sites pouvant être libres ou liés à l'autre site de la même ou d'une autre occurrence de la protéine $A$. En 4.3(b), le domaine abstrait est formé des vues locales de l'unique sorte de protéines : toutes les configurations pour les occurrences de la protéine $A$ sont considérées selon que chaque site soit libre ou lié. En 4.3(c) sont donnés des exemples de meilleure approximation d'ensemble d'états. Cela consiste à collecter les vues locales qui peuvent apparaître dans ces états. Réciproquement, En 4.3(d) sont donnés des exemples de concrétisations d'ensembles de vues locales. Ceci consiste à recomposer l'ensemble des états qui ne contiennent aucune occurrence des vues locales manquantes. Dans le cas particulier des vues locales, cela revient à prendre en compte tous les états composés uniquement des vues locales mises à disposition, sachant que chaque vue peut être utilisée zéro, une ou plusieurs fois.



(a) Exemple d'application de la fonction $\gamma_{\mathcal{P}} \circ \alpha_{\mathcal{P}}$.



(b) Exemple d'application de la fonction $\alpha_{\mathcal{P}} \circ \gamma_{\mathcal{P}}$.

Figure 4.4: Suite de l'exemple donné en Fig. 4.3. Un exemple d'application de la composée de fonctions $\gamma_{\mathcal{P}} \circ \alpha_{\mathcal{P}}$ est montré en 4.4(a). Celui-ci montre que l'abstraction ne permet pas de distinguer des ensembles d'anneaux d'occurrences de la protéine $A$ et ce quels que soient leurs tailles et leurs nombres. En 4.4(b) donne un exemple d'application de la composée de fonctions $\alpha_{\mathcal{P}} \circ \gamma_{\mathcal{P}}$. Cette fonction calcule que la vue locale avec le site $g$ libre et le site $d$ lié ne peut pas apparaître dans une espèce biochimique qui ne contiendrait pas la vue avec le site $g$ lié et le site $d$ libre.

que $X$. Il est impossible ainsi de distinguer ces deux ensembles en terme de propriétés abstraites. En revanche, la composée $\alpha_{\mathcal{P}} \circ \gamma_{\mathcal{P}}$ témoigne d'une certaine combinatoire dans le domaine abstrait. Elle associe à chaque propriété abstraite, la plus petite propriété abstraite qui est satisfaite par le même ensemble d'états concrets. Appliquer cette composée permet donc de raffiner une propriété abstraite, par déduction, et ce sans perdre le moindre état concret.

**Exemple 4.2.5** *Appliquée à l'ensemble formé d'un seul état composé uniquement d'un anneau de taille* 1*, la composée de fonctions $\gamma_{\mathcal{P}} \circ \alpha_{\mathcal{P}}$ donne l'ensemble des états formés uniquement d'anneaux d'occurrences de la protéine A. En effet, la meilleure approximation d'un anneau de taille* 1*, est l'ensemble de vues locales composé uniquement de la vue dont les deux sites sont liés. Or, voir également en Fig. 4.3(d), la concrétisation de cet ensemble de vues locales est l'ensemble de tous les états formés uniquement d'anneaux. Ainsi le niveau d'abstraction ne permet de distinguer, ni le nombre d'occurrences, ni la taille des anneaux d'occurrences de la protéine A.*

**Exemple 4.2.6** *Appliquée à l'ensemble formé exactement des deux vues locales, la première avec le site g libre et le site d lié, la seconde avec les deux sites liés, la composée de fonctions $\alpha_{\mathcal{P}} \circ \gamma_{\mathcal{P}}$ retourne l'ensemble formé d'une seule vue locale, celle avec les deux sites liés. En effet, la première vue ne peut apparaître dans un état sans que celui-ci ne contienne une occurrence de la vue locale avec le site d libre et le site g lié. De ce fait, elle ne peut apparaître dans aucun état de la concrétisation de l'ensemble formé par ces deux vues locales et n'est donc pas un élément de la meilleure approximation de l'ensemble de ces états. Ainsi, un état abstrait peut contenir des motifs d'intérêt, qui ne peuvent apparaître dans des complexes biochimiques sans contenir des occurrences de motifs d'intérêt interdits. Retirer ces motifs ne change pas l'ensemble des états concrets qui satisfont la propriété abstraite, mais cette étape peut requérir un temps de calcul substantiel.*

## 4.3 Transferts de point-fixes

Le plus petit point fixe qui définit l'ensemble des espèces biochimiques accessibles dans un réseau réactionnel, pour un état initial donné, peut se calculer au niveau des propriétés abstraites grâce à la correspondance de Galois ($\alpha_{\mathcal{P}}$,$\gamma_{\mathcal{P}}$).

Pour cela, il faut tout d'abord construire la *contre-partie abstraite* de la fonction $\mathbb{F}$, qui agira, non pas sur des ensembles d'états concrets, mais directement sur les propriétés abstraites. Cette contre-partie abstraite se définit de manière systématique : il suffit, pour chaque propriété abstraite $X^\sharp$, de considérer l'ensemble des états concrets $\gamma_{\mathcal{P}}(X^\sharp)$ qui vérifient la propriété $X^\sharp$, puis d'appliquer la fonction $\mathbb{F}$ à cet ensemble et enfin d'appliquer à ce résultat la fonction $\alpha_{\mathcal{P}}$ pour en calculer la meilleure approximation. C'est même la manière correcte la plus précise de procéder : la fonction $\alpha_{\mathcal{P}} \circ \mathbb{F} \circ \gamma_{\mathcal{P}}$ est, en effet, la meilleure contre-partie abstraite de la fonction $\mathbb{F}$ [41]. Elle permet de déléguer le calcul des états accessibles au domaine abstrait en contre-partie d'une perte éventuelle de précision. Pour ce faire, il suffit de remarquer que la fonction $\alpha_{\mathcal{P}} \circ \mathbb{F} \circ \gamma_{\mathcal{P}}$ est croissante (comme composée de fonctions croissantes) et définie sur l'ensemble des parties d'un ensemble. Elle admet donc un plus petit point fixe qui sera noté $lfp\ (\alpha_{\mathcal{P}} \circ \mathbb{F} \circ \gamma_{\mathcal{P}})$. L'inclusion suivante : $lfp\ \mathbb{F} \subseteq \gamma_{\mathcal{P}}(lfp\ (\alpha_{\mathcal{P}} \circ \mathbb{F} \circ \gamma_{\mathcal{P}}))$ se prouve alors par induction [41]. Autrement dit le plus petit point fixe de la fonction $\alpha_{\mathcal{P}} \circ \mathbb{F} \circ \gamma_{\mathcal{P}}$ est une abstraction de l'ensemble des états accessibles du modèle considéré. C'est à dire que la propriété abstraite $lfp\ (\alpha_{\mathcal{P}} \circ \mathbb{F} \circ \gamma_{\mathcal{P}})$ est satisfaite par chaque état accessible du modèle.

Le calcul des itérations de la fonction $[\alpha_{\mathcal{P}} \circ \mathbb{F} \circ \gamma_{\mathcal{P}}]$ peut prendre beaucoup de temps. Il est possible d'ajuster le compromis entre précision et temps de calcul en remplaçant celle-ci par une fonction moins précise. En effet, pour toute fonction $\mathbb{F}^\sharp$ telle que $[\alpha_{\mathcal{P}} \circ \mathbb{F} \circ \gamma_{\mathcal{P}}](Y) \subseteq \mathbb{F}^\sharp(Y)$ pour tout ensemble de motifs $Y \subseteq \mathcal{P}$, l'inclusion $lfp\, \mathbb{F} \subseteq \gamma_{\mathcal{P}}(lfp\, (\mathbb{F}^\sharp))$ est également satisfaite [41].

Une telle fonction $\mathbb{F}^\sharp$ peut être dérivée à la main. Pour cela, il faut d'abord donner une définition plus explicite de la fonction $[\alpha_{\mathcal{P}} \circ \mathbb{F} \circ \gamma_{\mathcal{P}}]$. Appliquée à un sous-ensemble $Y \subseteq \mathcal{P}$ de motifs d'intérêt, cette fonction ajoute l'ensemble des nouveaux motifs d'intérêt qui peuvent apparaître dans un état accessible en une étape de réécriture à partir d'un état qui ne contient aucun motif de l'ensemble $\mathcal{P}$ qui ne serait pas dans l'ensemble de motifs $Y$. Or, une telle étape de réécriture est nécessairement induite par une règle-réaction, elle-même induite par une règle du modèle. Chaque nouveau motif $P$ doit donc apparaître dans le membre droit d'une règle-réaction et l'ensemble d'états singleton formé du membre gauche de cette règle-réaction ne doit contenir aucune occurrence de motifs de l'ensemble $\mathcal{P} \setminus Y$. Dans cette règle-réaction, l'occurrence du motif $P$ dont il est question et l'image du membre droit de la règle sous-jacente ont nécessairement au moins une occurrence de protéines en commun (sinon le motif $P$ apparaitrait également dans le membre gauche de la règle-réaction et ne serait dont pas un nouveau motif). Il est alors possible de fixer le motif $P$ au membre droit de cette règle, en unifiant les occurrences de protéines du motif $P$ et de la règle du modèle qui sont communes dans la règle-réaction. Ceci forme alors un raffinement du membre droit de la règle. Un raffinement de la règle peut alors être construit en ajoutant toute information présente dans le motif $P$ qui n'est pas déjà présente dans le membre droit initial de la règle, dans le membre gauche de la règle. Le résultat est une spécialisation de la règle à la production du motif $P$ à cette position particulière. Par construction, le membre gauche de la règle raffinée apparait dans un état dans la concrétisation de $Y$.

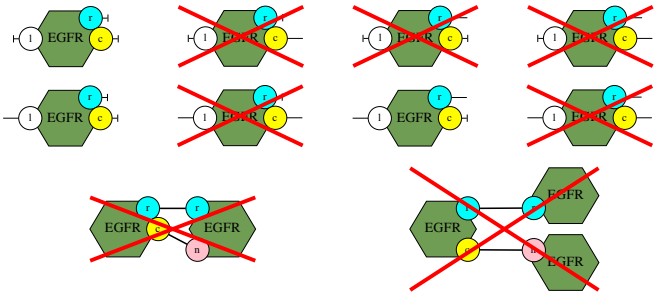
Ainsi, pour calculer les nouveaux motifs d'intérêt de l'ensemble $[\alpha_{\mathcal{P}} \circ \mathbb{F} \circ \gamma_{\mathcal{P}}](Y)$, il suffit de calculer tous les *chevauchements* possibles entre un nouveau motif d'intérêt potentiel (dans $\mathcal{P} \setminus Y$) et un membre droit d'une règle du modèle. Chaque chevauchement induit un raffinement de la règle correspondante. Si le membre gauche de la règle raffinée apparait dans un état de l'ensemble $\gamma_{\mathcal{P}}(Y)$, alors ce motif appartient bien à l'ensemble $[\alpha_{\mathcal{P}} \circ \mathbb{F} \circ \gamma_{\mathcal{P}}]Y$.

**Exemple 4.3.1** *Un exemple de cette construction est dessiné en Fig. 4.5. L'état abstrait actuel est donné en Fig. 4.5(a). Seules trois vues locales sont pour l'instant autorisées, celle avec aucun des sites l, r, c lié, celle avec seul le site l lié et celle avec seuls les sites l et r liés. Par ailleurs, ni les doubles liaisons, ni les chaînes d'au moins trois récepteurs ne sont autorisées à cet instant de l'analyse. La preuve que l'on peut construire une occurrence de protéines avec les trois sites l, r et c liés, est donnée en Fig. 4.5(b). Il suffit d'identifier cette vue locale à l'occurrence gauche du récepteur dans le membre droit de la règle (les plongements en pointillés bleus et marrons envoient ces deux occurrences de protéines sur une même occurrence de protéines) qui permet d'établir une liaison asymétrique entre deux récepteurs membranaires (voir en Fig. 3.1(e)). Ceci est possible car les sites en commun dans ces deux occurrences de protéines sont dans un état compatible : en effet, la vue locale demande que ces sites soient liés, alors que le membre droit de la règle précise à quels sites ils le sont. La règle est alors spécialisée à la production de la nouvelle vue locale pour ce chevauchement particulier entre la vue locale et le membre droit de la règle. Cela consiste à ajouter dans les membres gauche et droit de la règle l'information que le site l est lié. Pour conclure, il suffit alors d'exhiber un plongement (ici dessiné en pointillés rouges) en le membre gauche de la règle ainsi raffinée et une espèce biochimique, en vérifiant que cette espèce biochimique ne contient aucune occurrence des motifs d'intérêt non encore autorisés. Le dimer avec uniquement une liaison symétrique et dans lequel les deux occurrences du récepteur membranaire sont toutes deux liées à des occurrences du ligand et les sites Y48 et Y68 non phosphorylés et libres rempli parfaitement ces conditions. Ainsi la vue locale dans laquelle les trois sites l, r et c sont liés simultanément sera utilisable dès la prochaine itération de l'analyse.*

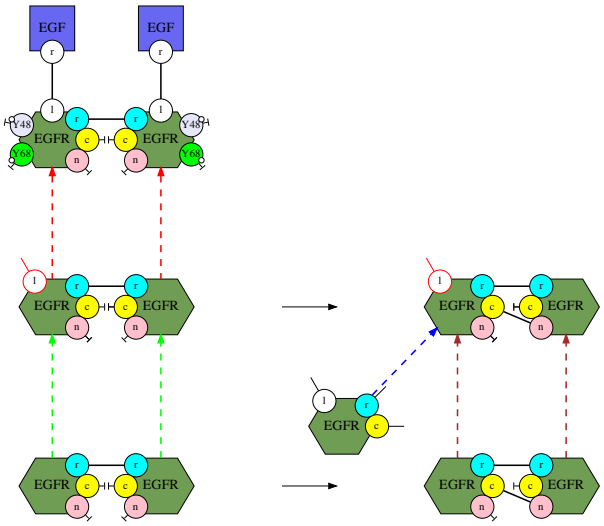
Lors du calcul de l'ensemble $[\alpha_{\mathcal{P}} \circ \mathbb{F} \circ \gamma_{\mathcal{P}}](Y)$, l'étape la plus coûteuse en temps de calcul est de vérifier que les membres gauches des règles raffinées peuvent apparaître dans un état de l'ensemble $\gamma_{\mathcal{P}}(Y)$. La section suivante a pour but de réduire ce coût moyennant une approximation supplémentaire.

## 4.4 Analyse par ensembles de motifs orthogonaux

Ajouter des hypothèses sur l'ensemble des motifs d'intérêt et simplifier le test de réalisabilité du membre gauche des raffinements de règles en le remplaçant par une condition nécessaire, mais pas toujours suffisante, permet




(a) Un état abstrait.




(b) La construction de la vue locale dans laquelle les trois sites sont liés peut être construite en une étape à partir de cet état abstrait.

Figure 4.5: Découverte d'un nouveau motif d'intérêt accessible dans le modèle des premières étapes de la voie de l'acquisition du facteur de croissance de d'épiderme (voir en Fig. 3.1 page 20). En 4.5(a), il est supposé qu'à ce moment de l'analyse, seules trois vues locales sont autorisées. Par ailleurs, il n'est permis de former ni des doubles liaisons entre récepteurs membranaires, ni des chaînes de trois récepteurs ou plus. En 4.5(b), la preuve que la vue locale doit être déclarée accessible à ce niveau d'abstraction est représentée sous forme de diagramme. Elle consiste à appliquer la règle de liaison asymétrique en identifiant la vue locale au récepteur de gauche dans le membre droit de la règle et en raffinant la règle en conséquence. Le membre gauche de la règle obtenue apparaît dans une espèce biochimique ne contenant aucun motif d'intérêt non encore découvert, ce qui conclut la preuve.

de rendre ce calcul plus efficace au prix d'une perte de précision de l'analyse. Ceci permet de définir une approximation correcte de la fonction $\alpha_{\mathcal{P}} \circ \mathbb{F} \circ \gamma_{\mathcal{P}}$.

### 4.4.1 Ensembles de motifs orthogonaux

Pour ce faire, l'ensemble des motifs d'intérêt peut être organisé sous la forme d'un ensemble fini d'ensembles finis de motifs orthogonaux [67]. Chaque *ensemble de motifs orthogonaux* est un arbre de décision raffinant progressivement un motif initial, dans le but de répondre à une question spécifique. Un ensemble de motifs orthogonaux est construit de manière à ce que toute occurrence du motif initial dans une espèce biochimique, puisse être complétée en exactement une occurrence d'un de ces raffinements. En conséquence, les raffinements du motif initial sont deux à deux incompatibles et ils recouvrent, en quelque sorte, tous les cas possibles pour le motif initial.

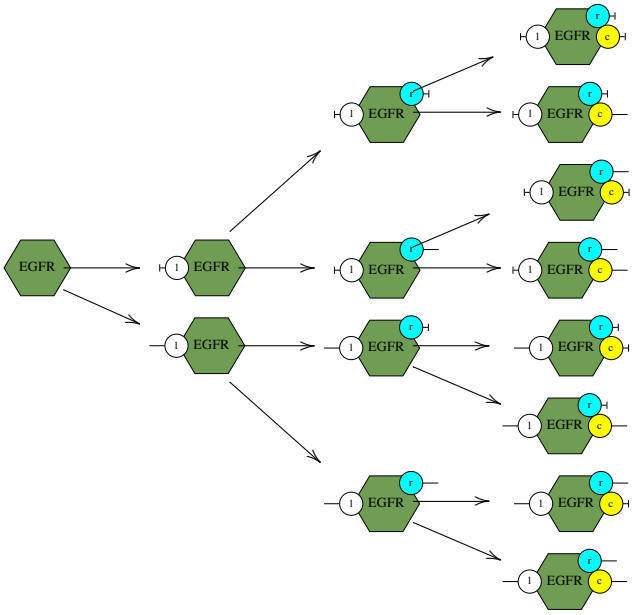
Le choix exact des ensembles de motifs orthogonaux repose sur une analyse préliminaire qui calcule, par inspection des règles du modèle, quelles questions intéressantes se posent. Trois catégories de questions sont considérées par défaut dans l'analyseur KaSa (mais il est possible de paramétrer l'analyse pour en désactiver une ou deux). La première infère des relations entre les états des différents sites de chaque type de protéines, cela correspond à l'analyse des vues locales [50]. La seconde permet de détecter des relations entre l'état des sites dans des occurrences de protéines qui partagent un lien [67] dans le but d'analyser les déplacements de complexes biochimiques lorsque ceux-ci sont codés par des transformations de l'état d'activation de sites fictifs. L'analyse permet alors de vérifier si oui ou non deux occurrences de protéines sont toujours localisées dans le même compartiment quand elles sont liées entre elles. La troisième permet de détecter si une même occurrence de protéines peut être liée simultanément à deux occurrences différentes de protéines ou si une même occurrence de protéines peut être liée au moins doublement à une autre occurrence de protéines [67]. Une quatrième sorte d'ensembles de motifs orthogonaux est en cours d'implantation. Elle se concentre sur la formation de complexes biochimiques cycliques : son but est de prouver l'absence de complexes biochimiques de taille non bornée [19].

Les ensembles finis de motifs orthogonaux peuvent être construits récursivement, en remplaçant un des motifs par plusieurs motifs le raffinant. Il suffit de choisir une information non spécifiée dans ce motif et de considérer tous les cas possibles pour cette information, d'où la représentation sous forme d'arbre de décision. L'ensemble de motifs orthogonaux est alors formé par les feuilles de cet arbre, alors que les nœuds de cet arbre représentent les motifs intermédiaires qui ont été remplacés par des motifs plus précis.

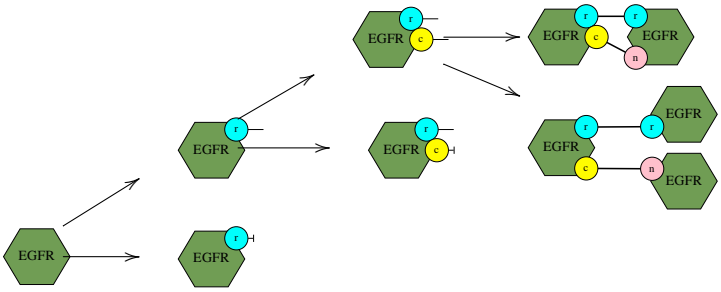
**Exemple 4.4.1** *L'ensemble des motifs d'intérêt introduit en Fig. 4.2 est inclus dans la réunion de deux ensembles de motifs orthogonaux. En effet, l'ensemble des vues locales peut être obtenu, en partant d'une occurrence de la protéine EGFR sans aucun site, en se demandant successivement si le site l est libre ou non, si le site r est libre ou non et si le site c est libre ou non. L'arbre de décision correspondant se trouve en Fig. 4.6(a). Les deux derniers motifs d'intérêt sont obtenus en se demandant si un récepteur peut établir des liaisons doubles. Partant d'une occurrence de la protéine EGFR sans aucun site, il faut se demander si le site r est libre ou non, puis dans le cas où le site r est lié, si le site c est lié ou non, et enfin, dans le cas où le site c est également lié, si ces deux sites sont liés à une même occurrence de récepteur membranaire ou à deux occurrences différentes. L'arbre de décision ainsi obtenu est donné en Fig. 4.6(b).*

### 4.4.2 Pas de calcul abstraits

Les différents ensembles de motifs orthogonaux collaborent au sein de l'analyse, qui effectue ainsi une induction mutuelle sur ces derniers. Ceci présente deux avantages par rapport à des analyses séparées ou en cascades (où chacune utiliserait le résultat des analyses précédentes). D'une part, il n'est pas nécessaire de définir quel ensemble de motifs orthogonaux doit être analysé avant quel autre. D'autre part, une induction mutuelle est strictement plus expressive. La collaboration entre les différents ensembles de motifs orthogonaux se produit lors du test de la nouvelle condition utilisée pour prouver que le membre gauche des règles raffinées n'est pas réalisable étant donné les motifs qui sont autorisés à un moment donné de l'analyse (voir en Fig. 4.5). Pour faire cette preuve, le raffinement d'une règle est construit de la manière habituelle. Il suffit ensuite de trouver une occurrence de protéines dans le membre gauche de la règle raffinée qui soit incompatible avec l'état actuel de l'analyse sur au moins un des ensembles de motifs orthogonaux pris en paramètre de l'analyse. Pour cela, la racine de l'ensemble de motifs orthogonaux doit être de la même sorte que l'occurrence de la protéine en question et l'information contextuelle de cette occurrence de protéines dans ce membre gauche de la règle raffinée ne doit être compatible avec aucun des motifs de cet ensemble de motifs orthogonaux déjà déjà déclarés potentiellement




(a) Ensemble de motifs orthogonaux pour les vues locales.




(b) Ensemble de motifs orthogonaux pour discuter des doubles liaisons.

Figure 4.6: Deux exemples d'ensemble de motifs orthogonaux. En 4.6(a), l'ensemble des vues locales (voir les huit premiers motifs en Fig. 4.2 page 27) sous forme d'arbre de décision. En 4.6(b), celui pour discuter de la présence potentielle de doubles liaisons entre des récepteurs et de la présence potentielle de trimers (voir les deux derniers motifs en Fig. 4.2).

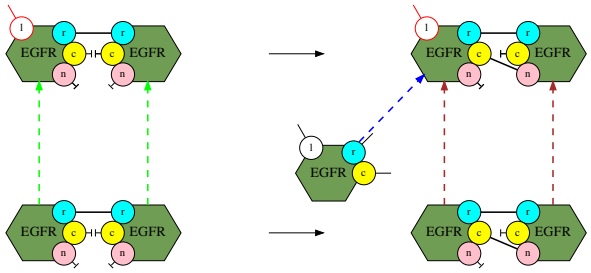



Figure 4.7: L'étape abstraite de la Fig. 4.5 revisitée avec la procédure de décision approchée. Au lieu de vérifier que chaque motif connexe du membre gauche de la règle raffinée se plonge dans un complexe biochimique dans la concrétisation de l'état abstrait, il suffit de s'assurer, pour chaque occurrence de protéines dans ce motif et chaque ensemble de motifs orthogonaux portant sur ce type de protéines si il contient un motif compatible déjà découvert par l'analyse.

accessibles par l'analyse. Dans le cas contraire, l'analyseur ne peut pas prouver que le motif est inaccessible. Le motif est alors considéré comme potentiellement accessible pour la suite de l'analyse. Il s'agit bien entendu d'une approximation.

**Exemple 4.4.2** *En Fig. 4.7, l'étape de calcul qui avait été décrite en Fig. 4.5 est rejouée en remplaçant le test de réalisabilité par cette procédure approchée. Au lieu de construire un plongement du membre gauche de la règle raffinée vers un complexe biochimique afin de vérifier qu'il ne contient pas de motif non encore autorisé, la nouvelle procédure se contente de vérifier pour chaque occurrence de protéines dans le membre gauche de la règle raffinée et pour chaque ensemble de motifs orthogonaux portant sur cette sorte de protéines si celui-ci contient un motif autorisé compatible avec cette occurrence. Dans ce cas, cela revient à vérifier qu'il existe bien une vue locale déjà autorisée dans laquelle les deux sites l et r sont liés, alors que le site c est libre et que le motif dans lequel le site r est lié et le site c est libre est autorisé dans le deuxième ensemble de motifs orthogonaux.*

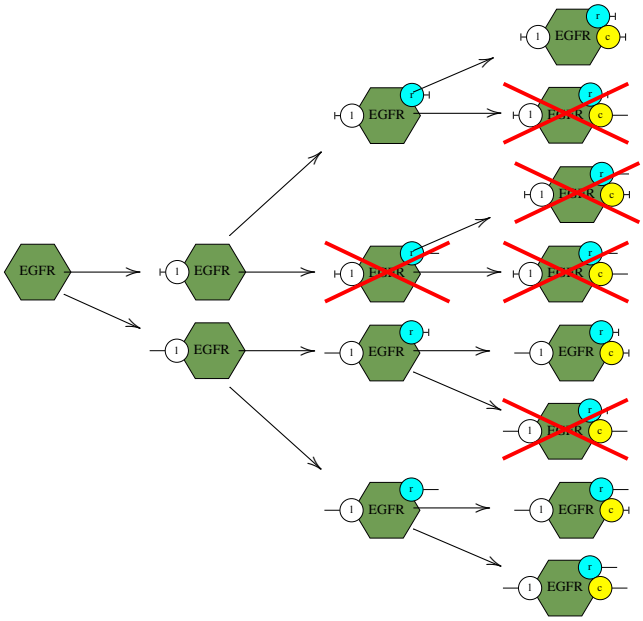
Outre le fait de ne pas vérifier l'existence d'un complexe biochimique qui pourrait compléter le collage obtenu entre les motifs connexes du membre gauche de la règle raffinée et les motifs déjà déclarés potentiellement accessibles par l'analyse, il est intéressant de remarquer que la procédure de décision approchée évite le calcul de tous les chevauchements entre les motifs d'intérêt non encore découverts par l'analyse, en se focalisant sur la racine de chaque ensemble de motifs orthogonaux. Ce sont les deux sources de pertes d'information dues à l'affaiblissement de la procédure de décision.

**Exemple 4.4.3** *Le résultat de l'itération pour le modèle formé des règles qui avaient été décrites en Fig. 3.1 pour les ensembles de motifs orthogonaux qui avaient été introduits en Fig. 4.6, est donné en Fig. 4.8. Cette itération a été initialisée avec une quantité arbitraire d'occurrences de protéines de chaque sorte, mais avec tous leurs sites libres. Pour ce qui est des vues locales (voir en Fig. 4.8(a)), seules 4 configurations sont possibles pour l'état des sites l, r, et c des récepteurs membranaires. Ainsi, le site c ne peut être lié sans que le site r ne le soit et le site r ne peut être lié sans que le site l ne le soit. De son côté, l'analyse des doubles liaisons (voir en Fig. 4.8(b)) montre qu'il est impossible de former des chaînes d'au moins trois récepteurs membranaires.*

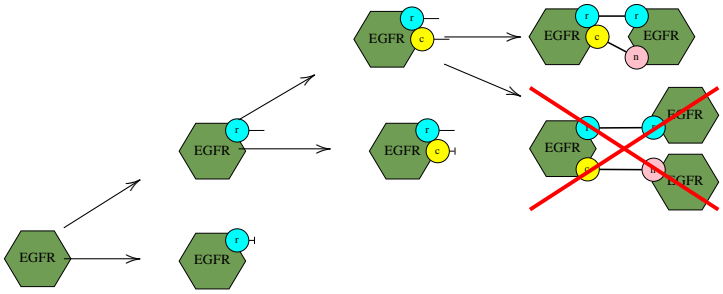
Il est important de rappeler que l'analyse ne donne qu'une sur-approximation des états accessibles. Ainsi, tout motif prouvé comme non accessible est bien inaccessible. Par contre, il n'y a aucune garantie qu'un motif non prouvé inaccessible puisse apparaître dans un état accessible depuis un des états initiaux.

### 4.4.3 Post-traitement et visualisation des résultats

L'itération de point-fixe est suivie d'une phase de traitement du résultat. Le but est essentiellement de rendre le résultat de l'analyse plus compréhensible pour l'utilisateur. Dans un premier temps, un parcours de chaque arbre de décision est effectué et chaque nœud dont tous les fils sont déclarés inaccessibles est déclaré inaccessible lui-aussi. Ensuite, tous les nœuds des arbres de décision sont explorés en répertoriant ceux dont les enfants n'ont pas tous le même statut. Ceci témoigne d'une propriété intéressante puisque dans ce cas, un des raffinements d'un motif accessible n'est pas accessible. Cette information est alors présentée sous la forme d'une implication, appelée *lemme de raffinement*, entre un motif (le nœud en question) et une liste de motifs (ses fils qui n'ont




(a) Analyse d'accessibilité pour l'ensemble des vues locales (voir en Fig. 4.6(a)).

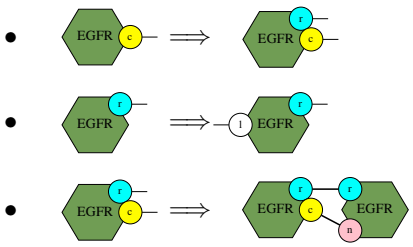



(b) Analyse d'accessibilité pour discuter de la présence éventuelle de trimers et de double liaisons.

Figure 4.8: Résultat de l'analyse pour les deux ensembles de motifs orthogonaux donnés en Fig. 4.6. Les motifs orthogonaux sont aux feuilles des arbres de décision. Ceux qui sont barrés en rouge n'apparaissent dans aucune exécution du modèle (pour n'importe quel état initial sans lien). Par construction de l'arbre de décision, les nœuds dont tous les enfants sont inaccessibles sont également inaccessibles et donc barrés eux-aussi.

pas été prouvés inaccessibles). Une telle implication s'interprète de la manière suivante : chaque occurrence du motif de la précondition dans un complexe biochimique accessible peut toujours se raffiner en l'un des motifs de la postcondition.

**Exemple 4.4.4** *Le résultat de l'analyse décrit en Fig. 4.8 donne lieu aux implications suivantes :*




*Cela prouve que dans une occurrence du récepteur membranaire, le site c ne peut être lié sans que le site r ne le soit également, et que le site r ne peut être lié sans que le site l ne le soit aussi. De plus, une occurrence du récepteur dont les sites r et c sont tous deux liés, est nécessairement liée doublement à une même occurrence du récepteur.*

Par ailleurs, l'analyseur vérifie pour chaque règle si son membre gauche est compatible avec le résultat de l'analyse (avec la procédure de décision simplifiée présentée Sec. 4.4.1). Les règles pour lesquelles ce n'est pas le cas sont reportées à l'utilisateur.

## 4.5 Étude de performance et utilisation concrète

Nous avons utilisé l'analyseur statique KaSa sur plusieurs modèles. Les résultats de ces analyses sont décrites en Fig. 4.9. Les onze premiers modèles sont des traductions directes des modèles qui sont fournis avec la distribution du logiciel BNGL [10]. Le modèle 'sos' décrit les premiers événements de l'acquisition du facteur de croissance de l'épiderme (il comprend entre autres les règles données en Fig. 3.1). Les modèles 'machine' et 'ensemble' sont deux versions de la voie de signalisation MAPK, publiées par Eric Deeds et Ryan Suderman [109]. Les versions du modèle 'korkut' concernent la voie de signalisation de la protéine Ras. Ce modèle a été conçu par John Bachman et Benjamin Gyori (Sorger lab) dans le cadre du projet DARPA Big Mechanism [32] en utilisant des outils de traitement automatique du langage naturel pour extraire des faits de la littérature, en assemblant ces faits en Kappa, puis en corrigeant manuellement le modèle obtenu [76]. Le modèle 'tgf' s'intéresse lui à la matrice extra-cellulaire de la protéine tgf-$\beta$. Cinq versions de ce modèle ont été analysées. Elles ont été assemblées à la main d'après la littérature par Nathalie Théret et Jean Coquet, puis corrigées avec l'aide de l'analyseur KaSa. Enfin, plusieurs versions du modèle de la voie de signalisation de la protéine Wnt, écrites par Héctor F. Medina Abarca (Fontana Lab) dans le cadre du projet DARPA Big Mechanism [32] ont été analysées. Ce modèle a également été assemblé manuellement après lecture humaine de la littérature. Dans ce dernier modèle, le grand nombre de règles vient du fait que des scripts ont été utilisés pour raffiner des règles d'interaction génériques afin d'ajuster leur cinétique en fonction d'information contextuelle sur l'état des occurrences de protéines qui interagissent et de leurs voisines.

Pour chaque modèle et chaque version de ce modèle, nous avons reporté le nombre total de règles dans le modèle, ainsi que nombre d'implications découvertes par l'analyse portant sur des relations soit entre au moins deux sites, soit entre les états de liaison et d'activation d'un même site. Nous avons également donné le nombre de règles qui ont été trouvées mortes par l'analyseur statique (du fait de l'approximation, l'analyseur peut manquer des règles mortes, par contre, toute règle détectée morte l'est). Nous donnons également le temps de calcul total de l'analyse, ce qui montre que KaSa passe à l'échelle même sur des modèles comportant un grand nombre de règles d'interaction.

Les informations trouvées par l'analyse statique sont utiles. Une même démarche a été suivie pour améliorer la qualité de ces modèles, qu'ils soient écrits à la main ou assemblés automatiquement par fouille automatique de la littérature. La première étape est la vérification des règles mortes. Ces règles sont souvent la conséquence, soit d'erreurs typographiques, soit d'états initiaux incomplets, soit de règles manquantes, soit de relations de causalité qui ne peuvent pas être satisfaites. La lecture des contraintes trouvées par l'analyseur permet de mieux comprendre leur origine. Elle permet également de vérifier que les invariants structurels auquel le modélisateur peut s'attendre sont bien vérifiés. L'étape suivante est d'étudier comment une occurrence de protéines peut

| modèle | nombre de règles | nombre de contraintes inférées | nombre de règles mortes détectées | temps d'analyse (secondes) |
|---|---|---|---|---|
| repressilator | 42 | 0 | 0 | 0.009 |
| egfr_net | 39 | 0 | 0 | 0.033 |
| egfr_net_red | 45 | 0 | 0 | 0.034 |
| fceri_fyn | 46 | 0 | 0 | 0.060 |
| fceri_fyn_lig | 48 | 0 | 0 | 0.064 |
| fceri_fyn_trimer | 362 | 0 | 36 | 0.496 |
| fceri_fyn_gamma2 | 59 | 0 | 0 | 0.095 |
| fceri_fyn_ji | 36 | 0 | 0 | 0.047 |
| fceri_fyn_ji_red | 32 | 0 | 0 | 0.042 |
| fceri_fyn_lyn_745 | 40 | 0 | 2 | 0.051 |
| fceri_fyn_trimer | 192 | 0 | 0 | 0.260 |
| sos | 20 | 0 | 0 | 0.016 |
| machine | 220 | 0 | 7 | 0.663 |
| ensemble | 233 | 0 | 0 | 0.593 |
| korkut (2017/01/13) | 3916 | 0 | 1610 | 13 |
| korkut (2017/01/17) | 12896 | 0 | 874 | 48 |
| korkut (2017/02/06) | 5750 | 0 | 884 | 90 |
| TGF (V19) | 97 | 0 | 10 | 0.368 |
| TGF (V20) | 99 | 0 | 10 | 0.490 |
| TGF (V21) | 211 | 0 | 0 | 1.05 |
| TGF (2017/04/01) | 235 | 0 | 0 | 0.875 |
| TGF (2018/04/19) | 292 | 0 | 0 | 1.13 |
| BigWnt (2015/12/28) | 356 | 0 | 1 | 2.66 |
| BigWnt (2016/09/28) | 1419 | 0 | 0 | 10 |
| BigWnt (2017/03/22) | 1486 | 0 | 12 | 13 |

Figure 4.9: Résultats expérimentaux (calculés sur un MacBook Pro avec une puce Intel Core i7-6567U (cadencée 3.3 GHz)). Pour chaque modèle et chaque version, le nombre de règles est donné, ainsi que le nombre de contraintes découvertes par l'analyse et le nombre de règles mortes trouvées (qui ne sont donc jamais utilisées dans le modèle). Le temps total de l'analyse est également fourni.

passer d'une configuration à une autre. L'analyse des *traces locales* [66, 59] calcule des systèmes de transitions à partir des vues locales. Ceci permet d'avoir une cartographie des changements de configuration de chaque occurrence de protéines en faisant abstraction de l'état des occurrences de protéines auxquelles cette occurrence est liée. En particulier, une étude de ces systèmes de transitions permet de calculer efficacement des transitions qui sont définitives : c'est à dire celles qui transforment la configuration d'une occurrence de protéines, sans retour possible, quel que soit le nombre de transitions ultérieures.

**Exemple 4.5.1** *Le système de transitions qui représente les traces locales des occurrences du récepteur membranaire dans le modèles des premières interactions qui interviennent dans l'activation du facteur de croissance de l'épiderme, est dessiné en Fig. 4.10. La succession entre les différentes configurations possibles y est clairement décrite. Ainsi, partant d'un récepteur avec les sites l, r, c et n libres (le site n a été ajouté pour rendre l'exemple plus intéressant), le site l peut devenir lié en premier, ensuite le site r peut devenir lié, ensuite soit le site c, soit le site n peut devenir lié. Par contre, ces deux liaisons sont exclusives : les sites c et n d'une occurrence de protéines ne peuvent pas être liés tous les deux simultanément. Par ailleurs, toutes les liaisons peuvent se défaire dans l'ordre inverse de leur création.*

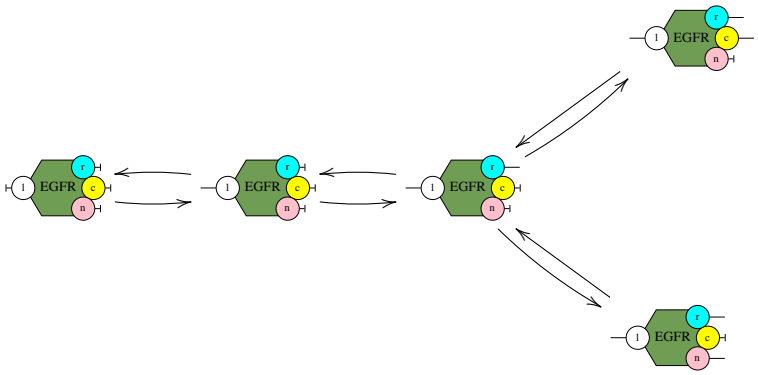



Figure 4.10: Le système de transitions pour les vues locales des occurrences du récepteur membranaire. L'état du site $n$ a été ajouté pour rendre l'exemple plus intéressant. Ce système de transitions explique les étapes que traverse ces occurrences lorsque leurs sites deviennent liés. Les transitions en sens inverse, qui correspondent aux règles de libération des sites sont aussi représentées.

# Chapter 5

# Flot d'information dans la sémantique stochastique d'un modèle Kappa

## 5.1 Système stochastique sous-jacent à un modèle Kappa

## 5.2 Cas d'études

Les modèles d'interaction entre protéines souffrent d'une très grande complexité combinatoire. Les protéines peuvent se lier entre elles et modifier leurs états, ce qui conduit potentiellement à la formation de très grands complexes biochimiques. Ceci empêche toute description extensionnelle des systèmes dynamiques engendré par ces modèles, que ce soit dans un cadre différentiel ou stochastique.

Le chapitre précédent portait sur une approche qui permettait de réduire le nombre de variables nécessaire pour décrire la dynamique de ces modèles dans le cadre différentiel. En suivant le flot d'information potentiel entre les différentes parties des espèces biochimiques, elle consiste à identifier quelles corrélations entre états de sites d'interaction ont une importance sur la dynamique du système, puis à oublier les autres en découpant les espèces biochimiques en petite unité d'information, appelées fragments. Alors qu'elle donne des résultats prometteurs dans le cadre différentiel, le présent chapitre montre au contraire qu'il est beaucoup plus difficile de réduire la dimension des modèles stochastiques engendrés par un modèle de réécriture de graphes à sites.

Ce chapitre commence par la description formelle de la sémantique stochastique d'un modèle Kappa. En Sect. **??** sont fournis trois exemples pour illustrer en quoi l'approche décrite dans le chapitre **??** pour la réduction des systèmes différentiels engendrés ne peut pas s'appliquer directement pour réduire les systèmes stochastiques sous-jacents. En Sect. **??** ...

## 5.3 Sémantique stochastique d'un modèle de réécritures de graphes à sites

## 5.4 Cas d'étude

Trois cas d'études sont considérés. Le premier, qui est décrit en Sect. 5.4.1, fait intervenir une sorte de protéine pouvant se lier indépendament à deux autres sortes de protéines. La première sorte de protéine peut alors être vue comme deux morceaux entièrement indépendant, ce qui conduit à une réduction du modèle correcte à la fois dans le cadre différentiel ou stochastique. Dans le second exemple, qui est donné en Sect. 5.4.2, il est montré que, par rapport à la réduction du système différentiel engendré par un modèle de réécriture de graphes à sites, des corrélations supplémentaires sont souvent nécessaires pour simuler fidèlement les systèmes stochastiques. Dans cette exemple, une règle de déliaison ne peut être simplifiée sans connaître précisément la distribution des différents états des deux protéines qu'elles séparent pour préserver la dynamique du système sous-jacent dans le cadre stochastique. Dans le cadre différentiel, cette information peut être oubliée tout en obtenant une réduction exacte du système engendré. Enfin, le troisième exemple, qui est expliqué en Sect. 5.4.3, montre l'existence de flot d'information entre différentes composantes connexes du membre gauche d'une règle, ce qui

une fois encore interdit certaines réductions qui sont pourtant correctes dans le cadre différentiel.

### 5.4.1 Un exemple d'indépendance entre deux liaisons

Voici un modèle dans lequel chaque occurrence d'une protéine peut être découpée en deux morceaux indépendants. La dimension de l'espace d'états de la chaîne de Markov sous-jacente peut alors être réduite en conséquence.

Soit une protéine $B$ munie de deux sites de liaison $a$ et $c$. Dans chaque occurrence de la protéine $B$, le site $a$ peut se lier au site $b$ des occurrences d'une autre protéine, appelée $A$, alors que le $c$ peut se liet au site $b$ des occurrences d'une troisième protéine, appelée $C$. Les règles de liaison et de déliaison sont dessinées en Fig. **??**.

$$A + B \underset{k_d^{AB}}{\overset{k^{AB}}{\rightleftharpoons}} AB \qquad B + C \underset{k_d^{BC}}{\overset{k^{BC}}{\rightleftharpoons}} BC$$

$$A + BC \underset{k_d^{AB}}{\overset{k^{AB}}{\rightleftharpoons}} ABC \qquad AB + C \underset{k_d^{BC}}{\overset{k^{BC}}{\rightleftharpoons}} ABC$$

Les constantes de réactions pour les règles d'association et de dissociation ne dépendent pas du fait que l'occurrence de la protéine $B$ soit liée, ou non, sur son autre site de liaison. Ce point est essentiel pour justifier la correction de la réduction du modèle qui va suivre.

Il est tentant d'oublier la corrélation éventuelle entre l'état de liaison des deux sites des occurrences de la protéine $B$ dans le système stochastique engendré par les règles de cet exemple ? Pour cela, il faut regarder ce qui se passe au niveau de l'évolution de la distribution des différents états du système. L'état du système peut être representer par un sextuplet d'entiers naturels $\langle n_A, n_B, n_C, n_{AB}, n_{BC}, n_{ABC}\rangle$ où la composante $n_X$ représente le nombre d'occurrences du complexe biochimique $X$ dans cet état, pour n'importe quel complexe biochimique $X$ du modèle. Le modèle ayant peu de complexes biochimiques différents. Ceux-ci sont représentés par la liste des occurrences des protéines qui les constituent. Ainsi, les notations A, B et C représentent les protéines correspondantes, alors que les notations $AB$, $BC$, $ABC$ représentent les complexes formés d'une occurrence de la protéine $B$ liée à une occurrence de la protéine $A$, d'une occurrence de la protéine $B$ liée à une occurrence de la protéine $C$ et d'une occurrence de la protéine $B$ liée à une occurrence de la protéine $A$ et à une occurrence de la protéine $C$. La probabilité $P_t(\sigma)$ que le système soit dans l'état $\sigma$ au temps $t$ est alors donnée par l'équation maîtresse suivante :

$$\begin{aligned}
&P_t(\langle n_A, n_B, n_C, n_{AB}, n_{BC}, n_{ABC}\rangle)' = \\
&\quad k^{AB}(n_A+1)(n_B+1)P_t(\langle n_A{+}1, n_B{+}1, n_C, n_{AB}{-}1, n_{BC}, n_{ABC}\rangle) \\
+\;& k_d^{AB}(n_{AB}+1)P_t(\langle n_A{-}1, n_B{-}1, n_C, n_{AB}{+}1, n_{BC}, n_{ABC}\rangle) \\
+\;& k^{AB}(n_A+1)(n_{BC}+1)P_t(\langle n_A{+}1, n_B, n_C, n_{AB}, n_{BC}{+}1, n_{ABC}{-}1\rangle) \\
+\;& k_d^{AB}(n_{ABC}+1)P_t(\langle n_A{-}1, n_B, n_C, n_{AB}, n_{BC}{-}1, n_{ABC}{+}1\rangle) \\
+\;& k^{BC}(n_B+1)(n_C+1)P_t(\langle n_A, n_B{+}1, n_C{+}1, n_{AB}, n_{BC}{-}1, n_{ABC}\rangle) \\
+\;& k_d^{BC}(n_{BC}+1)P_t(\langle n_A, n_B{-}1, n_C{-}1, n_{AB}, n_{BC}{+}1, n_{ABC}\rangle) \\
+\;& k^{BC}(n_{AB}+1)(n_C+1)P_t(\langle n_A, n_B, n_C{+}1, n_{AB}{+}1, n_{BC}, n_{ABC}{-}1\rangle) \\
+\;& k_d^{BC}(n_{ABC}+1)P_t(\langle n_A, n_B, n_C{-}1, n_{AB}{-}1, n_{BC}, n_{ABC}{+}1\rangle) \\
-\;& k^{AB}n_A(n_B+n_{BC})P_t(\langle n_A, n_B, n_C, n_{AB}, n_{BC}, n_{ABC}\rangle) \\
-\;& k_d^{AB}(n_{AB}+n_{ABC})P_t(\langle n_A, n_B, n_C, n_{AB}, n_{BC}, n_{ABC}\rangle) \\
-\;& k^{BC}(n_B+n_{AB})n_CP_t(\langle n_A, n_B, n_C, n_{AB}, n_{BC}, n_{ABC}\rangle) \\
-\;& k_d^{BC}(n_{BC}+n_{ABC})P_t(\langle n_A, n_B, n_C, n_{AB}, n_{BC}, n_{ABC}\rangle)
\end{aligned}$$

Afin d'oublier la corrélation entre l'état des deux sites de liaisons des occurrences de la protéine $B$, chaque état $\sigma = \langle n_A, n_B, n_C, n_{AB}, n_{BC}, n_{ABC}\rangle$ est abstrait en deux triplets $\beta^A(\sigma) := \langle n_A, n_B + n_{BC}, n_{AB} + n_{ABC}\rangle$ et $\beta^C(\sigma) := \langle n_C, n_B + n_{AB}, n_{BC} + n_{ABC}\rangle$. Ces deux triplets projettent l'état du système $\sigma$ en passant sous silence l'état de liaison respectivement des sites $c$ et $a$ des occurrences de la protéine $B$. L'évolution temporelle de la distribution de ces deux projections est alors noté $P_t^A(\sigma^A)$ et $P_t^C(\sigma^C)$. Ainsi, la quantité $P_tA(\sigma^A)$ représente la probabilité que le système soit dans un état $\sigma$ tel que $\beta^A(\sigma) = \sigma^A$ à l'instant $t$, alors la quantité $P_tC(\sigma^C)$ représente la probabilité que le système soit dans un état $\sigma$ tel que $\beta^C(\sigma) = \sigma^C$ à l'instant $t$. Il est alors possible de vérifier analytiquement que l'évolution de deux distributions de probabilités $P_t^A(\sigma^A)$ et $P_t^C(\sigma^C)$ vérifient les

deux équations maîtresses suivantes :

$$
\begin{aligned}
&P_t^A(\langle n_A, n_{B?}, n_{AB?}\rangle)' = \\
&\qquad k^{AB}(n_A+1)(n_{B?}+1)P_t^A(\langle n_A{+}1, n_{B?}{+}1, n_{AB?}{-}1\rangle) \\
&+\quad k_d^{AB}(n_{AB?}+1)P_t^A(\langle n_A{-}1, n_{B?}{-}1, n_{AB?}{+}1\rangle) \\
&-\quad (k^{AB}n_An_{B?}+k_d^{AB}n_{AB?})P_t^A(\langle n_A, n_{B?}, n_{AB?}\rangle)
\end{aligned}
$$

$$
\begin{aligned}
&P_t^C(\langle n_C, n_{?B}, n_{?BC}\rangle)' = \\
&\qquad k^{BC}(n_{?B}+1)(n_C+1)P_t^C(\langle n_C{+}1, n_{?B}{+}1, n_{?BC}{-}1\rangle) \\
&+\quad k_d^{BC}(n_{?BC}+1)P_t^C(\langle n_C{-}1, n_{?B}{-}1, n_{?BC}{+}1\rangle) \\
&-\quad (k^{BC}n_{?B}n_C+k_d^{AB}n_{?BC})P_t^C(\langle n_C, n_{?B}, n_{?BC}\rangle).
\end{aligned}
$$

De ce fait, même dans le cadre stochastique, il n'est pas nécessaire de décrire l'état du site de liaison $c$ dans les occurrences de la protéine $B$ pour connaître celui du site $a$, et réciproquement.

### 5.4.2 Un exemple avec une déliaison inconditionnelle

Le fait que les occurrences d'un complexe biochimique puisse être séparées en deux parties entièrement indépendantes (comme c'était le cas dans l'exemple précédent) est assez rare en pratique. Ce nouvel exemple contient une règle de déliaison inconditionnelle entre deux occurrences de protéines pouvant prendre plusieurs états. Dans la sémantique différentielle, cette règle n'induit pas de flot d'information et le système d'équations différentielles sous-jacent peut être réduit. En stochastique, la corrélation entre l'état des occurrences de protéines qui sont liées entre-elles a un impact sur la distribution de l'état des monomères. Ceci induit un flux d'information qui empêche de réduire de manière exacte le système stochastique sous-jacent.

Soit $A$ et $B$ deux sortes de protéine. Chaque occurrence de ces protéines peut être phosphorylée, ou non. De plus, une occurrence de la protéine $A$ peut se lier à une occurrence de la protéine $B$ pour former un dimère.

Le comportement de ce modèle peut être décrit par les règles de réécriture données en Fig. **??**.

$$
\begin{array}{lcl}
\circ A\dashv & \underset{k_{A-}}{\overset{k_{A+}}{\rightleftharpoons}} & \bullet A\dashv \\
\circ AB\circ & \underset{k_{A-}}{\overset{k_{A+}}{\rightleftharpoons}} & \bullet AB\circ \\
\circ AB\bullet & \underset{k_{A-}}{\overset{k_{A+}}{\rightleftharpoons}} & \bullet AB\bullet
\end{array}
\qquad
\begin{array}{lcl}
\vdash B\circ & \underset{k_{B-}}{\overset{k_{B+}}{\rightleftharpoons}} & \vdash B\bullet \\
\circ AB\circ & \underset{k_{B-}}{\overset{k_{B+}}{\rightleftharpoons}} & \circ AB\bullet \\
\bullet AB\circ & \underset{k_{B-}}{\overset{k_{B+}}{\rightleftharpoons}} & \bullet AB\bullet
\end{array}
\qquad
\begin{array}{lcl}
\circ A\dashv + \vdash B\circ & \underset{k_{A..B}}{\overset{k_{AB}}{\rightleftharpoons}} & \circ AB\circ \\
\bullet A\dashv + \vdash B\circ & \underset{k_{A..B}}{\overset{k_{AB}}{\rightleftharpoons}} & \bullet AB\circ \\
\circ A\dashv + \vdash B\bullet & \underset{k_{A..B}}{\overset{k_{AB}}{\rightleftharpoons}} & \circ AB\bullet \\
\bullet A\dashv + \vdash B\bullet & \underset{k_{A..B}}{\overset{k_{AB\star}}{\rightleftharpoons}} & \bullet AB\bullet
\end{array}
$$

En particulier, il y a plusieurs constantes de réaction pour les règles de liaison (voir en Fig. **??**), selon l'état de phosphorylation des occurrences de protéines qui se lient.

Hormi cette règle de liaison, les interactions potentielles sont purement locales. Ainsi, les constantes de réaction pour la phosphorylation des occurrences de la protéine $A$ (voir en Fig. **??**) et de la protéine $B$ (voir en Fig. **??**) sont indépendantes du fait que ces occurrences appartiennent à des dimères ou non, et la constante de réaction pour la déliaison de deux occurrences de protéine est indépendante de l'état de phosphorylation de ces deux occurrences de protéine (voir en Fig. **??**). Par contre, il existe plusieurs règles de liaison (voir en Fig. **??**), avec des constantes de réaction pouvant varier selon l'état de phosphorylation des deux occurrences de protéine qui se lient. Par exemple, si la constante de réaction $k_{AB\star}$ est choisie plus grande que la constante de réaction $k_{AB}$, alors la formation de dimère est facilitée quand les deux occurrences des protéines qui se lient sont préalablement phosphorylée. En conséquence, l'état de phosphorylation des occurrences de protéine au sein des dimères n'est pas indépendante. Lorsque l'occurrence de la protéine $A$ est phosphorylée, il y a plus de chance que l'occurrence de la protéine $B$ le soit également.

Il est tentant de vouloir abstraire cette corrélation. Cela reviendrait à oublier quelles occurrences de protéine sont liées entre-elles, et à suivre indépendamment les occurrences de protéines en ne gardant que leur état de phosphorylation et de liaison. Tous les interactions se traduisent directement à ce niveau d'abstraction, sauf celle de déliaison. En effet, une déliaison fait intervenir deux occurrences de protéines liées entre-elles. Comme les liaisons ne sont plus représentées. Du coup, il semble naturel de représenter les règles de liaisons par deux règles

indépendantes, l'une pour libérer les occurrences liées de la protéine $A$ et l'autre pour libérer les occurrences liées de le protéine $B$ (ici l'état de liaison peut être assimilé à un état interne puisque l'information de savoir quelle occurrence de protéine est liée à quelle autre, n'est pas maintenue).

Quelles sont les conséquences de cette modification des règles du modèles quant à la dynamique des systèmes engendrés ?

La sémantique différentielle du modèle initial est définie par le système d'équations suivant :

$$\begin{array}{rcl}
[\circ A\dashv]' &=& k_{A-}[\bullet A\dashv] + k_{A..B}([\circ AB\circ] + [\circ AB\bullet]) - (k_{A+} + k_{AB}([\vdash B\circ] + [\vdash B\bullet]))[\circ A\dashv] \\
[\bullet A\dashv]' &=& k_{A+}[\circ A\dashv] + k_{A..B}([\bullet AB\circ] + [\bullet AB\bullet]) - (k_{A-} + k_{AB}[\vdash B\circ] + k_{AB\star}[\vdash B\bullet])[\bullet A\dashv] \\
[\vdash B\circ]' &=& k_{B-}[\vdash B\bullet] + k_{A..B}([\circ AB\circ] + [\bullet AB\circ]) - (k_{B+} + k_{AB}([\circ A\dashv] + [\bullet A\dashv]))[\vdash B\circ] \\
[\vdash B\bullet]' &=& k_{B+}[\vdash B\circ] + k_{A..B}([\circ AB\bullet] + [\bullet AB\bullet]) - (k_{B-} + k_{AB}[\circ A\dashv] + k_{AB\star}[\bullet A\dashv])[\vdash B\bullet] \\
[\circ AB\circ]' &=& k_{A-}[\bullet AB\circ] + k_{B-}[\circ AB\bullet] + k_{AB}[\circ A\dashv][\vdash B\circ] - (k_{A+} + k_{B+} + k_{A..B})[\circ AB\circ] \\
[\bullet AB\circ]' &=& k_{A+}[\circ AB\circ] + k_{B-}[\bullet AB\bullet] + k_{AB}[\bullet A\dashv][\vdash B\circ] - (k_{A-} + k_{B+} + k_{A..B})[\bullet AB\circ] \\
[\circ AB\bullet]' &=& k_{A-}[\bullet AB\bullet] + k_{B+}[AB] + k_{AB}[A][B^\star] - (k_{A+} + k_{B-} + k_{A..B})[\circ AB\bullet] \\
[\bullet AB\bullet]' &=& k_{A+}[\circ AB\bullet] + k_{B+}[\bullet AB\circ] + k_{AB\star}[\bullet A\dashv][\vdash B\bullet] - (k_{A-} + k_{B-} + k_{A..B})[\bullet AB\bullet].
\end{array}$$

Dans ce système, la concentration en monomère de la protéine $A$ non phosphorylée est notée $[\circ A\dashv]$ et celle de la protéine phosphorylée est notée $[\bullet A\dashv]$. La concentration en monomère de la protéine $B$ non phosphorylée est notée $[\vdash B\circ]$ et celle de la protéine phosphorylée est notée $[\vdash B\bullet]$. La concentration en dimère entièrement non phosphorylé est notée $[\circ AB\circ]$, la concentration en dimère avec l'occurrence de la protéine $A$ phosphorylé, et non l'occurrence de la protéine $B$, est notée $[\bullet AB\circ]$, la concentration en dimère avec l'occurrence de la protéine $B$ phosphorylé, et non l'occurrence de la protéine $A$, est notée $[\circ AB\bullet]$ et la concentration en dimère doublement phosphorylé est notée $[\bullet AB\bullet]$.

Le modèle simplifié engendre lui le système différentiel suivant :

$$\begin{array}{rcl}
[\circ A\dashv]' &=& k_{A-}[\bullet A\dashv] + k_{A..B}[\circ A-] - (k_{A+} + k_{AB}([\vdash B\circ] + [\vdash B\bullet]))[\circ A\dashv] \\
[\bullet A\dashv]' &=& k_{A+}[\circ A\dashv] + k_{A..B}[\bullet A-] - (k_{A-} + k_{AB}[\vdash B\circ] + k_{AB\star}[\vdash B\bullet])[\bullet A\dashv] \\
[\vdash B\circ]' &=& k_{B-}[\vdash B\bullet] + k_{A..B}[-B\circ] - (k_{B+} + k_{AB}([\circ A\dashv] + [\bullet A\dashv]))[\vdash B\circ] \\
[\vdash B\bullet]' &=& k_{B+}[\vdash B\circ] + k_{A..B}[-B\bullet B] - (k_{B-} + k_{AB}[\circ A\dashv] + k_{AB\star}[\bullet A\dashv])[\vdash B\bullet] \\
[\circ A-]' &=& k_{A-}[\bullet A-] + k_{AB}[\circ A\dashv]([\vdash B\circ] + [\vdash B\bullet]) - (k_{A+} + k_{A..B})[\circ A-] \\
[\bullet A-]' &=& k_{A+}[\circ A-] + k_{AB}[\bullet A\dashv][\vdash B\circ] + k_{AB\star}[\bullet A\dashv][\vdash B\bullet] - (k_{A-} + k_{A..B})[\bullet A-] \\
[-B\circ]' &=& k_{B-}[-B\bullet B] + k_{AB}[\vdash B\circ]([\circ A\dashv] + [\bullet A\dashv]) - (k_{B+} + k_{A..B})[-B\circ] \\
[-B\bullet B]' &=& k_{B+}[-B\circ] + k_{AB}[\circ A\dashv][\vdash B\bullet] + k_{AB\star}[\bullet A\dashv][\vdash B\bullet] - (k_{B-} + k_{A..B})[-B\bullet B],
\end{array}$$

dans lequel[1] $[\circ A-]$ représente la concentration en protéine $A$ liée et non phosphorylée, $[\bullet A-]$ la concentration en protéine $A$ liée et phosphorylée, $[-B\circ]$ la concentration en protéine $B$ liée et non phosphorylée et $[-B\bullet B]$ la concentration en protéine $B$ liée et phosphorylée.

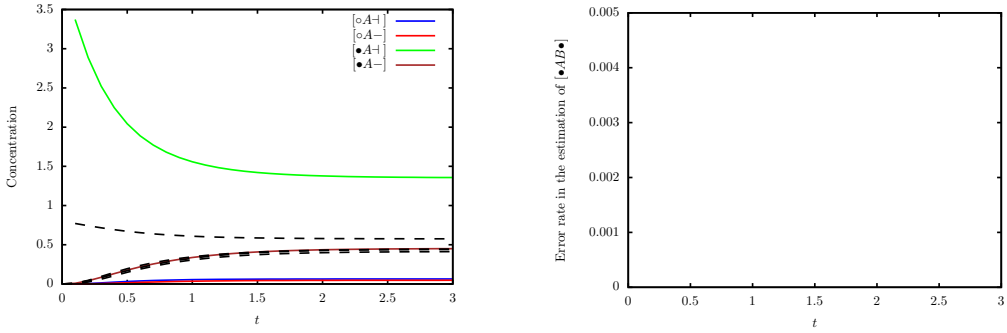
Le système d'équations engendré par le modèle initial et celui-engendré par le modèle simplifié sont liées formellement par les contraintes suivantes :

$$\begin{cases}
[\circ A-] := [AB] + [\circ AB\bullet], \\
[\bullet A-] := [\bullet AB\circ] + [\bullet AB\bullet], \\
[-B\circ] := [AB] + [\bullet AB\circ], \\
[-B\bullet B] := [\circ AB\bullet] + [\bullet AB\bullet].
\end{cases}$$

qui sont obtenues en prennant la définition extensionnelle des parties de complexes biochimiques.

$$\begin{array}{rcl}
[\circ A\dashv]' &=& k_{A-}[\bullet A\dashv] + k_{A..B}[\circ A-] - (k_{A+} + k_{AB}([\vdash B\circ] + [\vdash B\bullet]))[\circ A\dashv] \\
[\bullet A\dashv]' &=& k_{A+}[\circ A\dashv] + k_{A..B}[\bullet A-] - (k_{A-} + k_{AB}[\vdash B\circ] + k_{AB\star}[\vdash B\bullet])[\bullet A\dashv] \\
[\vdash B\circ]' &=& k_{B-}[\vdash B\bullet] + k_{A..B}[-B\circ] - (k_{B+} + k_{AB}([\circ A\dashv] + [\bullet A\dashv]))[\vdash B\circ] \\
[\vdash B\bullet]' &=& k_{B+}[\vdash B\circ] + k_{A..B}[-B\bullet B] - (k_{B-} + k_{AB}[\circ A\dashv] + k_{AB\star}[\bullet A\dashv])[\vdash B\bullet] \\
[\circ A-]' &=& k_{A-}[\bullet A-] + k_{AB}[\circ A\dashv]([\vdash B\circ] + [\vdash B\bullet]) - (k_{A+} + k_{A..B})[\circ A-] \\
[\bullet A-]' &=& k_{A+}[\circ A-] + k_{AB}[\bullet A\dashv][\vdash B\circ] + k_{AB\star}[\bullet A\dashv][\vdash B\bullet] - (k_{A-} + k_{A..B})[\bullet A-] \\
[-B\circ]' &=& k_{B-}[-B\bullet B] + k_{AB}[\vdash B\circ]([\circ A\dashv] + [\bullet A\dashv]) - (k_{B+} + k_{A..B})[-B\circ] \\
[-B\bullet B]' &=& k_{B+}[-B\circ] + k_{AB}[\circ A\dashv][\vdash B\bullet] + k_{AB\star}[\bullet A\dashv][\vdash B\bullet] - (k_{B-} + k_{A..B})[-B\bullet B],
\end{array}$$

[1] L'exposant $\diamond$ représente un état de phosphorylation non spécifié.




(a) Évolution de la concentration de certains motifs. (b)

Figure 5.1: À gauche, l'évolution de la concentration des différentes configuration de la protéine $A$, qu'elle soit phosphorylée, ou non, et quelle soit liée, ou non. Les courbes colorées sont obtenues en simulant numériquement le système d'équations initial, alors que les courbes le sont à partir du système d'équations réduit. Ce taux est défini comme la différence des deux valeurs (dans le système initial et dans le système réduit) divisé par la plus grande de ces deux valeurs. Toutes les constantes de réaction sont fixées à 1, sauf la constante $k_{AB\star}$ qui est fixée à 10. À l'origine, la concentration en monomère non phosphorylé de la protéine $A$ en monomère non phosphorylé de la protéine $B$ sont toutes deux fixées à 2 ; la concentration des autres composants potentiels est fixée à 0.

where[2] $[\circ A{-}] := [AB] + [\circ AB\bullet]$, $[\bullet A{-}] := [\bullet AB\circ] + [\bullet AB\bullet]$, $[{-}B\circ] := [AB] + [\bullet AB\circ]$, and $[{-}B\bullet B] := [\circ AB\bullet] + [\bullet AB\bullet]$, are satisfied.

Yet, this correlation forbids the reduction of the stochastic semantics. Let us explain why. In the stochastic semantics, a chemical soup can be denoted by a 8-tuple $\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle$ of natural numbers, where $n_X$ is the number of instance of $X$ in the chemical soup, for any $X \in \{A, A^\star, B, B^\star, AB, \bullet AB\circ, \circ AB\bullet, \bullet AB\bullet\}$. The probability $P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle)$ that the system is in a given state $\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n$ at time $t$ is given by the following master equation:

$$
\begin{aligned}
&P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle)' = \\
&\quad k_{A+}(n_A+1)P_t(\langle n_A{+}1, n_{A^\star}{-}1, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle) \\
+\quad & k_{A+}(n_{AB}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}{+}1, n_{\bullet AB\circ}{-}1, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle) \\
+\quad & k_{A+}(n_{\circ AB\bullet}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}{+}1, n_{\bullet AB\bullet}{-}1\rangle) \\
+\quad & k_{A-}(n_{A^\star}+1)P_t(\langle n_A{-}1, n_{A^\star}{+}1, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle) \\
+\quad & k_{A-}(n_{\bullet AB\circ}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}{-}1, n_{\bullet AB\circ}{+}1, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle) \\
+\quad & k_{A-}(n_{\bullet AB\bullet}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}{-}1, n_{\bullet AB\bullet}{+}1\rangle) \\
+\quad & k_{B+}(n_B+1)P_t(\langle n_A, n_{A^\star}, n_B{+}1, n_{B^\star}{-}1, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle) \\
+\quad & k_{B+}(n_{AB}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}{+}1, n_{\bullet AB\circ}, n_{\circ AB\bullet}{-}1, n_{\bullet AB\bullet}\rangle) \\
+\quad & k_{B+}(n_{\bullet AB\circ}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}{+}1, n_{\circ AB\bullet}, n_{\bullet AB\bullet}{-}1\rangle) \\
+\quad & k_{B-}(n_{B^\star}+1)P_t(\langle n_A, n_{A^\star}, n_B{-}1, n_{B^\star}{+}1, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle) \\
+\quad & k_{B-}(n_{\circ AB\bullet}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}{-}1, n_{\bullet AB\circ}, n_{\circ AB\bullet}{+}1, n_{\bullet AB\bullet}\rangle) \\
+\quad & k_{B-}(n_{\bullet AB\bullet}+1)P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}{-}1, n_{\circ AB\bullet}, n_{\bullet AB\bullet}{+}1\rangle) \\
+\quad & k_{AB}(n_A+1)(n_B+1)P_t(\langle n_A{+}1, n_{A^\star}, n_B{+}1, n_{B^\star}, n_{AB}{-}1, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle) \\
+\quad & k_{AB}(n_A+1)(n_{B^\star}+1)P_t(\langle n_A{+}1, n_{A^\star}, n_B, n_{B^\star}{+}1, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}{-}1, n_{\bullet AB\bullet}\rangle) \\
+\quad & k_{AB}((n_{A^\star}+1)(n_B+1))P_t(\langle n_A, n_{A^\star}{+}1, n_B{+}1, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}{-}1, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle) \\
+\quad & k_{AB\star}((n_{A^\star}+1)(n_{B^\star}+1))P_t(\langle n_A, n_{A^\star}{+}1, n_B, n_{B^\star}{+}1, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}{-}1\rangle) \\
&\quad \text{(continued on the next page)}
\end{aligned}
$$

[2]The superscript $\diamond$ stands for "whatever the phosphorilation state is".

$$
\begin{aligned}
&+\quad k_{A..B}(n_{AB}-1)P_t(\langle n_A{-}1, n_{A^\star}{-}1, n_B, n_{B^\star}, n_{AB}{+}1, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle)\\
&+\quad k_{A..B}(n_{\circ AB\bullet}-1)P_t(\langle n_A{-}1, n_{A^\star}, n_B, n_{B^\star}{-}1, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}{+}1, n_{\bullet AB\bullet}\rangle)\\
&+\quad k_{A..B}(n_{\bullet AB\circ}-1)P_t(\langle n_A, n_{A^\star}{-}1, n_B{-}1, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}{+}1, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle)\\
&+\quad k_{A..B}(n_{\bullet AB\bullet}-1)P_t(\langle n_A, n_{A^\star}{-}1, n_B, n_{B^\star}{-}1, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}{+}1\rangle)\\
&-\quad k_{A+}(n_A+n_{AB}+n_{\circ AB\bullet})P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle)\\
&-\quad k_{A-}(n_{A^\star}+n_{\bullet AB\circ}+n_{\bullet AB\bullet})P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle)\\
&-\quad k_{B+}(n_B+n_{AB}+n_{\bullet AB\circ})P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle)\\
&-\quad k_{B-}(n_{B^\star}+n_{AB^\star}+n_{\bullet AB\bullet})P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle)\\
&-\quad k_{AB}((n_A+n_{A^\star})(n_B+n_{B^\star})-n_{A^\star}n_{B^\star})P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle)\\
&-\quad k_{AB\star}n_{A^\star}n_{B^\star}P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle)\\
&-\quad k_{A..B}(n_{AB}+n_{\circ AB\bullet}+n_{\bullet AB\circ}+n_{\bullet AB\bullet})P_t(\langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle)
\end{aligned}
$$

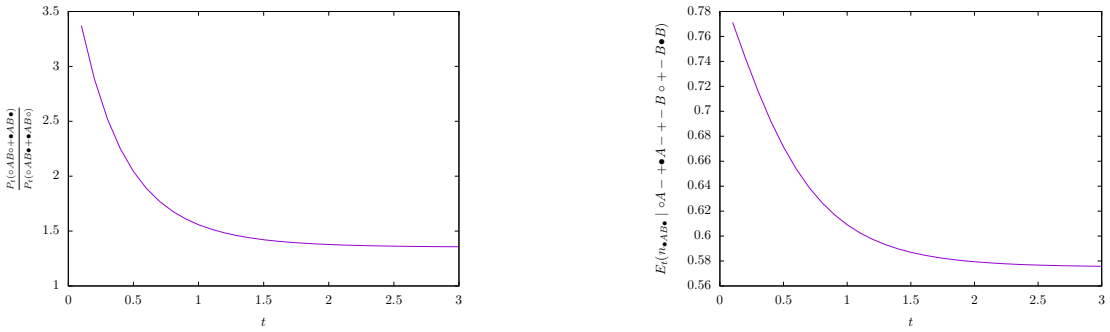
As in the example of Sect. 5.4.1, we would like to abstract away the correlation between the phosphorilation state of the proteins $A$ and the phosphorilation state of the proteins $B$ which belong to the same complex. Given a state $\sigma = \langle n_A, n_{A^\star}, n_B, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}\rangle$, we denote by $\beta(\sigma)$ the 8-tuple $\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle$ (such a tuple is called an abstract state). The probability $P_t^\sharp(\sigma^\sharp)$ that the system is in a state $\sigma$ such that $\beta(\sigma) = \sigma^\sharp$ at time $t$, satisfies the following equation:

$$
\begin{aligned}
&P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle)' =\\
&\qquad k_{A+}(n_A+1)P_t^\sharp(\langle n_{\circ A\dashv}{+}1, n_{\bullet A\dashv}{-}1, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle)\\
&+\quad k_{A+}(n_{\circ A-}+1)P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}{+}1, n_{\bullet A-}{-}1, n_{-B\circ}, n_{-B\bullet B}\rangle)\\
&+\quad k_{A-}(n_{A^\star}+1)P_t^\sharp(\langle n_{\circ A\dashv}{-}1, n_{\bullet A\dashv}{+}1, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle)\\
&+\quad k_{A-}(n_{\bullet A-}+1)P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}{-}1, n_{\bullet A-}{+}1, n_{-B\circ}, n_{-B\bullet B}\rangle)\\
&+\quad k_{B+}(n_B+1)P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}{+}1, n_{\vdash B\bullet}{-}1, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle)\\
&+\quad k_{B+}(n_{-B\circ}+1)P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}{+}1, n_{-B\bullet B}{-}1\rangle)\\
&+\quad k_{B-}(n_{B^\star}+1)P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}{-}1, n_{\vdash B\bullet}{+}1, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle)\\
&+\quad k_{B-}(n_{-B\bullet B}+1)P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}{-}1, n_{-B\bullet B}{+}1\rangle)\\
&+\quad k_{AB}(n_A+1)(n_B+1)P_t^\sharp(\langle n_{\circ A\dashv}{+}1, n_{\bullet A\dashv}, n_{\vdash B\circ}{+}1, n_{\vdash B\bullet}, n_{\circ A-}{-}1, n_{\bullet A-}, n_{-B\circ}{-}1, n_{-B\bullet B}\rangle)\\
&+\quad k_{AB}(n_A+1)(n_{B^\star}+1)P_t^\sharp(\langle n_{\circ A\dashv}{+}1, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}{+}1, n_{\circ A-}{-}1, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}{-}1\rangle)\\
&+\quad k_{AB}((n_{A^\star}+1)(n_B+1))P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}{+}1, n_{\vdash B\circ}{+}1, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}{-}1, n_{-B\circ}{-}1, n_{-B\bullet B}\rangle)\\
&+\quad k_{AB\star}((n_{A^\star}+1)(n_{B^\star}+1))P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}{+}1, n_{\vdash B\circ}, n_{\vdash B\bullet}{+}1, n_{\circ A-}, n_{\bullet A-}{-}1, n_{-B\circ}, n_{-B\bullet B}{-}1\rangle)\\
&+\quad k_{A..B}\tilde{E}_t(n_{AB} \mid \langle n_{\circ A\dashv}{-}1, n_{\bullet A\dashv}, n_{\vdash B\circ}{-}1, n_{\vdash B\bullet}, n_{\circ A-}{+}1, n_{\bullet A-}, n_{-B\circ}{+}1, n_{-B\bullet B}\rangle)\\
&+\quad k_{A..B}\tilde{E}_t(n_{\circ AB\bullet} \mid \langle n_A{-}1, n_{A^\star}, n_B, n_{B^\star}{-}1, n_{AB}{+}1, n_{\bullet AB\circ}, n_{\circ AB\bullet}, n_{\bullet AB\bullet}{+}1\rangle)\\
&+\quad k_{A..B}\tilde{E}_t(n_{\bullet AB\circ} \mid \langle n_A, n_{A^\star}{-}1, n_B{-}1, n_{B^\star}, n_{AB}, n_{\bullet AB\circ}{+}1, n_{\circ AB\bullet}{+}1, n_{\bullet AB\bullet}\rangle)\\
&+\quad k_{A..B}\tilde{E}_t(n_{\bullet AB\bullet} \mid \langle n_A, n_{A^\star}{-}1, n_B, n_{B^\star}{-}1, n_{AB}, n_{\bullet AB\circ}{+}1, n_{\circ AB\bullet}, n_{\bullet AB\bullet}{+}1\rangle)\\
&-\quad k_{A+}(n_A+n_{\circ A-})P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle)\\
&-\quad k_{A-}(n_{A^\star}+n_{\bullet A-})P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle)\\
&-\quad k_{B+}(n_B+n_{-B\circ})P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle)\\
&-\quad k_{B-}(n_{B^\star}+n_{-B\bullet B})P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle)\\
&-\quad k_{AB}((n_A+n_{A^\star})(n_B+n_{B^\star})-n_{A^\star}n_{B^\star})P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle)\\
&-\quad k_{AB\star}(n_{A^\star}n_{B^\star})P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle)\\
&-\quad k_{A..B}(n_{\bullet A-}+n_{\circ A-})P_t^\sharp(\langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle),
\end{aligned}
$$

where for any expression $X(\sigma)$ and any (abstract) state $\sigma^\sharp$, the expression $\tilde{E}_t(X(\sigma) \mid \sigma^\sharp)$ denotes the product between the conditional expectation $E_t(X(\sigma) \mid \sigma^\sharp)$ of the expression $X(\sigma)$ knowing that $\beta(\sigma) = \sigma^\sharp$, and the probability $P_t^\sharp(\sigma^\sharp)$ of being in a state $\sigma$ such that $\beta(\sigma) = \sigma^\sharp$.

Whenever $k_{AB} = k_{AB\star}$, we can check that the fact that $P_t(\sigma) = P_t(\sigma')$ for any pair of states $\sigma$,$\sigma'$ such that $\beta(\sigma) = \beta(\sigma')$ is an invariant. Thus, provided that $k_{AB} = k_{AB\star}$ and that there is no correlation between the phosphorilation state of the proteins $A$ and $B$ which are bound together at time $t = 0$, one can use the following properties:

$$
\begin{aligned}
E_t(n_{AB} \mid \langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle) &= \frac{n_{\circ A-}n_{-B\circ}}{\circ A-+\bullet A-}\\
E_t(n_{\circ AB\bullet} \mid \langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle) &= \frac{n_{\circ A-}n_{-B\bullet B}}{\circ A-+\bullet A-}\\
E_t(n_{\bullet AB\circ} \mid \langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle) &= \frac{n_{\bullet A-}n_{-B\circ}}{\circ A-+\bullet A-}\\
E_t(n_{\bullet AB\bullet} \mid \langle n_{\circ A\dashv}, n_{\bullet A\dashv}, n_{\vdash B\circ}, n_{\vdash B\bullet}, n_{\circ A-}, n_{\bullet A-}, n_{-B\circ}, n_{-B\bullet B}\rangle) &= \frac{n_{\bullet A-}n_{-B\bullet B}}{\circ A-+\bullet A-},
\end{aligned}
$$




(a) (b)

Figure 5.2: On the left, quotient between the probability of being in the state $AB + \bullet AB\bullet$ and the probability of being in the state $\circ AB\bullet + \bullet AB\circ$. On the right, conditional expectation of the number of fully phosphorilated complexes $\bullet AB\bullet$ knowing that all proteins are bound, and that there is exactly one phosphorilated protein $A$ and exactly one phosphorilated $B$. All rates are set to 1, except $k_{AB\star}$ which is set to 10. At time 0, the chemical soup is made of two proteins $A$ and two proteins $B$, none of these proteins being phosphorilated or bound.

so as to write conditional expectations of $n_{AB}$, $n_{\circ AB\bullet}$, $n_{\bullet AB\circ}$, and $n_{\bullet AB\bullet}$ as time-independent expressions of $n_{\circ A-}$, $n_{\bullet A-}$, $n_{-B\circ}$, and $n_{-B\bullet B}$.

Whenever $k_{AB} \neq k_{AB\star}$, these conditional expectations may be time-dependent. We show in Fig. 5.2(a) that the ratio between the probability of being in the state $AB + \bullet AB\bullet$ and the probability of being in the state $\circ AB\bullet + \bullet AB\circ$ is time-dependent. Moreover, we show in Fig. 5.2(b) that the conditional expectation of $n_{\bullet AB\bullet}$ knowing that we are in the (abstract) state $\circ A - +\bullet A - + - B \circ + - B\bullet B$ is time-dependent as well, which forbids doing the same simplification as in the differential semantics.

We have seen through this example that some reactions may operate simultaneously over two fragments. This leads to coupled semi-reactions. We have noticed that coupled semi-reactions raise no issue when reducing the differential semantics. We say that the application of semi-rules is fair in the differential semantics, since the proportion of the concentration of a given fragment that is consumed by a semi-reaction does not depend on the correlation between the states of the two fragments. This is not the case in the stochastic semantics: we have noticed that the stochastic semantics can be reduced only if the state of the two fragments are not correlated, otherwise the choice of the fragments on which coupled semi-reactions operate is entangled, which forbids the reduction. In other words, we say that in the differential semantics, we can abstract away the correlations which are not observed by rules, whereas in the stochastic semantics, we have to prove that the rules cannot enforce correlations between the state of some fragments and we use this property so as to reduce the dimension of the state space of the system. In the later case, the reduction is only valid when there is no correlation at time $t = 0$.

### 5.4.3 An example of distant control

In this section, we show that, in the stochastic semantics, one protein can control the behavior of another one even if they are not in the same connected component in the left hand side of a reaction.

We consider a kind of proteins, A, which bears two phosphorilation sites. Each phosphorilation site can be unphosphorilated, or phosphorilated. We use the symbol $\star$ to denote phosphorilated sites. The phosphorilation state of the first site is written as a superscript, whereas the phosphorilation state of the second site is written as a subscript. This way, a protein A having the first site phosphorilated and the second site unphosphorilated is denoted by $\mathrm{A}^\star$.

The behavior of a chemical soup can be described by the following set of reactions:

$$\mathrm{A} \underset{k^-}{\overset{k^+}{\rightleftharpoons}} \mathrm{A}^\star \qquad \mathrm{A}_\star \underset{k^-}{\overset{k^+}{\rightleftharpoons}} \mathrm{A}^\star_\star$$

$$\begin{aligned} \mathrm{A} + \mathrm{A}^\star &\xrightarrow{k_+} \mathrm{A}_\star + \mathrm{A}^\star \\ \mathrm{A}^\star + \mathrm{A}^\star &\xrightarrow{k_+} \mathrm{A}^\star_\star + \mathrm{A}^\star \\ \mathrm{A} + \mathrm{A}^\star_\star &\xrightarrow{k_+} \mathrm{A}_\star + \mathrm{A}^\star_\star \\ \mathrm{A}^\star + \mathrm{A}^\star_\star &\xrightarrow{k_+} \mathrm{A}^\star_\star + \mathrm{A}^\star_\star \end{aligned}$$

$$\begin{aligned} \mathrm{A}^\star_\star &\xrightarrow{k_-} \mathrm{A}^\star \\ \mathrm{A}_\star &\xrightarrow{k_-} \mathrm{A} \end{aligned}$$

We have assumed (see second column) that the kinetic of the phosphorilation of the second site of a protein depends on the number of the other proteins that are phosphorilated on their first site — that is to say that the proteins that are phosphorilated on their first site catalyzes the phosphorilation of the second site in the other proteins. We have also assumed that other reactions are purely local, that is to say that the kinetic of phosphorilation and dephosphorilation on the first site does not depend on the phosphorilation state of the second site (neither of the protein being phosphorilated, nor of the other proteins) (see first column), and that the kinetic of dephosphorilation of the second site does not depend on the phosphorilation state of the first site of the proteins in the soup (see third column).

In this example, we would like to abstract the correlation between the phosphorilation state of the two sites of each protein. This could be achieved, by splitting each complex into two parts, and by abstracting away which parts are connected together. It raises an issue for reducing the stochastic semantics. Indeed, one can notice that the reaction which activates the second site of protein favors the phosphorilation of the second site of the protein in the state A. For instance, if we assume that both the number of instances of the protein in state A and the number of instances of the protein in the state $\mathrm{A}^\star$ is equal to $m$, and that the number of instances of the protein in the state $\mathrm{A}^\star_\star$ is equal to $n$. Then, the cumulative activity of the following two reactions:

$$\mathrm{A} + \mathrm{A}^\star \xrightarrow{k_+} \mathrm{A}_\star + \mathrm{A}^\star \qquad \mathrm{A} + \mathrm{A}^\star_\star \xrightarrow{k_+} \mathrm{A}_\star + \mathrm{A}^\star_\star$$

is equal to $k_+(n+m)m$, whereas the cumulative activity of the following two reactions:

$$\mathrm{A}^\star + \mathrm{A}^\star \xrightarrow{k_+} \mathrm{A}^\star_\star + \mathrm{A}^\star \qquad \mathrm{A}^\star + \mathrm{A}^\star_\star \xrightarrow{k_+} \mathrm{A}^\star_\star + \mathrm{A}^\star_\star$$

is equal to $k_+(n+m-1)m$ (the subtraction by 1 comes from the fact that each reactant must be mapped to distinct instances of chemical species). Nevertheless, it does not forbid the reduction of the differential semantics: intuitively, the term 1 vanishes because we consider an infinite number of instances, within an infinite volume.

Let us check formally that the differential semantics of this model can be reduced and explain why we do not know how to abstract its stochastic semantics. This differential semantics is defined by the following system of differential equations:

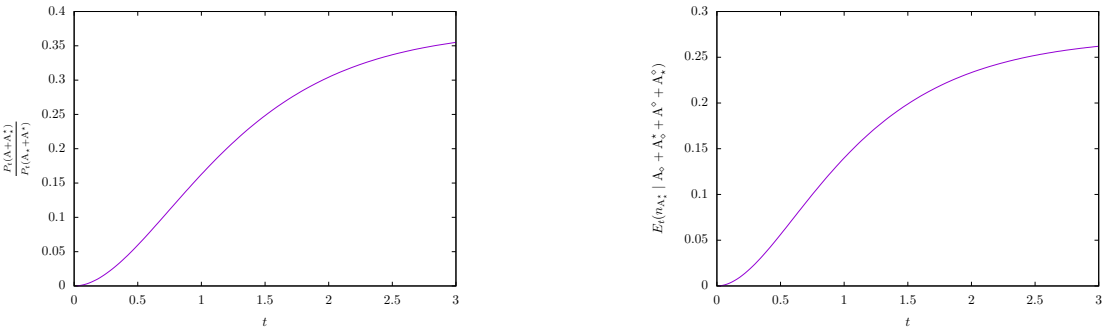
$$\begin{array}{rcl}
[\mathrm{A}]' &=& k^-[\mathrm{A}^\star] + k_-[\mathrm{A}_\star] - (k^+ + k_+([\mathrm{A}^\star] + [\mathrm{A}^\star_\star]))[\mathrm{A}] \\
[\mathrm{A}^\star]' &=& k^+[\mathrm{A}] + k_-[\mathrm{A}^\star_\star] - (k^- + k_+([\mathrm{A}^\star] + [\mathrm{A}^\star_\star]))[\mathrm{A}^\star] \\
[\mathrm{A}_\star]' &=& k^-[\mathrm{A}^\star_\star] + k_+[\mathrm{A}]([\mathrm{A}^\star] + [\mathrm{A}^\star_\star]) - (k^+ + k_-)[\mathrm{A}_\star] \\
[\mathrm{A}^\star_\star]' &=& k^+[\mathrm{A}_\star] + k_+[\mathrm{A}^\star]([\mathrm{A}^\star] + [\mathrm{A}^\star_\star]) - (k^- + k_-)[\mathrm{A}^\star_\star].
\end{array}$$

We notice that the correlation between the two sites can be abstracted away. Indeed, we notice that the following equations:

$$\begin{array}{rcl}
[\mathrm{A}_\diamond]' &=& k^-[\mathrm{A}^\star_\diamond] - k^+[\mathrm{A}_\diamond] \\
[\mathrm{A}^\star_\diamond]' &=& k^+[\mathrm{A}_\diamond] - k^-[\mathrm{A}^\star_\diamond] \\
[\mathrm{A}^\diamond]' &=& k_-[\mathrm{A}^\diamond_\star] - k_+[\mathrm{A}^\diamond][\mathrm{A}^\star_\diamond] \\
[\mathrm{A}^\diamond_\star]' &=& k_+[\mathrm{A}^\diamond][\mathrm{A}^\star_\diamond] - k_-[\mathrm{A}^\diamond_\star],
\end{array}$$

where $[\mathrm{A}_\diamond] := [\mathrm{A}] + [\mathrm{A}_\star]$, $[\mathrm{A}^\star_\diamond] := [\mathrm{A}^\star] + [\mathrm{A}^\star_\star]$, $[\mathrm{A}^\diamond] := [\mathrm{A}] + [\mathrm{A}^\star]$, and $[\mathrm{A}^\diamond_\star] := [\mathrm{A}_\star] + [\mathrm{A}^\star_\star]$, are satisfied.

We now wonder whether the same reduction can be used in the case of the stochastic semantics. In the stochastic semantics, a chemical soup can be denoted by a 4-tuple $\langle n_\mathrm{A}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle$ of natural numbers, where $n_X$ is the number of instance of $X$ in the chemical soup, for any $X \in \{\mathrm{A}, \mathrm{A}_\star, \mathrm{A}^\star, \mathrm{A}^\star_\star\}$. The probability $P_t(\langle n_\mathrm{A}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle)$ that the system is in a given state $\langle n_\mathrm{A}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle$ at time $t$ is given by the following




(a) (b)

Figure 5.3: On the left, quotient between the probability of being in the state $\mathrm{A} + \mathrm{A}^\star_\star$ and the probability of being in the state $\mathrm{A}_\star + \mathrm{A}^\star$. On the right, conditional expectation of the number of protein in the state $\mathrm{A}^\star$ knowing that (i) there are two proteins, (ii) exactly one protein is phosphorilated at its the first site, and (iii) exactly one protein (potentially the same) is phosphorilated at its second site. All rates are set to 1. At time 0, the chemical soup is made of two proteins $A$ fully unphosphorilated.

master equation:

$$
\begin{array}{rl}
& P_t(\langle n_{\mathrm{A}}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle)' = \\
& \quad k^+(n_{\mathrm{A}}+1)P_t(\langle n_{\mathrm{A}}{+}1, n_{\mathrm{A}^\star}{-}1, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle) \\
+ & \quad k^+(n_{\mathrm{A}_\star}+1)P_t(\langle n_{\mathrm{A}}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}{+}1, n_{\mathrm{A}^\star_\star}{-}1\rangle) \\
+ & \quad k^-(n_{\mathrm{A}^\star}+1)P_t(\langle n_{\mathrm{A}}{-}1, n_{\mathrm{A}^\star}{+}1, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle) \\
+ & \quad k^-(n_{\mathrm{A}^\star_\star}+1)P_t(\langle n_{\mathrm{A}}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}{-}1, n_{\mathrm{A}^\star_\star}{+}1\rangle) \\
+ & \quad k_+(n_{\mathrm{A}}+1)(n_{\mathrm{A}^\star}+n_{\mathrm{A}^\star_\star})P_t(\langle n_{\mathrm{A}}{+}1, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}{-}1, n_{\mathrm{A}^\star_\star}\rangle) \\
+ & \quad k_+(n_{\mathrm{A}^\star}+1)(n_{\mathrm{A}^\star}+n_{\mathrm{A}^\star_\star}-1)P_t(\langle n_{\mathrm{A}}, n_{\mathrm{A}^\star}{+}1, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}{-}1\rangle) \\
+ & \quad k_-(n_{\mathrm{A}_\star}+1)P_t(\langle n_{\mathrm{A}}{-}1, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}{+}1, n_{\mathrm{A}^\star_\star}\rangle) \\
+ & \quad k_-(n_{\mathrm{A}^\star_\star}+1)P_t(\langle n_{\mathrm{A}}, n_{\mathrm{A}^\star}{-}1, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}{+}1\rangle) \\
- & \quad k^+(n_{\mathrm{A}}+n_{\mathrm{A}_\star})P_t(\langle n_{\mathrm{A}}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle) \\
- & \quad k^-(n_{\mathrm{A}^\star}+n_{\mathrm{A}^\star_\star})P_t(\langle n_{\mathrm{A}}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle) \\
- & \quad k_+(n_{\mathrm{A}}(n_{\mathrm{A}^\star}+n_{\mathrm{A}^\star_\star})+n_{\mathrm{A}^\star}(n_{\mathrm{A}^\star}+n_{\mathrm{A}^\star_\star}-1))P_t(\langle n_{\mathrm{A}}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle) \\
- & \quad k_-(n_{\mathrm{A}_\star}+n_{\mathrm{A}^\star_\star})P_t(\langle n_{\mathrm{A}}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle).
\end{array}
$$

Given a state $\sigma = \langle n_{\mathrm{A}}, n_{\mathrm{A}^\star}, n_{\mathrm{A}_\star}, n_{\mathrm{A}^\star_\star}\rangle$, we denote by $\beta(\sigma)$ the 4-tuple $\langle n_{\mathrm{A}_\diamond}, n_{\mathrm{A}^\star_\diamond}, n_{\mathrm{A}^\diamond}, n_{\mathrm{A}^\diamond_\star}\rangle$ (such a tuple is called an abstract state). The probability $P^\sharp_t(\sigma^\sharp)$ that the system is in a state $\sigma$ such that $\beta(\sigma) = \sigma^\sharp$ at time $t$, satisfies the following equation:

$$
\begin{array}{rl}
& P^\sharp_t(\langle n_{\mathrm{A}_\diamond}, n_{\mathrm{A}^\star_\diamond}, n_{\mathrm{A}^\diamond}, n_{\mathrm{A}^\diamond_\star}\rangle)' = \\
& \quad k_+(n_{\mathrm{A}_\diamond}+1)P^\sharp_t(\langle n_{\mathrm{A}_\diamond}{+}1, n_{\mathrm{A}^\star_\diamond}{-}1, n_{\mathrm{A}^\diamond}, n_{\mathrm{A}^\diamond_\star}\rangle) \\
+ & \quad k^-(n_{\mathrm{A}^\star_\diamond}+1)P^\sharp_t(\langle n_{\mathrm{A}_\diamond}{-}1, n_{\mathrm{A}^\star_\diamond}{+}1, n_{\mathrm{A}^\diamond}, n_{\mathrm{A}^\diamond_\star}\rangle) \\
+ & \quad k_+(n_{\mathrm{A}^\diamond}+1)n_{\mathrm{A}^\star_\diamond}P^\sharp_t(\langle n_{\mathrm{A}_\diamond}, n_{\mathrm{A}^\star_\diamond}, n_{\mathrm{A}^\diamond}{+}1, n_{\mathrm{A}^\diamond_\star}{-}1\rangle) \\
+ & \quad k_-(n_{\mathrm{A}^\diamond_\star}+1)P^\sharp_t(\langle n_{\mathrm{A}_\diamond}, n_{\mathrm{A}^\star_\diamond}, n_{\mathrm{A}^\diamond}{-}1, n_{\mathrm{A}^\diamond_\star}{+}1\rangle) \\
- & \quad (k^+n_{\mathrm{A}_\diamond}+k^-n_{\mathrm{A}^\star_\diamond}+k_+n_{\mathrm{A}^\diamond}n_{\mathrm{A}^\star_\diamond}+k_-n_{\mathrm{A}^\diamond_\star})P^\sharp_t(\langle n_{\mathrm{A}_\diamond}, n_{\mathrm{A}^\star_\diamond}, n_{\mathrm{A}^\diamond}, n_{\mathrm{A}^\diamond_\star}\rangle) \\
- & \quad k_+\tilde{E}_t(n_{\mathrm{A}^\star} \mid \langle n_{\mathrm{A}_\diamond}, n_{\mathrm{A}^\star_\diamond}{+}1, n_{\mathrm{A}^\diamond}, n_{\mathrm{A}^\diamond_\star}{-}1\rangle) \\
+ & \quad k_+\tilde{E}_t(n_{\mathrm{A}^\star} \mid \langle n_{\mathrm{A}_\diamond}, n_{\mathrm{A}^\star_\diamond}, n_{\mathrm{A}^\diamond}, n_{\mathrm{A}^\diamond_\star}\rangle),
\end{array}
$$

where for any expression $X(\sigma)$ and any (abstract) state $\sigma^\sharp$, the expression $\tilde{E}_t(X(\sigma) \mid \sigma^\sharp)$ denotes the product between the conditional expectation $E_t(X(\sigma) \mid \sigma^\sharp)$ of the expression $X(\sigma)$ knowing that $\beta(\sigma) = \sigma^\sharp$ and the probability $P^\sharp_t(\sigma^\sharp)$ of being in a state $\sigma$ such that $\beta(\sigma) = \sigma^\sharp$.

In general, the conditional properties of the number of instances of proteins in the form $\mathrm{A}^\star$ having fixed a given abstract state, is time-dependent. We show in Fig. 5.3(a) that the ratio between the probability of being in the state $\mathrm{A} + \mathrm{A}^\star_\star$ and the probability of being in the state $\mathrm{A}_\star + \mathrm{A}^\star$ is time-dependent. Moreover, we show in Fig. 5.3(b) that the conditional expectation of $n_{\mathrm{A}^\star}$ knowing that we are in the (abstract) state $\mathrm{A}_\diamond + \mathrm{A}^\star_\diamond + \mathrm{A}^\diamond + \mathrm{A}^\diamond_\star$ is time-dependent, which forbids doing the same simplication as in the differential semantics.

| Model | early EGF | EGF/Insulin cross talk | SFB |
|---|---|---|---|
| Species | 356 | 2899 | $\sim 2.10^{19}$ |
| ODE fragments | 38 | 208 | $\sim 2.10^{5}$ |
| Stochastic fragments | 356 | 618 | $\sim 2.10^{19}$ |

Figure 5.4: Reduction factors for differential fragments [64, 47] and stochastic fragments. We try these reduction methods on three models. The first one is the model of the early events of the EGF pathway (see Sect. **??**); the second one, taken from [33, table 7], describes the cross-talk between another model of the early events of the EGF pathway and the insulin receptor; whereas the third one is a version of a pilot study on a larger section of the EGF pathway [46, 12, 107, 21].

We have seen through this example that, because a given instance of chemical species can only be used once as a reactant when applying a given chemical reaction, some corrective terms as $+1$ or $-1$ may appear in master equations. These corrective terms may forbid the reduction of stochastic semantics. Nevertheless, this is not an issue when reducing differential semantics, since these corrective terms vanish when we consider an infinite number of instances of proteins (within an infinite volume).

## 5.5 Conclusion

In this paper, we have illustrated through small examples why it is more difficult to reduce the dimension of the state space of stochastic semantics than the one of differential semantics. In the case of the differential semantics, it is possible to abstract away some correlations between the state of some fragments of chemical species, because these correlations are not observed by the (groups of) reactions. This is not so easy in the case of stochastic semantics, because a given reaction application may operate on several fragments simultaneously, in such a case the choice for the state of fragments on which semi-reactions are applied is driven by the correlation between the state of these two fragments (see Sect. 5.4.2). Moreover, stochastic semantics counts individuals which leads to some constant corrective terms (such as increment or decrement by 1) which also forbids exact reduction (see Sect. 5.4.3).

In Fig. 5.4, we give the number of chemical species, the number of differential fragments, and the number of stochastic fragments for three bigger models. The reduction factor for the differential semantics is very interesting, whereas there is almost no reduction in the stochastic case. A careful look into the models would show that this is due to coupled semi-reactions. Moreover, the reduction that arises in the second model is due to a protein which has two fully independent parts (as in Sect. 5.4.1).

This emphasizes how interesting the stochastic semantics is: the stochastic semantics does not only describe a limit behavior, but also showes the variability of a system and how robust a system is to stochastic variations. The counterpart is that it is very difficult to handle with (as a formal object) and to simplify.

# Chapter 6

# Symétrie dans les graphes à sites

Les symétries apparaissent sous des formes diverses dans les modèles décrits par des règles de réécritures. Certaines sont indissociables de la sémantique. C'est le cas des automorphismes qui représentent des équivalences entre agents au sein de graphes à sites. D'autres peuvent décrire des équivalences entre sites d'interaction, des équivalences entre sortes de protéines ou encore des équivalences spatiales qui permettent de voir la structure de certains complexes biochimiques modulo des isométries.

Dans [62], nous avons introduit un cadre de travail algébrique unifiant pour décrire, inférer, et utiliser des groupes de transformations de graphes à sites dans le langage Kappa. L'idée principale est de considérer des groupes de transformations qui agissent sur les graphes à sites avec une opération supplémentaire permettant de restreindre une transformation sur les sous-graphes d'un graphe. Il est alors possible de définir quand un ensemble de règle est symétrique par rapport à ces groupes de transformation et d'en déduire des propriétés au niveau du comportement global du modèle, que ce soit pour sa sémantique stochastique ou différentielle. Pour simplifier la présentation, nous nous concentrons ici aux symétries qui correspondent à des permutations de sites, qui sont décrites dans [24].

**Travaux voisins.** Plusieurs formalismes permettent de faire apparaître chaque site plusieurs fois dans l'interface des agents. C'est le cas du langage BNGL [9]. Dans le langage ReactC [87], des sites indistinguable peuvent être encodés à l'aide d'hyperliens. Ceci procure un moyen syntaxique de décrire des sites symétriques. En contre-partie, le calcul des l'ensembles des occurrences des motifs de réécriture devient prohibitif (les graphes ne sont plus rigides), ce qui pose de grands problèmes de complexité pour échantilloner les trajectoires de la sémantique stochastique.

En Kappa, la rigidité des graphes est un principe fondamental, et donc, l'utilisation de plusieurs instances d'un même site dans l'interface d'un agent est interdite. Les équivalences entre sites d'interaction peuvent toutefois être décrites à un plus haut-niveau de spécification [78] puis traduites automatiquement dans le noyau pure de Kappa. Cependant, il faut être conscient que cette traduction peut générer un très grand nombre de règles, là où une seule règle aurait été suffisante en utilisant des sites à occurrences multiples.

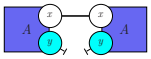
Dans ce chapitre, nous proposons une approche pour détecter automatiquement des équivalences entre sites d'interaction qui n'ont pas besoin d'être spécifiées à la main. Puis, nous utilisons ces équivalences pour réduire l'espace d'état ou le nombre d'espèces biochimiques des modèles considérés. Ceci permet de réduire la dimension des systèmes différentiels sous-jacents et facilite le calcul de propriétés sur la distribution des traces de la sémantique stochastique. Ce n'est en revanche pas crucial pour l'échantillonnage des traces de la sémantique stochastique, puisque la simulation travaille directement au niveau des graphes à sites [49].

Enfin, notre cadre de travail n'est pas limité à la prise en compte des équivalences entre sites d'interaction. Il peut être utilisé pour considérer des chaînes d'agents indépendamment de leur orientation ou des sous-groupes de permutations de sites comme c'est courant en chemoinformatique.

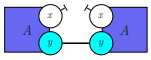
## Remerciements.

Ce chapitre reprends donc les principaux résultats qui avaient été établis avec Ferdinanda Camporesi [24] à propos de la détection et de la prise en compte d'équivalences entre sites d'interaction dans un noyau sans effets

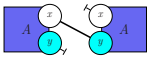
(a) Une occurrence de la protéine $A$ avec ses deux sites de liaison libres.



(b) Deux occurrences de la protéine $A$ liées par leurs deux sites $x$.

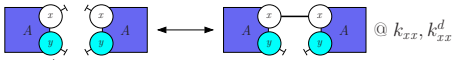


(c) Deux occurrences de la protéine $A$ liées par leurs deux sites $y$.

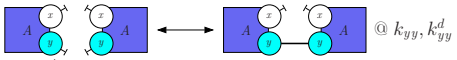


(d) Deux occurrences de la protéine $A$ liées respectivement par le site $x$ de l'une et le site $y$ de l'autre.

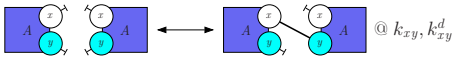
Figure 6.1: Les quatre sortes de complexes biochimiques du cas d'étude sur les symétries.



(a) Règles de liaison/déliaison entre les sites $x$ de deux occurrences de la protéine $A$ dont le site $y$ est libre.



(b) Règles de liaison/déliaison entre les sites $y$ de deux occurrences de la protéine $A$ dont le site $x$ est libre.



(c) Règles de liaison déliaison entre le site $x$ d'une occurrence de la protéine $A$ et le site $y$ d'une autre occurrence de la protéine $A$, sous la condition que le site $y$ de la première occurrence de protéine et le site $x$ de la seconde soient tous les deux libres.

Figure 6.2: Les six règles de réécriture du cas d'étude sur les symétries.

de bord du langage Kappa, pour réduire la combinatoire de ses modèles. L'extension à des groupes de symétries arbitraires et au langage complet qui est introduite dans [62] n'est évoquée qu'en conclusion de ce chapitre.

## 6.1 Le cas d'études

Nous illustrons la notion de symétries dans Kappa à travers un exemple jouet.

### 6.1.1 Modèle

Nous considérons un example dans lequel certains sites les sites d'une protéine ont exactement les mêmes capacités d'interaction.

#### 6.1.1.1 Les constituants du modèle et ses règles d'interaction

On considère une unique sorte de protéine, $A$. On suppose que chaque occurrence de cette protéine dispose exactement de deux sites de liaison qui seront notés $x$ et $y$. Les sites de deux occurrences différentes de la protéine $A$ peuvent se lier arbitrairement de manière réversible. Il y a donc trois types de liaison, $x$-$x$, $x$-$y$ et $y$-$y$, selon les sites qui sont impliqués dans la liaison. De plus, il est fait l'hypothèse que dans chaque occurrence de cette protéine, au plus un site peut être lié à la fois. Sous ces hypothèses, il existe exactement quatre sortes de complexes biochimiques[1]. Celles-ci sont toutes les quatre dessinées en Fig. 6.1. L'occurrence d'un complexe biochimique est ainsi formée soit d'une occurrence de la protéine $A$ avec les deux sites libres (voir en Fig. 6.1(a)), soit de deux occurrences de la protéine $A$ liées par leurs deux sites $x$ (voir en Fig. 6.1(b)), par leurs deux sites $y$ (voir en Fig. 6.1(c) ou par le site $x$ de l'une et le site $y$ de l'autre (voir en Fig. 6.1(d)).

Les règles de liaison et de déliaison entre les occurrences de la protéine $A$ sont décrites en Fig. 6.2. La règle bidirectionnelle dessinée en Fig. 6.2(a) stipule que deux occurrences de la protéine $A$ dont tous les sites sont libres peuvent lier leurs sites $x$ respectifs avec la constante de réaction $k_{xx}$ et qu'un tel lien peut se défaire avec la constante de réaction $k^d_{xx}$. La règle bidirectionnelle donnée en Fig. 6.2(b) précise que deux occurrences de la protéine $A$ dont tous les sites sont libres peuvent lier leurs sites $y$ respectifs avec la constante de réaction $k_{yy}$ et qu'un tel lien peut se briser avec la constante de réaction $k^d_{yy}$. Enfin, la règle bidirectionnelle dessinée

[1] voir la section 2.2 pour la notion de complexe biochimique.

$$
\begin{cases}
\frac{\mathrm{d}P_t(q_{6,0,0,0})}{\mathrm{d}t} = \frac{2k^d_{xx}}{2}P_t(q_{4,1,0,0}) + \frac{2k^d_{yy}}{2}P_t(q_{4,0,1,0}) + k^d_{xy}\cdot P_t(q_{4,0,0,1}) - \frac{30(k_{xx}+k_{yy}+k_{xy})}{2}P_t(q_{6,0,0,0}) \\
\frac{\mathrm{d}P_t(q_{4,1,0,0})}{\mathrm{d}t} = \frac{30k_{xx}}{2}P_t(q_{6,0,0,0}) + \frac{4k^d_{xx}}{2}P_t(q_{2,2,0,0}) + \frac{2k^d_{yy}}{2}P_t(q_{2,1,1,0}) + k^d_{xy}\cdot P_t(q_{2,1,0,1}) - \frac{12(k_{xx}+k_{yy}+k_{xy})+2k^d_{xx}}{2}P_t(q_{4,1,0,0}) \\
\frac{\mathrm{d}P_t(q_{4,0,1,0})}{\mathrm{d}t} = \frac{30k_{yy}}{2}P_t(q_{6,0,0,0}) + \frac{2k^d_{xx}}{2}P_t(q_{2,1,1,0}) + \frac{4k^d_{yy}}{2}P_t(q_{2,0,2,0}) + k^d_{xy}\cdot P_t(q_{2,0,1,1}) - \frac{12(k_{xx}+k_{yy}+k_{xy})+2k^d_{yy}}{2}P_t(q_{4,0,1,0}) \\
\frac{\mathrm{d}P_t(q_{4,0,0,1})}{\mathrm{d}t} = \frac{30k_{xy}}{2}P_t(q_{6,0,0,0}) + \frac{2k^d_{xx}}{2}P_t(q_{2,1,0,1}) + \frac{2k^d_{yy}}{2}P_t(q_{2,0,1,1}) + 2k^d_{xy}\cdot P_t(q_{2,0,0,2}) - \left(\frac{12(k_{xx}+k_{yy}+k_{xy})}{2} + k^d_{xy}\right)P_t(q_{4,0,0,1}) \\
\frac{\mathrm{d}P_t(q_{2,2,0,0})}{\mathrm{d}t} = \frac{12k_{xx}}{2}P_t(q_{4,1,0,0}) + \frac{6k^d_{xx}}{2}P_t(q_{0,3,0,0}) + \frac{2k^d_{yy}}{2}P_t(q_{0,2,1,0}) + k^d_{xy}\cdot P_t(q_{0,2,0,1}) - \frac{2k_{xx}+2k_{yy}+2k_{xy}+4k^d_{xx}}{2}P_t(q_{2,2,0,0}) \\
\frac{\mathrm{d}P_t(q_{2,1,1,0})}{\mathrm{d}t} = \frac{12k_{xx}}{2}P_t(q_{4,0,1,0}) + \frac{12k_{yy}}{2}P_t(q_{4,1,0,0}) + \frac{4k^d_{xx}}{2}P_t(q_{0,2,1,0}) + \frac{4k^d_{yy}}{2}P_t(q_{0,1,2,0}) + k^d_{xy}\cdot P_t(q_{0,1,1,1}) - \frac{2k_{xx}+2k_{yy}+2k_{xy}+2k^d_{xx}+2k^d_{yy}}{2}P_t(q_{2,1,1,0}) \\
\frac{\mathrm{d}P_t(q_{2,1,0,1})}{\mathrm{d}t} = \frac{12k_{xx}}{2}P_t(q_{4,0,0,1}) + \frac{12k_{xy}}{2}P_t(q_{4,1,0,0}) + \frac{4k^d_{xx}}{2}P_t(q_{0,2,0,1}) + \frac{2k^d_{yy}}{2}P_t(q_{0,1,1,1}) + 2k^d_{xy}\cdot P_t(q_{0,1,0,2}) - \left(\frac{2k_{xx}+2k_{yy}+2k_{xy}+2k^d_{xx}}{2} + k^d_{xy}\right)P_t(q_{2,1,0,1}) \\
\frac{\mathrm{d}P_t(q_{2,0,2,0})}{\mathrm{d}t} = \frac{12k_{yy}}{2}P_t(q_{4,0,1,0}) + \frac{2k^d_{xx}}{2}P_t(q_{0,1,2,0}) + \frac{6k^d_{yy}}{2}P_t(q_{0,0,3,0}) + k^d_{xy}\cdot P_t(q_{0,0,2,1}) - \frac{2k_{xx}+2k_{yy}+2k_{xy}+4k^d_{yy}}{2}P_t(q_{2,0,2,0}) \\
\frac{\mathrm{d}P_t(q_{2,0,1,1})}{\mathrm{d}t} = \frac{12k_{yy}}{2}P_t(q_{4,0,0,1}) + \frac{12k_{xy}}{2}P_t(q_{4,0,1,0}) + \frac{4k^d_{xx}}{2}P_t(q_{0,1,1,1}) + \frac{4k^d_{yy}}{2}P_t(q_{0,0,2,1}) + 2k^d_{xy}\cdot P_t(q_{0,0,1,2}) - \left(\frac{2k_{xx}+2k_{yy}+2k_{xy}+2k^d_{yy}}{2} + k^d_{xy}\right)P_t(q_{2,0,1,1}) \\
\frac{\mathrm{d}P_t(q_{2,0,0,2})}{\mathrm{d}t} = \frac{12k_{xy}}{2}P_t(q_{4,0,0,1}) + \frac{2k^d_{xx}}{2}P_t(q_{0,1,0,2}) + \frac{2k^d_{yy}}{2}P_t(q_{0,0,1,2}) + 3k^d_{xy}\cdot P_t(q_{0,0,0,3}) - \left(\frac{2k_{xx}+2k_{yy}+2k_{xy}}{2} + 2\cdot k^d_{xy}\right)P_t(q_{2,0,0,2}) \\
\frac{\mathrm{d}P_t(q_{0,3,0,0})}{\mathrm{d}t} = \frac{2k_{xx}}{2}P_t(q_{2,2,0,0}) - \frac{6k^d_{xx}}{2}P_t(q_{0,3,0,0}) \\
\frac{\mathrm{d}P_t(q_{0,2,1,0})}{\mathrm{d}t} = \frac{2k_{xx}}{2}P_t(q_{2,1,1,0}) + \frac{2k_{yy}}{2}P_t(q_{2,2,0,0}) - \frac{4k^d_{xx}+2k^d_{yy}}{2}P_t(q_{0,2,1,0}) \\
\frac{\mathrm{d}P_t(q_{0,2,0,1})}{\mathrm{d}t} = \frac{2k_{xx}}{2}P_t(q_{2,1,0,1}) + \frac{2k_{xy}}{2}P_t(q_{2,2,0,0}) - \left(\frac{4k^d_{xx}}{2} + k^d_{xy}\right)P_t(q_{0,2,0,1}) \\
\frac{\mathrm{d}P_t(q_{0,1,2,0})}{\mathrm{d}t} = \frac{2k_{xx}}{2}P_t(q_{2,0,2,0}) + \frac{2k_{yy}}{2}P_t(q_{2,1,1,0}) - \frac{2k^d_{xx}+4k^d_{yy}}{2}P_t(q_{0,1,2,0}) \\
\frac{\mathrm{d}P_t(q_{0,1,1,1})}{\mathrm{d}t} = \frac{2k_{xx}}{2}P_t(q_{2,0,1,1}) + \frac{2k_{yy}}{2}P_t(q_{2,1,0,1}) + \frac{2k_{xy}}{2}P_t(q_{2,1,1,0}) - \left(\frac{2k^d_{xx}+2k^d_{yy}}{2} + k^d_{xy}\right)P_t(q_{0,1,1,1}) \\
\frac{\mathrm{d}P_t(q_{0,1,0,2})}{\mathrm{d}t} = \frac{2k_{xx}}{2}P_t(q_{2,0,0,2}) + \frac{2k_{xy}}{2}P_t(q_{2,1,0,1}) - \left(\frac{2k^d_{xx}}{2} + 2k^d_{xy}\right)P_t(q_{0,1,0,2}) \\
\frac{\mathrm{d}P_t(q_{0,0,3,0})}{\mathrm{d}t} = \frac{2k_{yy}}{2}P_t(q_{2,0,2,0}) - \frac{6k^d_{yy}}{2}P_t(q_{0,0,3,0}) \\
\frac{\mathrm{d}P_t(q_{0,0,2,1})}{\mathrm{d}t} = \frac{2k_{yy}}{2}P_t(q_{2,0,1,1}) + \frac{2k_{xy}}{2}P_t(q_{2,0,2,0}) - \left(\frac{4k^d_{yy}}{2} + k^d_{xy}\right)P_t(q_{0,0,2,1}) \\
\frac{\mathrm{d}P_t(q_{0,0,1,2})}{\mathrm{d}t} = \frac{2k_{yy}}{2}P_t(q_{2,0,0,2}) + \frac{2k_{xy}}{2}P_t(q_{2,0,1,1}) - \left(\frac{2k^d_{yy}}{2} + 2k^d_{xy}\right)P_t(q_{0,0,1,2}) \\
\frac{\mathrm{d}P_t(q_{0,0,0,3})}{\mathrm{d}t} = \frac{2k_{xy}}{2}P_t(q_{2,0,0,2}) - 3k^d_{xy}\cdot P_t(q_{0,0,0,3})
\end{cases}
$$

Figure 6.3: Équation maîtresse pour le cas d'étude sur les symétries.

en Fig. 6.2(c) spécifie que deux occurrences de la protéine $A$ dont tous les sites sont libres, peuvent se lier respectivement par le site $x$ de l'une et le site $y$ de l'autres avec la constante de réaction $k_{xy}$ et qu'un tel lien peut se casser avec la constante de réaction $k^d_{xy}$.

#### 6.1.1.2 Le comportement du modèle

D'un point de vue qualitatif, les sites $x$ et $y$ des occurrences de la protéine $A$ ont exactement les mêmes capacités d'interaction. On peut étudier de manière empirique ce que cela implique au niveau de la sémantique du modèle, et ce pour les deux types de sémantiques quantitatives, à savoir la sémantique stochastique et la sémantique différentielle.

**6.1.1.2.1 Équation maîtresse.** Pour regarder le comportement de la sémantique stochastique du modèle, nous proposons de nous concentrer sur la solution de son équation maîtresse [97]. Celle-ci décrit pour chaque état potentiel du système sous-jacent, l'évolution temporelle de la probabilité que le système soit dans cet état. Comme il y a quatre sortes de complexes biochimiques, l'état du système sera représenté par un vecteur de quatre entiers naturels. Ainsi la variable $q_{i,j,k,l}$ représente l'état dans lequel il y a exactement $i$ occurrence(s) de la protéine $A$ sans sites liés (voir en Fig. 6.1(a)), $j$ paires d'occurrences de la protéine $A$ liées par leurs deux sites $x$ respectifs (voir en Fig. 6.1(b)), $k$ paires d'occurrences de la protéine $A$ liées par leurs deux sites $y$ respectifs (voir en Fig. 6.1(c)), et $l$ paires d'occurrences de la protéine $A$ liées l'une par son site $x$ et l'autre par son site $y$ (voir en Fig. 6.1(d)). L'équation maîtresse nécessite une variable par état potentiel. Pour faire tenir les équations sur une seule page, il est nécessaire de ne considérer qu'un nombre très limité de nombre d'occurrences de la protéine $A$. De ce fait, le système débutera avec la probabilité 1 dans un état avec uniquement six occurrences de la protéine $A$, avec tous leurs sites libres. Ainsi, comme les règles d'interaction conservent le nombre d'occurrences des protéines, seuls seront considérés les états i,j,k,l pour lesquels $i + 2 \cdot (j + k + l) = 6$. Le système est dont constitué de vingt variables, en l'occurrence, $q_{6,0,0,0}$, $q_{4,1,0,0}$, $q_{4,0,1,0}$, $q_{4,0,0,1}$, $q_{2,2,0,0}$, $q_{2,1,1,0}$, $q_{2,1,0,1}$, $q_{2,0,2,0}$, $q_{2,0,1,1}$, $q_{2,0,0,2}$, $q_{0,3,0,0}$, $q_{0,2,1,0}$, $q_{0,2,0,1}$, $q_{0,1,2,0}$, $q_{0,1,1,1}$, $q_{0,1,0,2}$, $q_{0,0,3,0}$, $q_{0,0,2,1}$, $q_{0,0,1,2}$ et $q_{0,0,0,3}$.

L'équation maîtresse de notre cas d'étude sur les symétries est donnée en Fig. 6.3. L'évolution temporelle de la distribution de probabilité de certains états est-elle donnée en Fig. 6.4 pour les paramètres cinétiques $k_{xx} = k_{xy} = 1$, $k_{yy} = k^d_{xx} = k^d_{yy} = 2$, et $k^d_{xy} = 4$.

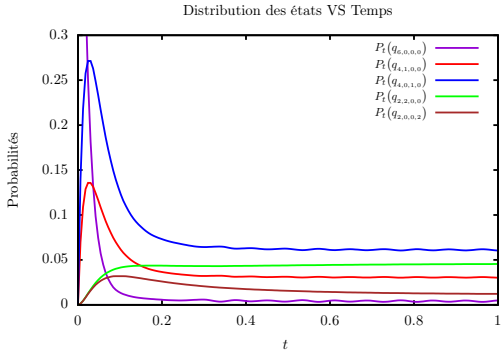



Figure 6.4: Évolution de la distribution de certains états en fonction du temps, avec les paramètres cinétiques $k_{xx} = k_{xy} = 1$, $k_{yy} = k^d_{xx} = k^d_{yy} = 2$, et $k^d_{xy} = 4$, et la distribution d'états initiale $P_0(q_{6,0,0,0}) = 1$.

**6.1.1.2.2 Sémantique différentielle.** La sémantique différentielle du modèle est définie par le systèmes d'équations différentielles donné en Fig. 6.5. Ces équations décrivent l'évolution de la concentration des différents complexes biochimique sous les hypothèses de la loi d'action de masse. Les trajectoires de ce système en prenant pour état initial, l'état où les monomères de la protéine $A$ ont pour concentration 6 et les dimers ont pour concentration 0 sont décrites en Fig. 6.6.

$$\begin{cases} \frac{\mathrm{d}[A]}{\mathrm{d}t} = 2k^d_{xy}[A.x - y.A] + 4\frac{k^d_{yy}}{2}2[A.y - y.A] + 4\frac{k^d_{xx}}{2}2[A.x - x.A] - (2\frac{k_{yy}}{2} + 2\frac{k_{xx}}{2} + 2\frac{k_{xy}}{2})[A] \\ \frac{\mathrm{d}[A.x-x.A]}{\mathrm{d}t} = \frac{k_{xx}}{2}[A]^2 - 2\frac{k^d_{xx}}{2}2[A.x - x.A] \\ \frac{\mathrm{d}[A.y-y.A]}{\mathrm{d}t} = \frac{k_{xy}}{2}[A]^2 - 2\frac{k^d_{yy}}{2}2[A.y - y.A] \\ \frac{\mathrm{d}[A.x-y.A]}{\mathrm{d}t} = \frac{k_{yy}}{2}[A]^2 - k^d_{xy}[A.x - y.A] \end{cases}$$

Figure 6.5: Système d'équations différentielles. La solution de ce système décrit l'évolution des concentrations des différents complexes biochimique en fonction du temps, sous les hypothèses de la loi d'Action de Masse.

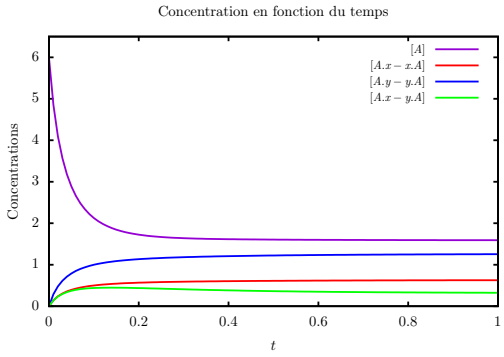



Figure 6.6: Évolution de la concentration des différents complexes biochimiques en fonction du temps, avec les paramètres cinétiques $k_{xx} = k_{xy} = 1$, $k_{yy} = k^d_{xx} = k^d_{yy} = 2$, et $k^d_{xy} = 4$, et l'état initiat $[A] = 6$ et $[A.x - x.A] = [A.x - y.A] = [A.y - y.A] = 0$.

#### 6.1.1.3 Symétries et propriétés comportementales

Sans considérer les aspects quantitatifs, les sites $x$ et $y$ des occurrences de la protéine $A$ sont équivalents. Dès qu'une règle permet de lier le site $x$ d'une occurence de la protéine $A$ à un site, une autre règle permet de lier le site $y$ de cette occurence de protéine au même site. De même, quand une règle permet de casser un lien entre le site $x$ d'une occurrence de protéine et un autre site, une autre règle permet de casser ce lien, si ce lien avait été porté par le site $y$ ce cette occurence de protéines.

Il est donc légitime de se poser la question suivante : quelles conséquences les équivalences entre sites peuvent-elles impliquer en terme de comportement du système sous-jacent et pour quelles conditions supplémentaires sur les paramètres cinétiques du modèle et son état initial (ou sa distribution d'états initiale) ?

### 6.1.2 Modèle simplifié

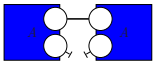
Il semble naturel de vouloir oublier la différence entre des sites équivalents. Dans notre modèle, cela revient à ignorer la différence entre les sites $x$ et $y$ dans les occurrences de la protéine $A$. À ce niveau d'abstraction, les occurrences de la protéine $A$ auront donc chacune deux sites d'interaction, désormais supposés indistingables.

#### 6.1.2.1 Complexes biochimiques et règles d'interaction.

Comme dessiné en Fig. 6.7, il ne reste alors que deux sortes de complexes biochimiques, les monomères qui sont formés d'une occurrence de la protéine $A$ avec ses deux sites de liaison libres et les dimères qui sont constitués de deux occurrences de la protéine $A$ liées exactement par une liaison entre un site de l'une et un site de l'autre.

Par ailleurs, les règles se simplifient de la même manière. Puisqu'il n'y a plus de distinctions entre les sites de liaison de chaque occurence de la protéine $A$, les trois règles de liaison peuvent se résumer en une seule, ainsi que les règles de déliaison. Les deux règles du modèle simplifié sont données en Fig. 6.8. La constante

(a) Une occurrence de la protéine $A$ avec ses deux sites libres.



(b) Deux occurrences de la protéine $A$ liées entre elles. .

Figure 6.7: Les deux sortes de complexes biochimiques dans le modèle simplifié. Le nom des sites a été retiré car ils sont maintenant supposés indistinguables.

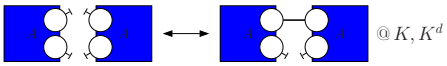



Figure 6.8: La règle bidirectionnelle du modèle simplifié. Deux occurrences de la protéine $A$ peuvent se lier à une constante de réaction $K$ (peu importe à quels sites) quand tous leurs sites sont libres. De plus, ce lien peut se rompre avec une constante de réaction $K^d$.

de réaction pour la règle de liaison est notée $K$, alors que celle pour la règle de déliaison $K^d$ (dans le modèle simplifié).

#### 6.1.2.2 États des systèmes stochastiques et différentiels sous-jacent.

Les états du système stochastique sous-jacent sont donc décrits par le nombre d'occurrences du monomère de la protéine $A$ (qui sont formées d'une occurrence de la protéine $A$) et le nombre d'occurrences du dimère de la protéine $A$ (qui sont formées de deux occurrences de la protéine $A$ liées entre-elles). Un tel état sera noté $Q_{i,j}$ avec $i$ le nombre d'occurrences de monomère et $j$ le nombre d'occurrences du dimère de la protéine $A$.

Quant à eux, les états du système différentiel sous-jacent sont décrits par deux concentrations. La concentration en monomère est notée $[A_1]$ alors que celle en dimère est notée $[A_2]$.

#### 6.1.2.3 Systèmes dynamiques sous-jacent

**6.1.2.3.1 Équation maîtresse.** L'équation maîtresse du modèle simplifié est donnée en Fig. 6.9. La solution de ce système d'équations définit donc l'évolution de la distribution de probabilités du nombre de dimers au cours du temps dans la sémantique stochastique. Seuls les états accessibles à partir de six monomères ont été considérés, c'est à dire les quatre états potentiels $Q_{6,0}$, $Q_{4,1}$, $Q_{2,2}$ et $Q_{0,3}$.

$$\begin{cases} \frac{\mathrm{d}P_t(Q_{6,0})}{\mathrm{d}t} = \frac{2K^d}{2}P_t(Q_{4,1}) - \frac{30K}{2}P_t(Q_{6,0}) \\ \frac{\mathrm{d}P_t(Q_{4,1})}{\mathrm{d}t} = \frac{30K}{2}P_t(Q_{6,0}) + \frac{4K^d}{2}P_t(Q_{2,2}) - \left(\frac{12K}{2} + \frac{2K^d}{2}\right)P_t(Q_{4,1}) \\ \frac{\mathrm{d}P_t(Q_{2,2})}{\mathrm{d}t} = \frac{12K}{2}P_t(Q_{4,1}) + \frac{6K^d}{2}P_t(Q_{0,3}) - \left(\frac{2K}{2} + \frac{4K^d}{2}\right)P_t(Q_{2,2}) \\ \frac{\mathrm{d}P_t(Q_{0,3})}{\mathrm{d}t} = \frac{2K}{2}P_t(Q_{2,2}) - \frac{6K^d}{2}P_t(Q_{0,3}) \end{cases}$$

Figure 6.9: Chemical Master Equation of the reduced model.

**6.1.2.3.2 Sémantique différentielle.** La sémantique différentielle est quant à elle donnée en Fig. 6.10. Elle donne l'évolution de la concentration en monomère, $[A_1]$, et en dimère, $[A_2]$, au cours du temps, sous les hypothèses de la loi d'action de masse.

### 6.1.3 Comparaison des dynamiques des deux modèles

#### 6.1.3.1 Quotient

Les composants de modèle simplifié ont été définis de manière intentionnelle sous la forme de graphe à sites. Quelques écarts ont été pris par rapport à la syntaxe de Kappa. En effet, celle-ci ne permet pas de répéter des sites dans l'interface d'une protéine. Cette description intentionnelle permet des raisonnements intuitifs sur le comportement des composants du modèle simplifié. Toutefois, une caractérisation extensionnelle de

$$\begin{cases} \frac{\mathrm{d}[A_1]}{\mathrm{d}t} = \frac{4K^d}{2}[A_2] - \frac{2K}{2}[A_1]^2 \\ \frac{\mathrm{d}[A_2]}{\mathrm{d}t} = \frac{K}{2}[A_1]^2 - \frac{2K^d}{2}[A_2] \end{cases}$$

Figure 6.10: système différentiel pour le modèle simplifié.

ces composante est requise pour relier formellement les composants du modèle simplifié au composants du modèle initial et ainsi justifier rigoureusement ces raisonnements. Il suffit pour cela d'interpréter les complexes biochimiques et les états du modèle simplifié respectivement comme des classes d'équivalences de complexes biochimiques et d'états du modèle initial. Les complexes biochimiques sont regroupés en deux classes, les monomères et les dimères. Ainsi, dans le cadre différentiel, les variables du modèle simplifié correspondent à la somme des concentrations des espèce biochimiques dans chaque classe d'équivalences de complexes biochimiques, comme indiqué en Fig. 6.13. De la même manière, les états du système stochastique initial peuvent être regroupés selon le nombre d'occurrences de monomère et de dimère. Du coup, à partir d'un état initial formé de 6 occurrences de la protéine $A$ aves ses deux sites sont libres, il est possible d'attendre des états dans exactement 4 classes d'équivalence, selon qu'il y ait 0, 1, 2 ou 3 dimères. Ceci peut être étendu aux distributions d'états qui apparaissent dans l'équation maîtresse : la probabilité d'avoir exactement $i$ monomères et $j$ dimères est alors vue comme la somme des probabilité des états du modèle initial ayant exactement $i$ monomères et $j$ dimères. Cette relation est exprimée en Fig. 6.12 pour les états accessibles à partir de 6 occurrences de la protéine $A$ avec les deux sites de liaison libres.

Il existe donc deux manières de calculer le comportement du modèle simplifié : soit en résolvant directement les équations du modèle simplifié, soit en résolvant le modèle initial et en calculant le comportement du modèle simplifié à l'aide de la caractérisation extensionnelle de ses composants.

En Fig. 6.14 est montrée la comparaison empirique entre ces deux méthodes pour plusieurs jeux de paramètres, que ce soit dans la cadre stochastique ou différentiel. La colonne de gauche est consacrée aux solutions des équations maîtresses alors que la colonne de droite porte sur celles des équations aux concentrations. Quatre jeux de paramètres sont considérés. Même si les paramètres n'ont pas la même signification dans le cadre stochastique et dans celui différentiel, la comparaison est faîte avec les mêmes valeurs numériques. Dans chaque graphique, les lignes continues sont obtenues en considérant le modèle initial et en projetant les résultats à l'aide de la caractérisation extensionnelle des composants du modèle simplifié alors que les courbes obtenues directement avec le modèle simplifié sont dessinées avec des tirets.

Le but de ces simulations numériques est de tester l'importance de trois contraintes. La première contrainte concerne les constantes de réaction pour les règles de déliaison. La seconde concerne l'état (ou la distribution d'états) initial(le). Enfin, la troisième concerne les constantes de réaction pour les règles de liaison.

- **Contrainte sur les constantes de déliaison.** Intuitivement, pour que les sites $x$ et $y$ des occurrences de la protéine $A$ tiennent un rôle symétrique, il est nécessaire que chaque lien puisse se défaire avec la même vitesse. Notons que les règles de déliaison pour les liens symétriques (entre deux sites $x$ ou entre deux sites $y$) peuvent s'appliquer deux fois, mais, par convention, leurs constantes de réaction sont divisées par 2. Aussi pour chaque type de lien puisse se défaire avec la même cinétique, il faut que les trois constantes $k^d_{xx}$, $k^d_{yy}$ et $k^d_{xy}$ soient égales. Enfin, pour que le modèle simplifié puisse simuler ce comportement, il faut prendre la constante $K^d$ égale à la valeur commune des trois constantes de déliaison.

- **Contrainte sur l'état initial ou la distribution initiale.** La seconde contrainte concerne l'état initial (ou la distribution des états initiaux). Elle a pour but de tester empiriquement l'importance que les sites $x$ et $y$ soient, ou non, indistingables dans la distribution initiale (ou dans l'état initial). Pour cela, sous cette hypothèse, on ne considérera que des occurrences de la protéine $A$ dont tous les sites $x$ et $y$ sont libres (puisque c'est la seule conformation de la protéine $A$ dans laquelle les deux sites sont dans le même état).

- **Contrainte sur les constantes de liaison.** La troisième contrainte concerne les règles de liaison. On peut se demander s'il est important que les liaisons entre deux occurrences libres de la protéine $A$ peuvent s'établir de manière équiprobable entre les 4 positions potentielles ($x$-$x$, $y$-$y$, $x$-$y$ et $y$-$x$). On a ici distingué les deux occurrences de la protéine $A$ dans les dimères. Il faut se rappeler que les trois constantes de réaction correspondantes sont, par convention, divisées par 2. Cependant les règles de liaison symétrique offrent chacune deux manières de créer la liaison correspondante (car les deux occurrences de

$$
\begin{cases}
\frac{\mathrm{d}P_t(q_{6,0,0,0})}{\mathrm{d}t} = \frac{2k^d_{xx}}{2}P_t(q_{4,1,0,0}) + \frac{2k^d_{yy}}{2}P_t(q_{4,0,1,0}) + k^d_{xy}\cdot P_t(q_{4,0,0,1}) - \frac{30(k_{xx}+k_{yy}+k_{xy})}{2}P_t(q_{6,0,0,0}) \\
\frac{\mathrm{d}P_t(q_{4,1,0,0})}{\mathrm{d}t} = \frac{30k_{xx}}{2}P_t(q_{6,0,0,0}) + \frac{4k^d_{xx}}{2}P_t(q_{2,2,0,0}) + \frac{2k^d_{yy}}{2}P_t(q_{2,1,1,0}) + k^d_{xy}\cdot P_t(q_{2,1,0,1}) - \frac{12(k_{xx}+k_{yy}+k_{xy})+2k^d_{xx}}{2}P_t(q_{4,1,0,0}) \\
\frac{\mathrm{d}P_t(q_{4,0,1,0})}{\mathrm{d}t} = \frac{30k_{yy}}{2}P_t(q_{6,0,0,0}) + \frac{2k^d_{xx}}{2}P_t(q_{2,1,1,0}) + \frac{4k^d_{yy}}{2}P_t(q_{2,0,2,0}) + k^d_{xy}\cdot P_t(q_{2,0,1,1}) - \frac{12(k_{xx}+k_{yy}+k_{xy})+2k^d_{yy}}{2}P_t(q_{4,0,1,0}) \\
\frac{\mathrm{d}P_t(q_{4,0,0,1})}{\mathrm{d}t} = \frac{30k_{xy}}{2}P_t(q_{6,0,0,0}) + \frac{2k^d_{xx}}{2}P_t(q_{2,1,0,1}) + \frac{2k^d_{yy}}{2}P_t(q_{2,0,1,1}) + 2k^d_{xy}\cdot P_t(q_{2,0,0,2}) - \left(\frac{12(k_{xx}+k_{yy}+k_{xy})}{2} + k^d_{xy}\right)P_t(q_{4,0,0,1}) \\
\frac{\mathrm{d}P_t(q_{2,2,0,0})}{\mathrm{d}t} = \frac{12k_{xx}}{2}P_t(q_{4,1,0,0}) + \frac{6k^d_{xx}}{2}P_t(q_{0,3,0,0}) + \frac{2k^d_{yy}}{2}P_t(q_{0,2,1,0}) + k^d_{xy}\cdot P_t(q_{0,2,0,1}) - \frac{2k_{xx}+2k_{yy}+2k_{xy}+4k^d_{xx}}{2}P_t(q_{2,2,0,0}) \\
\frac{\mathrm{d}P_t(q_{2,1,1,0})}{\mathrm{d}t} = \frac{12k_{xx}}{2}P_t(q_{4,0,1,0}) + \frac{12k_{yy}}{2}P_t(q_{4,1,0,0}) + \frac{4k^d_{xx}}{2}P_t(q_{0,2,1,0}) + \frac{4k^d_{yy}}{2}P_t(q_{0,1,2,0}) + k^d_{xy}\cdot P_t(q_{0,1,1,1}) - \frac{2k_{xx}+2k_{yy}+2k_{xy}+2k^d_{xx}+2k^d_{yy}}{2}P_t(q_{2,1,1,0}) \\
\frac{\mathrm{d}P_t(q_{2,1,0,1})}{\mathrm{d}t} = \frac{12k_{xx}}{2}P_t(q_{4,0,0,1}) + \frac{12k_{xy}}{2}P_t(q_{4,1,0,0}) + \frac{4k^d_{xx}}{2}P_t(q_{0,2,0,1}) + \frac{2k^d_{yy}}{2}P_t(q_{0,1,1,1}) + 2k^d_{xy}\cdot P_t(q_{0,1,0,2}) - \left(\frac{2k_{xx}+2k_{yy}+2k_{xy}+2k^d_{xx}}{2} + k^d_{xy}\right)P_t(q_{2,1,0,1}) \\
\frac{\mathrm{d}P_t(q_{2,0,2,0})}{\mathrm{d}t} = \frac{12k_{yy}}{2}P_t(q_{4,0,1,0}) + \frac{2k^d_{xx}}{2}P_t(q_{0,1,2,0}) + \frac{6k^d_{yy}}{2}P_t(q_{0,0,3,0}) + k^d_{xy}\cdot P_t(q_{0,0,2,1}) - \frac{2k_{xx}+2k_{yy}+2k_{xy}+4k^d_{yy}}{2}P_t(q_{2,0,2,0}) \\
\frac{\mathrm{d}P_t(q_{2,0,1,1})}{\mathrm{d}t} = \frac{12k_{yy}}{2}P_t(q_{4,0,0,1}) + \frac{12k_{xy}}{2}P_t(q_{4,0,1,0}) + \frac{4k^d_{xx}}{2}P_t(q_{0,1,1,1}) + \frac{4k^d_{yy}}{2}P_t(q_{0,0,2,1}) + 2k^d_{xy}\cdot P_t(q_{0,0,1,2}) - \left(\frac{2k_{xx}+2k_{yy}+2k_{xy}+2k^d_{yy}}{2} + k^d_{xy}\right)P_t(q_{2,0,1,1}) \\
\frac{\mathrm{d}P_t(q_{2,0,0,2})}{\mathrm{d}t} = \frac{12k_{xy}}{2}P_t(q_{4,0,0,1}) + \frac{2k^d_{xx}}{2}P_t(q_{0,1,0,2}) + \frac{2k^d_{yy}}{2}P_t(q_{0,0,1,2}) + 3k^d_{xy}\cdot P_t(q_{0,0,0,3}) - \left(\frac{2k_{xx}+2k_{yy}+2k_{xy}}{2} + 2\cdot k^d_{xy}\right)P_t(q_{2,0,0,2}) \\
\frac{\mathrm{d}P_t(q_{0,3,0,0})}{\mathrm{d}t} = \frac{2k_{xx}}{2}P_t(q_{2,2,0,0}) - \frac{6k^d_{xx}}{2}P_t(q_{0,3,0,0}) \\
\frac{\mathrm{d}P_t(q_{0,2,1,0})}{\mathrm{d}t} = \frac{2k_{xx}}{2}P_t(q_{2,1,1,0}) + \frac{2k_{yy}}{2}P_t(q_{2,2,0,0}) - \frac{4k^d_{xx}+2k^d_{yy}}{2}P_t(q_{0,2,1,0}) \\
\frac{\mathrm{d}P_t(q_{0,2,0,1})}{\mathrm{d}t} = \frac{2k_{xx}}{2}P_t(q_{2,1,0,1}) + \frac{2k_{xy}}{2}P_t(q_{2,2,0,0}) - \left(\frac{4k^d_{xx}}{2} + k^d_{xy}\right)P_t(q_{0,2,0,1}) \\
\frac{\mathrm{d}P_t(q_{0,1,2,0})}{\mathrm{d}t} = \frac{2k_{xx}}{2}P_t(q_{2,0,2,0}) + \frac{2k_{yy}}{2}P_t(q_{2,1,1,0}) - \frac{2k^d_{xx}+4k^d_{yy}}{2}P_t(q_{0,1,2,0}) \\
\frac{\mathrm{d}P_t(q_{0,1,1,1})}{\mathrm{d}t} = \frac{2k_{xx}}{2}P_t(q_{2,0,1,1}) + \frac{2k_{yy}}{2}P_t(q_{2,1,0,1}) + \frac{2k_{xy}}{2}P_t(q_{2,1,1,0}) - \left(\frac{2k^d_{xx}+2k^d_{yy}}{2} + k^d_{xy}\right)P_t(q_{0,1,1,1}) \\
\frac{\mathrm{d}P_t(q_{0,1,0,2})}{\mathrm{d}t} = \frac{2k_{xx}}{2}P_t(q_{2,0,0,2}) + \frac{2k_{xy}}{2}P_t(q_{2,1,0,1}) - \left(\frac{2k^d_{xx}}{2} + 2k^d_{xy}\right)P_t(q_{0,1,0,2}) \\
\frac{\mathrm{d}P_t(q_{0,0,3,0})}{\mathrm{d}t} = \frac{2k_{yy}}{2}P_t(q_{2,0,2,0}) - \frac{6k^d_{yy}}{2}P_t(q_{0,0,3,0}) \\
\frac{\mathrm{d}P_t(q_{0,0,2,1})}{\mathrm{d}t} = \frac{2k_{yy}}{2}P_t(q_{2,0,1,1}) + \frac{2k_{xy}}{2}P_t(q_{2,0,2,0}) - \left(\frac{4k^d_{yy}}{2} + k^d_{xy}\right)P_t(q_{0,0,2,1}) \\
\frac{\mathrm{d}P_t(q_{0,0,1,2})}{\mathrm{d}t} = \frac{2k_{yy}}{2}P_t(q_{2,0,0,2}) + \frac{2k_{xy}}{2}P_t(q_{2,0,1,1}) - \left(\frac{2k^d_{yy}}{2} + 2k^d_{xy}\right)P_t(q_{0,0,1,2}) \\
\frac{\mathrm{d}P_t(q_{0,0,0,3})}{\mathrm{d}t} = \frac{2k_{xy}}{2}P_t(q_{2,0,0,2}) - 3k^d_{xy}\cdot P_t(q_{0,0,0,3})
\end{cases}
$$

Figure 6.11: Équation maîtresse.

$$\begin{cases} P_t(Q_{6,0}) = P_t\big(q_{6,0,0,0,0}\big) \\ P_t(Q_{4,1}) = P_t\big(q_{4,1,0,0,0}\big) + P_t\big(q_{4,0,1,0,0}\big) + P_t\big(q_{4,0,0,0,1}\big) \\ P_t(Q_{2,2}) = P_t\big(q_{2,2,0,0,0}\big) + P_t\big(q_{2,1,1,0,0}\big) + P_t\big(q_{2,1,0,0,1}\big) + P_t\big(q_{2,0,2,0,0}\big) + P_t\big(q_{2,0,1,1,0}\big) + P_t\big(q_{2,0,0,0,2}\big) \\ P(Q_{0,3}) = P_t\big(q_{0,3,0,0,0}\big) + P_t\big(q_{0,2,1,0,0}\big) + P_t\big(q_{0,2,0,0,1}\big) + P_t\big(q_{0,1,2,0,0}\big) + P_t\big(q_{0,1,1,0,1}\big) + P_t\big(q_{0,1,0,0,2}\big) \\ \qquad\qquad + P_t\big(q_{0,0,3,0,0}\big) + P_t\big(q_{0,0,2,0,1}\big) + P_t\big(q_{0,0,1,0,2}\big) + P_t\big(q_{0,0,0,0,3}\big) + P_t\big(q_{0,0,0,1,2}\big) + P_t\big(q_{0,0,1,0,2}\big). \end{cases}$$

Figure 6.12: Définition extensionnelle des variables de l'équation maîtresse du modèle simplifié.

$$\begin{cases} [A_1] = [A] \\ [A_2] = [A.x - x.A] + [A.y - y.A] + [A.x - y.A] \end{cases}$$

Figure 6.13: Définition extensionnelle des variables du modèle simplifié pour le cadre différentiel.

la protéine $A$ jouent exactement le même rôle dans la règle), alors que la règle de liaison asymétrique peut établir deux liaisons différentes si l'on distingue les deux occurrences de la protéine $A$ auxquelles on applique la règle de liaison. De ce fait, la troisième contrainte spécifie que les constantes $k_{xx}$ et $k_{yy}$ sont égales et que la constante $k_{xy}$ est égale au double de la contrainte $k_{xx}$. Par ailleurs, pour simuler ces règles, la constante $K$ sera choisie comme la somme des trois constantes $k_{xx}$, $k_{yy}$ et $k_{xy}$. En effet, les trois règles de liaison du modèle initial, ainsi que celle du modèle simplifié peut s'appliquer dans deux positions chacune par paire d'occurrences libres de la protéine $A$, et d'autre part ces quatre constantes sont toutes divisées par 2 par convention.

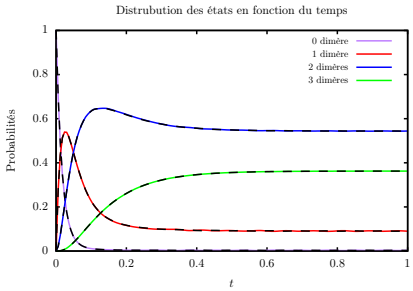
La correspondance entre le modèle initial et le modèle simplifiée est testée, pour la sémantique stochastique et pour la sémantique différentielle, sous quatre scénarios. Dans le premier, les trois contraintes de symétrie sont satisfaites, alors que les autres scénarios invalident exactement une de ces contraintes.

La description de chaque scénario et des résultats obtenus est donnée ci-dessous :

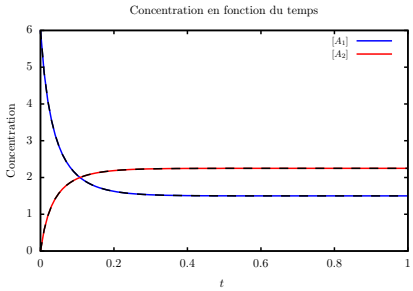
1. Dans le premier jeu de paramètres, les trois contraintes sont satisfaites. Les constantes $k_{xx}$ et $k_{yy}$ sont fixées à 1, alors que les constantes $k_{xy}$, $k^d_{xx}$, $k^d_{yy}$ et $k^d_{xy}$ sont fixées à 2. Le système stochastique sous-jacent débute de l'état formé de 6 occurrences de la protéine $A$ sans lien, et le système différentiel de l'état où les monomères sont présent en concentration 6, et les dimères absents. Les courbes montrent que la distribution des états du modèle simplifié correspondent exactement (voir en Fig. 6.14(a)) : les mêmes valeurs sont obtenues en calculant l'évolution de la distribution des états dans le modèle initial et en regroupant les états selon le nombre d'occurrences de dimère (courbes continues) ou en calculant la distribution des états dans le modèle simplifié directement (tirets). Le même phénomène se retrouve dans les systèmes différentiels sous-jacents. Les courbes en Fig. 6.14(b) montrent que les mêmes concentrations sont obtenues en regroupant la concentration des dimères dans le modèle initial ou en calculant directement cette concentration dans le modèle simplifié.

2. Dans le second jeu de paramètres, la contrainte sur les constantes de dissociation n'est plus satisfaite. Les constantes d'association et la distribution des états initiaux (ou l'état initial dans le cadre différentiel) sont gardées telles quelles, ainsi que les constantes de dissociation $k^d_{xx}$ et $k^d_{yy}$. Mais la constante de dissociation $k^d_{xy}$ est maintenant fixée à 4. Les trois constantes de déliaison du modèle initial n'étant pas égales, il n'y a pas de manière intuitive de fixer la constante de déliaison du modèle simplifié. Elle est prise arbitrairement égale à la moyenne entre la constante de déliaison des liens symétriques et celle des liens asymétriques, à savoir 3. On peut alors constater à la fois dans le cadre stochastique et dans le cadre différentiel un écart entre le comportement du modèle initial et du modèle simplifié (voir en Fig. 6.14(c) et en Fig. 6.14(d)). Une telle disparité aurait été observée quelque soit la valeur de la constante $K^d$.

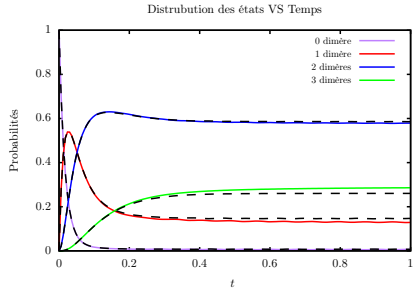
3. Le troisième jeu de paramètres a pour but de tester l'importance éventuelle de l'état initial. Par rapport au premier jeu de paramètres, seule la distribution des états initiaux (ou l'état initial dans le cadre différentiel) est modifiée. Le système stochastique débute dans un état avec deux occurrences de la protéine $A$ sans lien, et deux occurrences du dimère symétrique de la protéine $A$ formées par des liens sur les sites $x$. Quant au système différentiel, il démare d'un état où les monomères et les dimères $x-x$ sont en concentration 2, alors que les deux autres formes de dimère sont en concentration 0. l'évolution des distributions d'états des deux



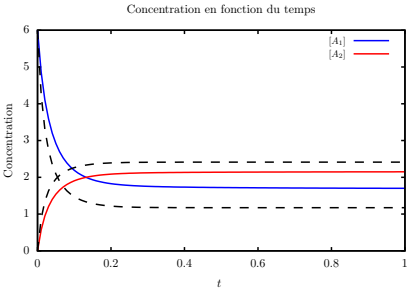

(a) $k_{xx} = k_{yy} = 1$, $k_{xy} = k_{xx}^d = k_{yy}^d = k_{xy}^d = 2$ et $P_0(q_{6,0,0,0}) = 1$ ; $K = 4$, $K^d = 2$, et $P_0(Q_{6,0}) = 1$.



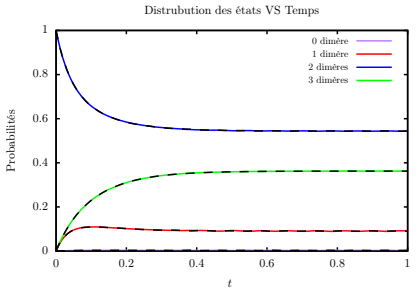

(b) $k_{xx} = k_{yy} = 1$, $k_{xy} = k_{xx}^d = k_{yy}^d = k_{xy}^d = 2$, initialement $[A] = 6$ et $[A.x{-}x.A] = [A.x{-}y.A] = [A.y{-}y.A] = 0$ ; $K = 4$, $K^d = 2$, initialement $[A_1] = 6$ et $[A_2] = 0$.



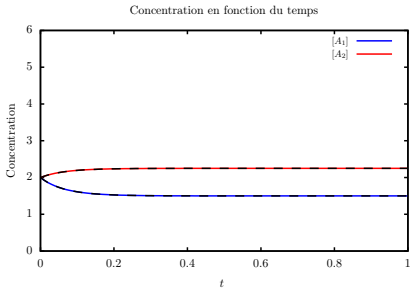

(c) $k_{xx} = k_{yy} = 1$, $k_{xy} = k_{xx}^d = k_{yy}^d = 2$, $k_{xy}^d = 4$ et $P_0(q_{6,0,0,0}) = 1$ ; $K = 4$, $K^d = 3$ et $P_0(Q_{6,0}) = 1$.



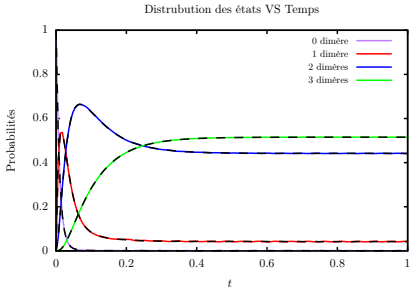

(d) $k_{xx} = k_{yy} = 1$, $k_{xy} = k_{xx}^d = k_{yy}^d = 2$, $k_{xy}^d = 4$, initialement $[A] = 6$ sans dimère ; $K = 4$, $K^d = 3$, initialement $[A_1] = 6$ et $[A_2] = 0$.



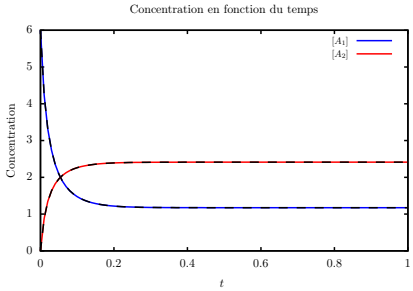

(e) $k_{xx} = k_{yy} = 1$, $k_{xy} = k_{xx}^d = k_{yy}^d = k_{xy}^d = 2$, $P_0(q_{2,2,0,0}) = 1$ ; $K = 4$, $K^d = 2$, $P_0(Q_{2,2}) = 1$.




(f) $k_{xx} = k_{yy} = 1$, $k_{xy} = k_{xx}^d = k_{yy}^d = k_{xy}^d = 2$, initialement $[A] = [A.x{-}x.A] = 2$ et $[A.y{-}y.A] = [A.x{-}y.A] = 0$ ; $K = 4$, $K^d = 2$, initialement $[A_1] = [A_2] = 2$.




(g) $k_{xx} = 1$, $k_{yy} = 2$, $k_{xy} = 4$, $k_{xx}^d = k_{yy}^d = k_{xy}^d = 2$ et $P_0(q_{6,0,0,0}) = 1$ ; $K = 7$, $K^d = 2$ et $P_0(Q_{6,0}) = 1$.




(h) $k_{xx} = 1$, $k_{yy} = 2$, $k_{xy} = 4$, $k_{xx}^d = k_{yy}^d = k_{xy}^d = 2$ initialement $[A] = 6$ sans dimère ; $K = 7$, $K^d = 2$, initialement $[A_1] = 6$ et $[A_2] = 0$.

Figure 6.14: Comparaison entre le comportement du modèle initial (traits continus) et celui du modèle simplifié (tirets). à gauche, évolution de la distribution de probabilités du nombre de dimères pour différents paramètres cinétiques et différentes distributions d'états initiaux dans le système stochastique sous-jacent. à droite, évolution de la concentration en dimère pour différents paramètres cinétiques et différentes concentrations initiales.

systèmes stochastiques sous-jacents est dessinée en Fig. 6.14(e), alors que l'évolution des concentrations en monomère et en dimère dans les deux systèmes différentiels sous-jacents est donnée en Fig. 6.14(f). À la fois, le cadre stochastique et dans le cadre différentiel, les courbes coïncident entre le modèle initial et le modèle simplifié, ce qui suggère que la correction de la simplification du modèle ne dépend pas de la distribution initiale (ou de l'état initial dans le cadre stochastique).

4. Enfin, le quatrième jeu de paramètres vise à tester l'impact de la contrainte sur les constantes d'association. Les paramètres sont les mêmes que dans le premier jeu, sauf la constante $k_{yy}$ qui est fixée à 2 et la constante $k_{xy}$ qui est fixée à 4. De plus, la constante d'association dans le modèle simplifié, $K$, reste la somme des constantes d'association dans le modèle initial, c'est à dire 7. Là encore, le comportement du modèle initial et celui du modèle simplifié coincident, que ce soit pour la sémantique stochastique ou la sémantique différentielle (voir en Fig. 6.14(g) et en Fig. 6.14(h)).

Ainsi, si l'égalité entre les constantes de dissociation semble cruciale pour pouvoir simplifier le modèle de manière exacte, le choix de la distribution initiale (ou de l'état initial) et des constantes d'association ne semblent pas importantes. Pour les constantes d'association, ceci s'explique par la simplicité du cas d'étude, puisque que les règles de liaison ne lient que des monomères et que ces monomères sont invariants par échange de leurs sites $x$ et $y$. De manière général, les conditions sur les constantes cinétique portent sur des règles qui effectuent des actions similaires sur des motifs équivalents à permutation de sites près. La correction du modèle simplifiée peut alors s'expliquer par une bisimulation en-avant [23] sur l'espace des états du système stochastique sous-jacent ou de son analogue différentiel [29].

#### 6.1.3.2 Invariants quantitatifs

Dans le paragraphe 6.1.2, le fait que certains sites puissent partager exactement les mêmes capacités d'interaction a été exploité pour simplifier un modèle jouet de manière exacte. Les composés du modèle simplifié ont été reliés formellement à ceux du modèle initial et les mêmes résultats ont été obtenus en exécutant le modèle initial avant d'utiliser ces relations formelles pour en déduire l'évolution des composants du modèle simplifié ou en exécutant directement le modèle simplifié.

Une autre classe de propriétés intéressante est celle des invariants quantitatifs. Il semble naturel de supposer que lorsque deux sites sont équivalents dans toutes les occurrences d'une protéine, alors, dans le cadre stochastique, les états obtenus en permutant ces deux sites dans une ou plusieurs occurrences de cette protéine dans un état donné soient équiprobables. De même, dans le cadre différentiel, la concentration des espèces biochimiques obtenues en permutant ces deux sites dans une ou plusieurs occurrences de cette protéine dans un complexe biochimique devrait être égales. En fait, ces quantités ne sont en pratique pas égales, mais plutôt proportionnelles. Pour mieux comprendre ce phénomène, nous pouvons étudier ce qui se passe dans un jeu de "pile" ou "face" avec deux pièces. Lorsque les deux pièces sont jetées, elles tombent toutes deux sur le côté "pile" avec une probabilité un quart, toutes deux sur le côté "face" avec une probabilité un quart et l'une sur le côté "pile" et l'autre sur le côté "face" avec une probabilité un demi. Pourtant il est possible d'intervertir les côtés "pile" et "face" d'une ou des deux pièces sans changer la nature du jeu. Les côtés "pile" et "face" ont donc un rôle symétrique. Les trois configurations ne sont pas équiprobables. Leur probabilité est proportionnelle au nombre de manière de retourner les pièces sans changer la configuration globale. Ainsi, la configuration comportant les deux côtés "pile" et "face", reste inchangée lorsque les deux pièces sont retournées simultanément (c'est la seule manière non triviale de laisser cette configuration inchangée), alors qu'un tirage double est changé dès que l'on retourne au moins une des deux pièces, d'où un rapport de 2 sur 1 (en tenant compte de la transformation triviale qui consiste à ne changer aucune pièce). Ainsi, il faut tenir compte du nombre des transformations qui laissent un état ou un complexe biochimique inchangé pour connaître les rapports de proportionnalité entre les probabilités d'être dans deux états symétriques ou les concentrations de deux complexes symétriques.

Nous étudions maintenant l'importance de trois contraintes pour assurer l'invariance de ces rapports de proportionnalité. Ce sont en fait les trois contraintes testées pour l'adéquation entre le modèle initial et le modèle simplifié, mais pour des raisons un peu différentes.

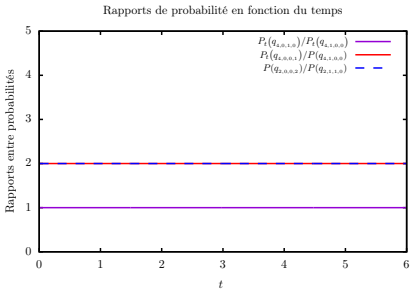
- **Contrainte sur les constantes de déliaison.** Intuitivement, pour préserver des rapports de proportionnalité entre les probabilités des états symétriques et entre les concentrations des différents types de dimères de la protéine $A$, il faut que chaque lien puisse se défaire à la même vitesse. Notons que les règles de déliaison pour les liens symétriques (entre deux sites $x$ ou entre deux sites $y$) peuvent s'appliquer deux fois, mais, par convention, leurs constantes de réaction sont divisées par 2. Aussi pour chaque type de lien puisse se défaire avec la même cinétique, il faut que les trois constantes $k^d_{xx}$, $k^d_{yy}$ et $k^d_{xy}$ soient égales.

- **Contrainte sur l'état initial ou la distribution initiale.** La seconde contrainte concerne l'état initial (ou la distribution des états initiaux). Pour que des rapports de proportionnalités se retrouvent tout au long de l'exécution du système, ils doivent également être vérifiés dans la distribution initiale (dans le cadre stochastique) ou dans l'état initial (dans le cadre différentiel). Comme ces rapports de proportionnalité ne sont pas à priori connus, le plus simple est de considérer uniquement des monomères dans les états initiaux. Ainsi, dans la version stochastique, on considérera une distribution initiale avec un état avec uniquement des occurrences de monomères avec probabilité 1 et, dans le cadre différentiel, un état initial avec uniquement des monomères.

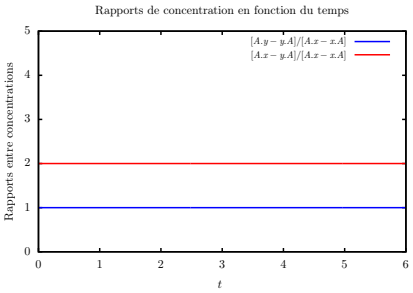
- **Contrainte sur les constantes de liaison.** La troisième contrainte concerne les règles de liaison. On peut se demander s'il est important que les liaisons entre deux occurrences libres de la protéine $A$ peuvent s'établir de manière équiprobable entre les 4 positions potentielles ($x$-$x$, $y$-$y$, $x$-$y$ et $y$-$x$). On a ici distingué les deux occurrences de la protéine $A$ dans les dimères. Il faut se rappeler que les trois constantes de réaction correspondantes sont, par convention, divisées par 2. Cependant les règles de liaison symétrique offrent chacune deux manières de créer la liaison correspondante (car les deux occurrences de la protéine $A$ jouent exactement le même rôle dans la règle), alors que la règle de liaison asymétrique peut établir deux liaisons différentes si l'on distingue les deux occurrences de la protéine $A$ auxquelles on applique la règle de liaison. De ce fait, la troisième contrainte spécifie que les constantes $k_{xx}$ et $k_{yy}$ sont égales et que la constante $k_{xy}$ est égale au double de la contrainte $k_{xx}$. Par ailleurs, pour simuler ces règles, la constante $K$ sera choisie comme la somme des trois constantes $k_{xx}$, $k_{yy}$ et $k_{xy}$. En effet, les trois règles de liaison du modèle initial, ainsi que celle du modèle simplifié peut s'appliquer dans deux positions chacune par paire d'occurrences libres de la protéine $A$, et d'autre part ces quatre constantes sont toutes divisées par 2 par convention.

Dans le cas d'étude, ces conditions doivent assurer, dans le cadre stochastique, que les probabilités d'être dans deux états obtenus en remplaçant une occurrence d'un type de dimère par une occurrence d'un autre type restent dans le même rapport de proportionnalité au cours du temps. Dans le cadre différentiel, ces conditions doivent assurer que les concentrations entre les différentes sortes de dimères restent dans le même rapport de proportionnalité aux cours du temps. En Fig. 6.15 est vérifié de manière empirique si de tels rapports de proportionnalité se manifestent pour plusieurs jeux de paramètres, que ce soit dans le cadre stochastique (colonne de gauche) ou différentiel (colonne de droite), en faisant varier la valeur des constantes de dissociations, la distribution initiale (ou l'état initial), et les constantes d'association. Dans le cadre stochastique, l'évolution de trois rapports est considéré : le rapport entre l'état composé de quatre occurrences du monomère de la protéine $A$ et d'une occurrence du dimère symétrique $x$-$x$ de la protéine $A$ et celui comprenant quatre occurrences du monomère de la protéine $A$ et une occurrence du dimère symétrique $y$-$y$ ; le rapport entre l'état composé de quatre occurrences du monomère de la protéine $A$ et d'une occurrence du dimère symétrique $x$-$x$ de la protéine $A$ et celui comportant quatre occurrences du monomère de la protéine $A$ et une occurrence du dimère asymétrique de la protéine $A$ et le rapport entre l'état composé de deux occurrences du monomère de la protéine $A$ et de deux occurrences du dimère asymétrique de la protéine $A$ et celui formé de deux occurrences du monomère de la protéine $A$, d'une occurrence de chacune des sortes de dimères symétriques de la protéine $A$. Dans le cadre différentielle, est considérée l'évolution des rapports entre les concentrations des différentes sortes des dimères de la protéine $A$.

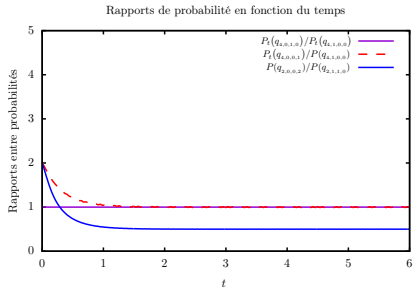
1. Dans le premier jeu de paramètres, les constantes d'association sont choisies pour que les liaisons se forment de manière équitable entre les sites $x$ et $y$ des occurrences de la protéine $A$. Les constantes de dissociation prennent toutes les trois la même valeur afin de préserver les rapports de proportionnalité quels qu'ils soient. Enfin, aucun dimère n'est présent ni dans la distribution initiale des états du système stochastique, ni dans l'état initial du système différentiel, de sortes que et la distribution initiale du système stochastique, et l'état initial du système différentiel, satisfont n'importe quels rapports de proportionnalité entre l'abondance des différentes sortes de dimères de la protéine $A$. En Fig. 6.15(a), les courbes montrent de manière empirique que la probabilité d'être dans l'état avec quatre occurrences de la protéine $A$ sous forme de monomère et deux occurrences de la protéine $A$ liées par leurs sites $x$ est toujours la même que celle d'être dans l'état avec quatre occurrences de la protéine $A$ sous forme de monomère et deux occurrences de la protéine $A$ liées par leurs sites $y$. Par ailleurs, cette probabilité est toujours la moitié de celle d'être dans l'état avec quatre occurrences de la protéine $A$ sous forme de monomère et deux occurrences de la protéine $A$ liées par le site $x$ de l'un et le site $y$ de l'autre. Ces rapports de proportionnalité correspondent à ceux déjà observer dans le jeu de "pile" ou "face". Enfin, la



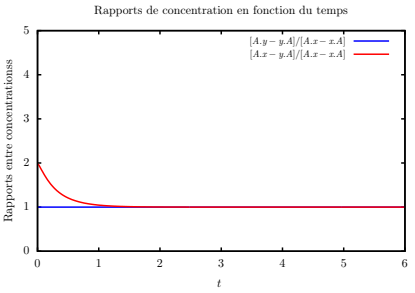

(a) $k_{xx} = k_{yy} = 1$, $k_{xy} = k^d_{xx} = k^d_{yy} = k^d_{xy} = 2$ et $P_0(q_{6,0,0,0}) = 1$.



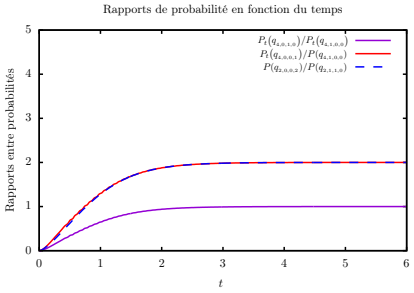

(b) $k_{xx} = k_{yy} = 1$, $k_{xy} = k^d_{xx} = k^d_{yy} = k^d_{xy} = 2$, initialement $[A] = 6$ et $[A.x - x.A] = [A.x - y.A] = [A.y - y.A] = 0$.



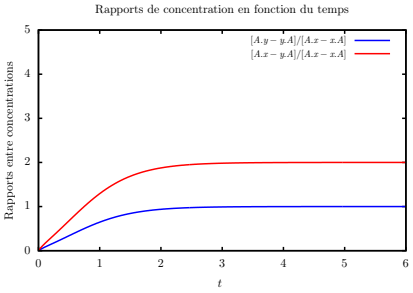

(c) $k_{xx} = k_{yy} = 1$, $k_{xy} = k^d_{xx} = k^d_{yy} = 2$, $k^d_{xy} = 4$ et $P_0(q_{6,0,0,0}) = 1$.



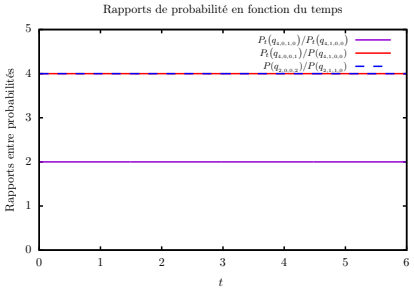

(d) $k_{xx} = k_{yy} = 1$, $k_{xy} = k^d_{xx} = k^d_{yy} = 2$, $k^d_{xy} = 4$, initialement $[A] = 6$ sans dimère.



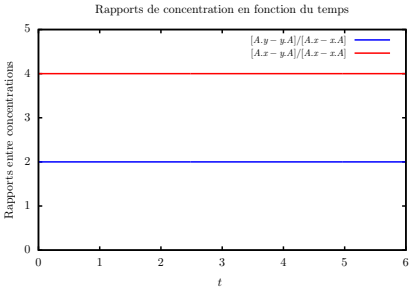

(e) $k_{xx} = k_{yy} = 1$, $k_{xy} = k^d_{xx} = k^d_{yy} = k^d_{xy} = 2$, $P_0(q_{2,2,0,0}) = 1$.




(f) $k_{xx} = k_{yy} = 1$, $k_{xy} = k^d_{xx} = k^d_{yy} = k^d_{xy} = 2$, initialement $[A] = [A.x - x.A] = 2$ et $[A.y - y.A] = [A.x - y.A] = 0$.




(g) $k_{xx} = 1$, $k_{yy} = 2$, $k_{xy} = 4$, $k^d_{xx} = k^d_{yy} = k^d_{xy} = 2$ et $P_0(q_{6,0,0,0}) = 1$.




(h) $k_{xx} = 1$, $k_{yy} = 2$, $k_{xy} = 4$, $k^d_{xx} = k^d_{yy} = k^d_{xy} = 2$ initialement $[A] = 6$ sans dimère.

Figure 6.15: Relations de proportionnalité dans le modèle initial. à gauche, évolution de rapports entre la probabilité d'être dans certains états pour différents paramètres cinétiques et différentes distributions d'états initiaux dans le système stochastique sous-jacent. à droite, évolution de rapports entre la concentration de différents complexes biochimiques pour différents paramètres cinétiques et différentes concentrations initiales.

probabilité d'être dans un état avec deux occurrences de la protéine $A$ sous forme de monomère et deux occurrences de dimères asymétriques est toujours deux fois plus grande que celle d'être dans un état avec deux occurrences de la protéine $A$ sous forme de monomère et d'une occurrence de chaque forme de dimère symétrique. Ce rapport s'explique que la formation d'un dimère asymétrique est deux fois plus probable que celle d'un type donnée de dimère symétrique (d'où un facteur 4). Cependant, il faut diviser ce facteur par 2, car deux dimères symétriques différents peuvent être obtenu en prenant le premier formé par une liaison entre deux sites $x$ et le second entre deux sites $y$, ou l'inverse, soit deux fois plus de possibilités.

En ce qui concerne le comportement de la solution du système différentiel sous-jacent, les courbes en Fig. 6.15(b) montrent que les concentrations des deux sortes de dimère symétrique restent toujours égales, alors que la concentration en dimère asymétrique est toujours égale au double de cette valeur commune.

2. Dans le second jeu de paramètres, la contrainte sur les constantes de dissociation est relachée. Les constantes de dissociations sont ainsi fixées de manière arbitraire à $k^d_{xx} = 2$, $k^d_{yy} = 2$ et $k^d_{xy} = 4$. Alors que les autres paramètres restent les mêmes que dans le premier jeu de paramètres. Les courbes dessinées en Fig. 6.15(c) montrent que la probabilité d'être dans l'état avec quatre occurrences de la protéine $A$ sous forme de monomère et deux occurrences de la protéine $A$ liées par leurs sites $x$ est toujours la même que celle d'être dans l'état avec quatre occurrences de la protéine $A$ sous forme de monomère et deux occurrences de la protéine $A$ liées par leurs sites $y$. Par contre, les autres probabilités étudiées ne sont pas proportionnelles. Initialement, la probabilité d'être dans un état avec une occurrence de dimère asymétrique et quatre occurrences de la protéine $A$ sous la forme de monomère est deux fois plus grande que celle d'être dans un état avec une occurrence de dimère asymétrique et quatre occurrences de la protéine $A$ sous la forme de monomère, mais ce rapport tends vers un avec le temps. D'autre part, la probabilité la probabilité d'être dans un état avec deux occurrences de dimère asymétrique et deux occurrences de la protéine $A$ sous la forme de monomère est deux fois plus grande au début que la probabilité d'être dans un état avec une occurrence de chacun des types de dimère symétrique et deux occurrences de la protéine $A$ sous la forme de monomère, mais ce rapport tends vers un demi avec le temps. Dans le système différentiel, la concentration des deux types de dimère symétrique est toujours la même, alors que la concentration en dimère asymétrique est initialement deux fois plus grande, pour finalement converger vers cette même valeur. Les rapports initiaux sont dictés par les constantes d'association, ce sont les mêmes que pour le premier jeu de paramètres. En revanche, les rapports à la limite sont dictés par le rapport entre les constantes respectives d'association et de dissociation (et la répétition des occurrences du dimère asymétrique pour le facteur un demi additionel).

3. Dans le troisième jeu de paramètres, c'est la distribution d'état initiale (dans le cadre stochastique) et l'état initial (dans le cadre différentiel) qui sont fixés arbitrairement. Aucun rapport de proportionnalité ne se dessine, que ce soit dans le cadre stochastique (voir en Fig. 6.15(e)) ou dans le cadre différentiel (voir en Fig. 6.15(f)). Initialement, les rapports de probabilités et de concentrations sont imposés par les abondances au début de l'exécution des deux systèmes. à la limite, ces rapports sont fixés par le rapport entre les constantes respectives d'association et de dissociation, ce qui explique qu'ils sont les mêmes que pour le premier jeu de paramètres.

4. Enfin, le quatrième jeu de paramètres est obtenu en fixant de manière arbitraire les constantes d'association. Ceci a pour effet de déplacer les rapports de probabilités et les rapports de concentrations. Par contre, comme c'était attendu, ces rapports sont constants au cours du temps, que ce soit pour la sémantique stochastique ou la sémantique différentielle (voir en Fig. 6.14(g) et en Fig. 6.14(h)).

L'invariance de ces rapports de proportionnalité, pour certains jeux de paramètres est révélateur de la présence d'une relation de groupage faible. Celle-ci permet de quotienter l'exécution d'une chaîne de Markov pour certaines distributions d'états initiales tout en restant Markovien [22]. De même, cela permet de réduire le système différentiel sous-jacent pour certains états initiaux.

Dans le cas du premier jeu de paramètres, le quotient opère même à plus bas-niveau, puisqu'il s'exprime au niveau réactionnel: à chaque réaction, peut être associée une réaction symétrique obtenue en échangeant les instances des sites $x$ et $y$ dans certaines occurrences de la protéine $A$, ce qui permet de montrer l'existence d'une bisimulation arrière [23] sur système de transition stochastique sous-jacent ou de réduire le système différentiel sous-jacent.

### 6.1.4 Symétries et réduction de modèles

Pour conclure avec cet exemple jouet, nous avons vu que des symétries pouvaient apparaître dans les complexes biochimiques, dans les états du système, dans les distribitions d'états ou même dans l'effet des règles. Ces symétries ont un impact sur notre capacité à réduire la dynamique du système. En effet, lorsque l'action des règles est symétrique, il est possible d'induire une relation d'équivalence pour quotienter l'ensemble des états, ce qui suggère l'existence d'une bisimulation avant. De plus, lorsque la distribution des états initiaux est également symétrique, alors le système non réduit comporte des invariants statistiques (le quotient entre la probabilité de deux états symétiques est égal à l'inverse du quotient de leur nombre d'automorphismes), ce qui suggère l'existence d'une bisimulation arrière. Contrairement aux raisonnements sur les rapport entre constantes de liaison et de déliaison, c'est un raisonnement compositionnel qui considère les réactions, classe de symétrie par classe de symétrie. C'est pourquoi ce type de raisonnement sera privilégié dans la suite de ce chapitre.

## 6.2 Permutations de sites dans des graphes à sites

Pour rester simple, le cadre général ne sera pas décrit. Au lieu de cela, la présentation se concentre sur le cas particulier où les symétries prennent la forme de permutation de sites. Le but de cette section est de définir l'effet de la permutation de sites dans les différents éléments du langage Kappa, c'est à dire les motifs, les plongements, les règles et l'application des règles.

Les permutations de sites forment un groupe. Ainsi il est possible de les composer et chaque permutation admet une permutation inverse. De plus, ce groupe est engendré par les transpositions de sites (qui consistent à échanger uniquement deux sites). Aussi, seul l'effet des transpositions de sites sera décrit, l'effet des autres permutations pouvant être déduit en décomposant ses permutations en séquences de transpositions.

### 6.2.1 Transposition de deux sites dans un complexe biochimique

(a) Transposition des sites dans espèces biochimiques formées d'une seule instance de la protéine $A$.

(b) Transposition des sites de l'occurrence gauche de la protéine $A$ dans les espèces biochimiques formées de deux instances de la protéine $A$.

(c) Transposition des sites de l'occurrence droite de la protéine $A$ dans les espèces biochimiques formées de deux instances de la protéine $A$.

(d) Transposition des sites des deux occurrences de la protéine $A$ dans les espèces biochimiques formées de deux instances de la protéine $A$.

Figure 6.16: Effet des transpositions de site sur les espèces biochimiques du modèle jouet.

**Exemple 6.2.1** *En Figs. 6.16 et 6.17 sont montrés des exemples de transpositions de sites dans des graphes à sites. L'action d'une permutation de site qui transforme un graphe à site en un autre est notée par une double flèche ondulée colorée. Les couleurs représentent différentes sortes de permutation. Dans le cas particulier des transpositions de sites, cette flèche est bidirectionnelle puisque cette action est involutive (appliquer deux fois une transposition revient à ne rien changer). L'effet des transpositions sur les graphes à site est dessiné en échangeant la couleur et le nom des sites qui sont permutés, sans changer leurs états de liaison (les sites n'ont pas d'états d'activation dans ce modèle, sinon eux-aussi auraient été conservés).*

*En Fig. 6.16, les transpositions de sites sont appliquées aux complexes biochimiques de l'exemple étudié dans ce chapitre (voir en Fig. 6.2). En Fig. 6.16(a) est montré l'effet des transpositions de sites sur les espèces biochimiques qui ne sont formées que d'une instance de la protéine A. Il n'y a donc qu'une transposition possible, puisqu'il n'y a qu'une instance de la protéine A. Cette transposition est dessinée en orange. Les deux sites de l'instance de la protéine A sont nécessairement libres, sinon ce ne serait pas un monomère. Du coup, la transposition des deux sites laisse la configuration de l'espèce biochimique inchangée (puisqu'en effet la position des sites dans une instance de protéine n'a pas de signification particulière en Kappa).*

*Pour les espèces chimiques formées de deux instances de la protéine A, il est possible de permuter les sites de l'instance de gauche, les sites de l'instance de droite, ou les deux. En Fig. 6.16(b) est montré, en vert, l'effet de la transposition des sites de gauche. En Fig. 6.16(c) est montré, en rouge, l'effet de la transposition des sites de droite. Enfin en Fig. fig:trans:dimboth est montré, en marron, l'effet de la combinaison des deux transpositions. L'ordre n'a pas d'influence sur le résultat. Hormis dans le cas où les sites de deux occurrences de la protéine A sont échangées dans le dimer formé par une liason entre le site x d'une occurrence de la protéine A et le site y de l'autre occurrence de la protéine A, l'application des transpositions de sites changent l'espèce biochimique. Essentiellement, les liaisons entre deux sites identiques deviennent des liaisons entre deux sites différents lorsque les sites sont évhangés dans une seule instance de la protéine A (voir en Figs. 6.16(b) et 6.16(c)), alors que réciproquement, les liaisons entre deux sites différents deviennent des liaisons entre deux sites identiques (le site en question dépend de quel côté la transposition est appliquée). Lorsque les transpositions de sites sont appliquées simultanément aux deux occurrences de la protéine A (voir en Fig. 6.16(d)), l'effet est de transformer un dimer formé par une liaison entre les deux sites $x^d$es instances de la protéine A en un dimer formé par une liaison entre les deux sites $y^d$es instances de la protéine A, et réciproquement. Dans le cas d'un dimer formé d'une liaison entre le site x d'une occurrence de la protéine A et le site y de l'autre occurrence, le rôle des deux instances est échangé. Ceci ne change pas l'espèce biochimique. Le résultat de la transformation est isomorphe au complexe biochimique initial.*

### 6.2.2 Transposition de deux sites dans un motif

Il est possible d'échanger deux sites dans une instance de protéine dans un motif. Pour cela il suffit de choisir une instance de protéine dans un motif, et deux sites dans l'interface de la sorte de protéine correspondante. Si l'instance de protéine ne contient aucun de ces deux sites, alors le motif reste tel qu'il est. Si l'instance de protéine contient exactement un des deux sites, alors celui-ci est remplacé par l'autre site en gardant ses éventuels états d'activation et de liaison. Enfin, si l'instance de protéine contient les deux sites, alors l'un est remplacé par l'autre, et réciproquement. De ce fait, les deux sites échangent leurs éventuels états d'activation et de liaison. Le motif reste inchangé si les deux sites avaient les mêmes états.

### 6.2.3 Restriction d'une transposition de deux sites au domaine d'un plongement

### 6.2.4 Effet d'une transposition de sites sur une règle

### 6.2.5 Effet d'une transposition de sites sur une réaction induite par une règle

Now we define the restriction of a permutation of sites that can be applied to the image of an embedding, into a permutation of sites that can be applied to the domain of this embedding.

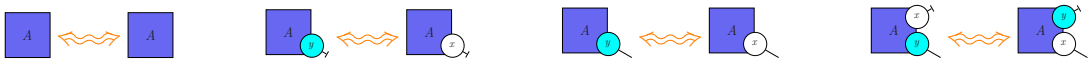
**Definition 6.2.1 (permutation propagation)** *Let $h$ be an embedding between two site graphs $G$ and $H$. Let $\sigma_H \in [H]$ be a permutation of sites that can be applied to the site graph $H$.*

*We define $h.'\sigma_H$ as the permutation of sites $(f_{h_n})_{n\in[G]}$, which can be applied to the site graph $G$.*

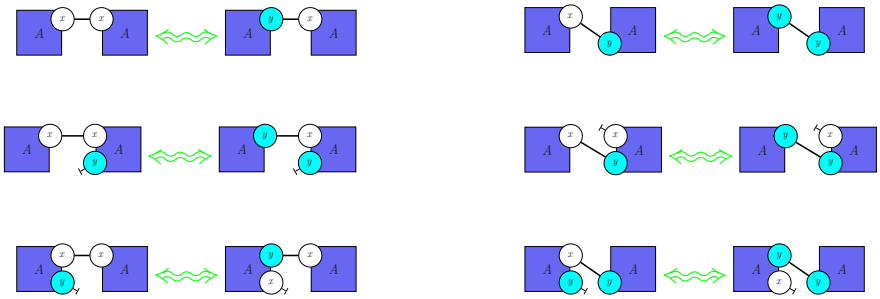
We are left to define the action of a permutation of sites over an embedding. Since permutations of sites preserve the set of agents of the site graphs to which they are applied (only the set of sites and the states of these sites may change), we can define the image of an embedding between two site graphs, as the embedding between the image of these two site graphs, that is induced by the same function over agents.

**Propriété 6.2.1** *Let $h$ be an embedding between two site graphs $G$ and $H$. Let $\sigma_H \in [H]$ be a permutation of the sites of the site graph $H$.*

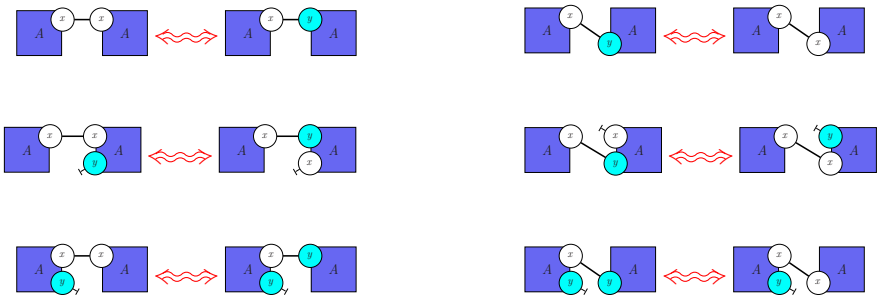
*The function between $[G]$ and $[H]$ that induces the embedding $h$, also induces an embedding between the site graph $(h.'\sigma_H).G$ and the site graph $\sigma_H.H$.*



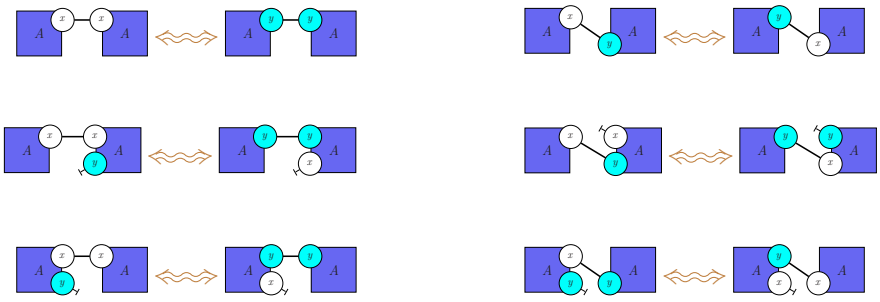
(a) Transposition des sites sur les motifs avec une seule instance de la protéine $A$ partiellement définie.



(b) Transposition des sites de l'occurrence gauche de la protéine $A$ dans les espèces biochimiques formées de deux instances de la protéine $A$ partiellement définies.



(c) Transposition des sites de l'occurrence droite de la protéine $A$ dans les espèces biochimiques formées de deux instances de la protéine $A$ partiellement définies.



(d) Transposition des sites des deux occurrences de la protéine $A$ dans les espèces biochimiques formées de deux instances de la protéine $A$ partiellement définies.

Figure 6.17: Effet des transpositions de sites dans les motifs connexes qui ne sont pas des espèces biochimiques.

**Definition 6.2.2 (action of a permutation over an embedding)** *Let $h$ be an embedding between two site graphs $G$ and $H$. Let $\sigma_H \in [[H]]$ be a permutation of the sites of the graph $H$.*

*We define $\sigma_H.''h$ as the embedding between the site graph $(h.'\sigma_H).G$ and $\sigma_H.H$ that is induced by the same function between $[G]$ and the site graph $[H]$ as $h$.*

Now we give more details about the example in Fig. **??**.

**Exemple 6.2.2** *We illustrate Defs. 6.2.1 and 6.2.2 in Fig.* **??**. *The signature is the same as in Ex.* **??**: *there are two kinds of agents $A$ and $B$. The agents of type $A$ have three binding sites $x$, $y$ and $z$, while the agents of type $B$ have three binding sites $u$, $v$, $w$.*

*We consider an embedding $h$ between two site graphs $G$ and $H$, where:*

1. *the site graph $G$ is made of an agent $1$ of type $A$ in which the site $x$ is bound (with no further information), the site $y$ is free, and we have no information about the binding state of the site $z$.*
2. *the site graph $H$ is made of two agents: the agent $2$ of type $A$ and the agent $3$ of type $B$. The site $y$ of the agent $2$ and the site $v$ of the agent $3$ are free. There is a link between the site $x$ of the agent $2$ and the site $w$ of the agent $1$ and another link between the site $z$ of the agent $2$ and the site $u$ of the agent $3$.*
3. *the embedding $h$ maps the agent of the site graph $G$ to the agent $2$ of the site graph $H$.*

*We consider the permutation of sites $\sigma_3 \in [H]$ that is defined as $\sigma_3 := (f_n)_{n\in\{2,3\}}$ where:*

- $f_2 = [x \mapsto z, y \mapsto y, z \mapsto x]$;
- $f_3 = [u \mapsto v, v \mapsto u, w \mapsto w]$.

*Thus the action of the permutation of sites $\sigma_3$ on the site graph $G$ swaps the states of the sites $x$ and $z$ in agent $2$ and the states of the sites $u$ and $v$ in agent $3$. As a result the site $y$ of the agent $2$ and the site $u$ of the agent $3$ become free. The site $x$ of agent $2$ and the site $v$ of the agent $3$ are now bound together, as well as the site $z$ of the agent $2$ and the site $w$ of the agent $3$.*

*The permutation of sites $\sigma_3$ can be restricted to the domain of the embedding $h$. By Def. 6.2.2, the resulting permutation of sites $h.'\sigma_3$ is the permutation in $[G]$ that is defined by $h.'\sigma_3 = (f'_1)$ where:*

$$f'_1 = f_{h(1)}.$$

*It follows that:*

$$f'_1 = [x \mapsto z, y \mapsto y, z \mapsto x].$$

*Thus, the permutation of sites $h.'\sigma_3$ swaps the states of the site $x$ and $z$ in the unique agent of the site graph $G$. It follows that in the unique agent of the site graph $(h.'\sigma_3).G$, the binding state of the site $x$ is not specified, the site $y$ is free, and the site $z$ is bound (with no further information). We can check that the embedding $\sigma_3.''h$ embeds the site graph $(h.'\sigma_3).G$ into the site graph $\sigma_3.H$.*

**Propriété 6.2.2** *The tuple $[] = (([G]), ., .', .'')$ is a valid set of transformations over site graphs (as defined in Defs.* **??** *and* **??**).

*Moreover each embedding is $[]$-compatible (as defined in Def.* **??**).

### 6.2.6 Retour sur le cas d'étude

Pour conclure ce chapitre, cette approche est appliquée sur le modèle jouet qui avait été introduit Sec. 6.1.1.1 page 52.

Dans ce modèle, quite à intervertir les sites $x$ et $y$ dans une ou plusieurs occurrences de la protéine $A$, les règles se répartissent en deux catégories : les règles d'association et les règles de dissociation. Ces deux catégories sont dessinées en Fig. 6.19, ainsi que les permutations de sites qui permettent de passer d'une règle à une autre au sein de chaque catégorie. Dans chacune des ces deux catégories, deux règles correspondent à la même, quite à échanger le rÃ´le des deux occurrences de la protéine $A$. Il s'agit des deux règles d'association asymétrique pour lier le site $x$ d'une des occurrences de la protéine $A$ au site $y$ de l'autre occurrence de la protéine $A$ (les deux règles sur la ligne du milieu en Fig. 6.19(a)) et des deux règles de dissociation asymétriques

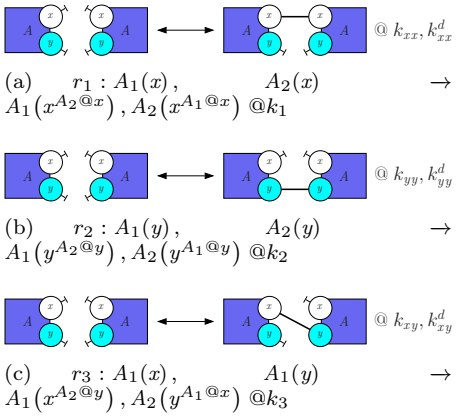



Figure 6.18: Under which conditions are the sites $x$ and $y$ equivalent in each occurrence of the protein $A$ ?

(les deux règles sur la ligne du milieu en Fig. 6.19(b)). Ainsi, pour que les sites $x$ et $y$ aient un rÃ´le équivalent dans le modèle, il suffit que les constantes corrigées des quatre règles d'association soient égales et que celles des quatre règles de dissociation le soient également. Ainsi, en tenant compte, du coefficient de correction des constantes de réaction et des répétitions des règles au sein de chaque catégorie de règles, cela donne les contraintes suivantes :

$$\frac{k_{xx}}{2 \cdot 1} = \frac{k_{yy}}{2 \cdot 1} = \frac{k_{xy}}{2 \cdot 2}$$

et

$$\frac{k_{xx}}{2 \cdot 1} = \frac{k_{yy}}{2 \cdot 1} = \frac{k_{xy}}{1 \cdot 2}.$$

Dans ces fractions, le premier facteur des dénominateurs représente la correction des constantes de réaction alors que le deuxième facteur correspond au nombre de formes de la règle (modulo permutation des agents) qui apparaissent dans les catégories de règles.

Il s'en suit les contraintes suivantes:

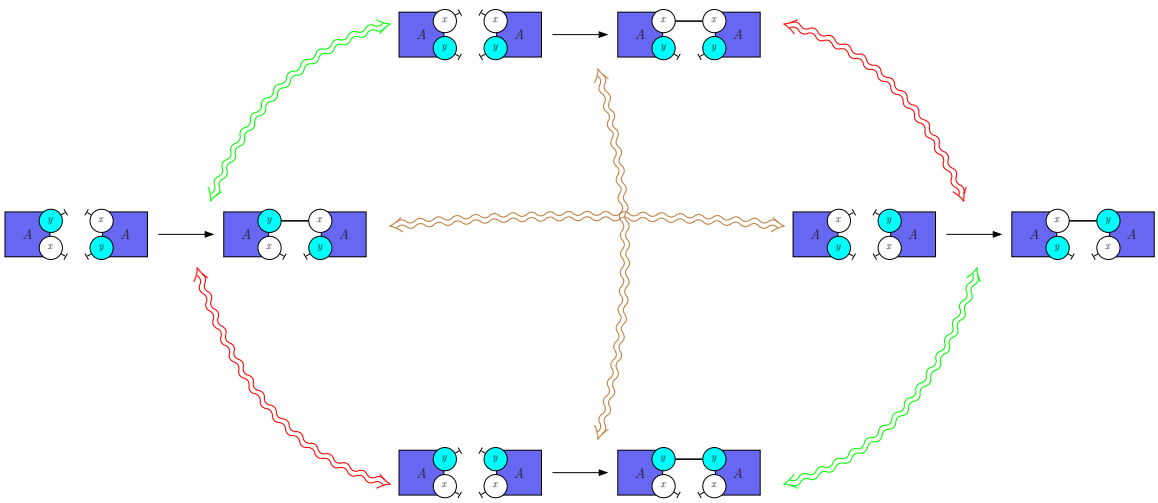
$$\begin{cases} k_{xx} = kyy \\ k_{xy} = 2k_{xx} \\ k^d_{xx} = k^d_{yy} \\ k^d_{xx} = k^d_{xy} \end{cases}$$

Sous ces conditions, les sites $x$ et $y$ sont équivalents, ce qui induit une bisimulation dans les deux sens sur à la fois sur le systèmes stochastique sous-jacent et le système différentiel sous-jacent.

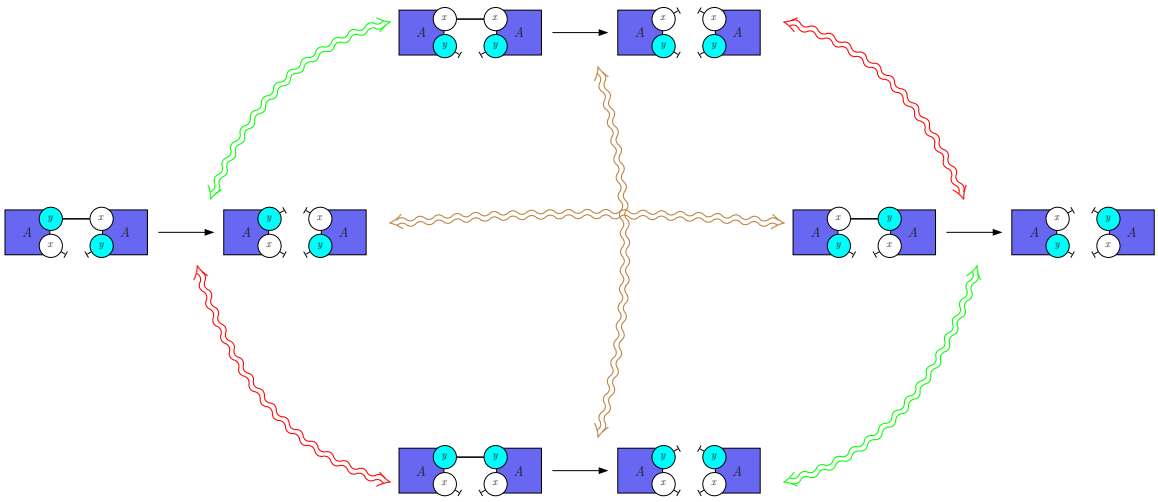
Cette bisimulation permet de réduire ces systèmes en oubliant la différence entre les sites $x$ et $y$, et ce quel que soit la distribution initiale des états (dans le cadre stochastique) ou les concentrations initiales (dans le cadre différentiel).

Par ailleurs, cette bisimulation permet de caractériser des sous-espaces de distributions d'états (dans le cadre stochastique) et des sous-espaces de concentrations (dans le cas différentiel) stables.

L'examen des différents complexes et de leur relation vis à vis des permutations de sites équivalent permet de trouver un sous-espace stable de concentrations. Quitte à permuter l'état des sites $x$ et $y$ dans une ou plusieurs instances de la protéine $A$, l'ensemble de tous les complexes biochimiques du modèle jouet se classe en deux catégories : les monomères et les dimères. Celles-ci sont ressinées en Fig. 6.20 sous la forme de deux orbites. Les permutations qui permettent de passer d'un complexe biochimique à un autre au sein de chaque catégorie sont représentées par des doubles flèches ondulées. PlutÃ´t que d'intervertir l'état des sites, l'action de ses permutations a été représenté en échangeant, de manière équivalente, les sites. Étant donné que l'ordre des sites dans une occurrence de protéine n'a pas de signification particulière en Kappa, il existe une seule forme de monomère. Leur concentration n'est donc pas contrainte dans le sous-espace stable de concentrations qui découle de l'équivalence entre les sites $x$ et $y$ dans les occurrences de la protéine $A$. Il existe en revanche trois formes de dimères : les dimères symétriques avec une liaison entre deux sites $x$ ; les dimères symétriques avec une liaison entre deux sites $y$ et les dimères asymétriques avec une liaison entre un site $x$ d'une occurrence de la

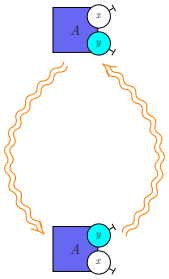


(a) Les règles d'association sous forme d'orbite.

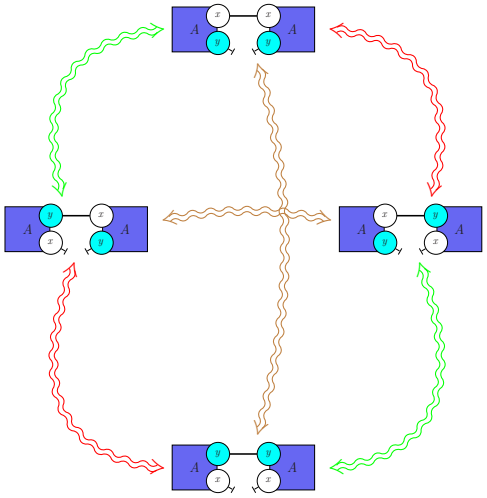


(b) Les règles de dissociation sous forme d'orbite.

Figure 6.19: Quitte à permuter les sites $x$ et $y$ dans certaines occurrences de la protéine $A$, les règles peuvent être regroupées en deux catégories. Les règles d'association sont représentées en Fig. 6.19(a), alors que les règles de dissociation sont représentées en Fig. 6.19(b). L'effet des permutations de sites est représenté par des doubles flèches ondulées. Les permutations triviales, qui consiste à ne rien changer ne sont pas représentées. La permutation des sites $x$ et $y$ dans l'occurrence gauche de la protéine est représentée en vert ; celle dans l'occurrence de droite est représentée en rouge ; la permutation simultannée des sites $x$ et $y$ dans les deux occurrences de la protéine $A$ est représentée en marron.



(a) Orbite des occurrences de monomère.



(b) Orbite des occurrences de dimère.

Figure 6.20: Quitte à permuter les sites $x$ et $y$ dans les occurrences de la protéine $A$, les complexes biochimiques se classent en deux catégories, les monomères (voir en Fig. 6.20(a)) et les dimères (voir en Fig. 6.20(b)). Pour les monomères, la seule transformation possible consiste à permuter les sites de l'unique instance de la protéine (ce qui est dessiné avec une double flèche ondulée orange). Ceci ne change par la conformation de la protéine puisque l'ordre des sites n'a pas d'importance dans le langage Kappa. Pour les dimères, le rÃ´le des sites peuvent être échangés dans l'occurrence gauche (ce qui est dessiné avec une double flèche ondulée verte), dans l'occurrence droite (rouge) ou dans les deux simultanément (noire). Quite à changer l'ordre des agents et des sites de liaisons, les deux représentations graphiques des dimères asymétriques sont équivalentes du point de vue de la sémantique du langage Kappa.

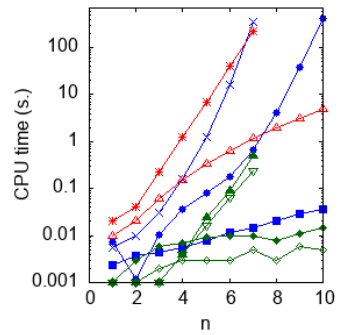
| model | sites | rules | species | | reactions | |
|---|---|---|---|---|---|---|
| | | | original | reduced | original | reduced |
| kinase/phosphatase | $n$ | $6n$ | $2+4^n$ | $2+\binom{n+3}{3}$ | $6n4^{n-1}$ | $2n\binom{n+2}{2}$ |
| multiple phosphorylation | $n$ | $n2^n$ | $2^n$ | $n+1$ | $n2^n$ | $2n$ |
| mult. phosphoryl. with counter | $n$ | $2n^2$ | $2^n$ | $n+1$ | $n2^n$ | $2n$ |

Figure 6.21: Key attributes of our models with respect to the parameter $n$.

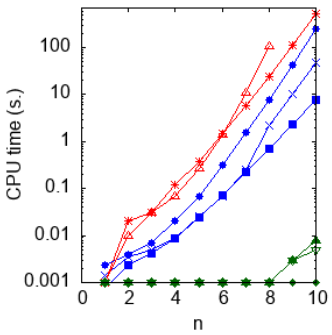
protéine $A$ et le site $y$ d'une autre occurrence de la protéine $A$. Quitte à échanger le rÃ´le des deux occurrences des protéines $A$, cette dernière forme de dimère apparaît deux fois. En conséquence, le sous-espace stable de concentrations qui découle de l'équivalence entre les sites $x$ et $y$ doit satisfaire les contraintes suivantes : les dimères asymétriques doivent constituer la moitié de la concentration totale en dimère et chacune des deux formes de dimère symétrique doit constituer un quart de la concentration totale en dimère. Il s'agit des mêmes proportions que pour le jeu de "pile" ou "face".

Dans le cadre stochastique, il faut regarder, pour déduire un sous-espace de distributions stable, les catégories d'états du système. Quite à permuter l'état des sites $x$ et $y$, dans une ou plusieurs occurrences de la protéine $A$, deux états sont équivalents s'ils contiennent le même nombre d'occurrence de monomère et le même nombre d'occurrence de dimère.

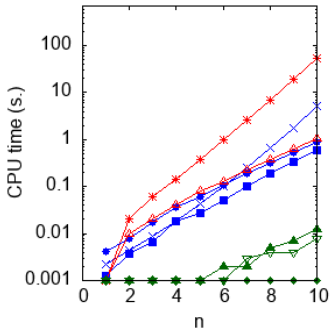
Reprenant la notation i,j,k,l pour désigner un état avec $i$ occurrences de monomère, $j$ occurrences du dimère formé d'une liaison entre deux sites $x$, $k$ occurrences du dimère formé d'une liaison entre deux sites $y$ et $l$ occurrences du dimère formé d'une liaison entre le site $x$ d'une occurrence de la protéine $A$ et le site $y$ d'une autre occurrence de la protéine $A$, deux états i',j',k',l' et i",j",k",l" sont donc équivalents losque $i' = i''$ et $j'+k'+l' = j''+k''+l''$. Reste à calculer les rapports de proportionnalité entre les probabilités de deux tels états dans le sous espace de distributions stable défini par l'équivalence entre les sites $x$ et $y$.



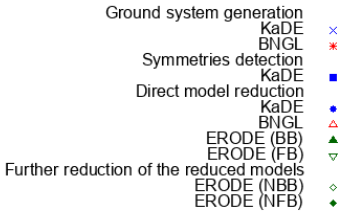

(a) kinetase/phosphatase model.




(b) multi-phosphorylation site model.




(c) multi-phosphorylation site model with counter.




(d) legend.

Figure 6.22: Comparison between the time performances of KaDe, BNGL, and ERODE, on a MacBookPro with a `2,8 GHz Intel Core i7` CPU and a `16 Go 1600 MHz DDR3` memory and with a 10 minutes time-out.

## 6.3 Benchmarks

We test the reduction power and the time efficiency of our framework on three families of models offering various conditions about the ratio of the number of Kappa rules to the number of reactions and about the ratio of the number of different bio-molecular species configurations to the number of their equivalence classes. In KaDe, the computation time for generating networks (or ODEs) depends mainly on the number of rules and the number of equivalence classes of bio-molecular species configurations. The data-structure described in [?] is used to generate reactions efficiently. More examples, including most of the BNGL test suite, are provided in Supplementary Information [?].

The first family involves a kinase, a phosphatase, and a target protein. The target protein has $n$ sites ($n$ is left as a parameter). The kinase may bind and unbind to each non-phosphorylated site of the target protein. The kinase may phosphorylate a site when releasing it. Conversely, the phosphatase may bind and unbind to each phosphorylated site of the target protein. The phosphatase may also dephosphorylate a site when releasing it. We assume that every site has the same mechanistic properties and that the rate of reactions does not depend on the state of the other sites in the target protein.

The second and third families of models are inspired by the protein Kai. This protein plays a crucial role in the control of the circadian clock oscillations. We consider a protein with $n$ sites ($n$ is left as a parameter) which may each be phosphorylated, or not. The kinase and the phosphatase are not described explicitly. We assume that the rate constants of phosphoralylation (resp. dephosphorylation) of a site in a protein depend on the number of sites that are already phosphorylated in this protein. In the third family of models, a trick suggested by Pierre Boutillier is used to reduce drastically the number of rules that are required to describe the models. We use a fictitious site that is bound to a chain of fictitious proteins the length of which encodes the number of phosphorylated sites. When a site is phosphorylated, a new protein is inserted in the chain and removed when a site gets dephosphorylated. Thus the phosphorylation level of a protein can be checked by looking at the length of this chain, without having to enumerate the different combinations fot the sites that are phosphorylated.

In Fig. 6.21, we give the number of rules, species and reactions, for each family of models for the parameter $n$ ranging from 1 to 10, as well as the number of reactions and species when equivalent sites are considered. In

Fig. 6.22, we compare the computation time to generate the original and the reduced networks with BNGL and KaDe. The generation of reduced models with KaDe (which does not require explicit annotation of equivalent sites) is much faster than the one of the unreduced networks. KaDe and BNGL generate exactly the same reduced networks. Lastly, we apply the fast version of ERODE of the bisimulation inference algorithm [**?**] on the original networks and the complete version on the reduced ones [**?**]. But we found not further reduction this way. In [**?**], we observe as good results on the BNGL test suite.

# Chapter 7

# Conclusion

Après un bref passage en revue de l'écosystème Kappa et de l'utilisation de l'interprétation abstraite pour extraire les propriétés des réseaux d'interactions biomoléculaires, le langage Kappa a été présenté plus en détail, ainsi qu'une analyse statique pour détecter parmi un ensemble de motifs d'intérêt lesquels peuvent potentiellement apparaître dans des complexes biochimiques dans une trace d'exécution d'un modèle.

Du point de vue de l'utilisateur, cette analyse permet de trouver – ou de retrouver – des propriétés structurelles sur les différentes configurations des occurrences des protéines au sein des complexes biochimiques : elle détecte quelles sont les relations entre l'état des sites des occurrences d'une protéine (Est-ce que tel site peut être lié sans que tel autre le soit ? Est-ce que ce site peut être lié sans être phosphorylé ?) ; elle permet de vérifier si deux occurrences de protéines liées entre-elles sont, oui ou non, nécessairement localisées au même endroit au sein d'une hiérarchie statique de compartiments ; elle analyse si une occurrence de protéines peut être doublement liée à une autre ou si elle peut être liée à deux occurrences différentes de protéines. En plus, de permettre la détection de règles mortes, qui ne pourront jamais être appliquées dans le modèle, le résultat est présenté graphiquement sous la forme de lemmes de raffinement, ce qui le rend compréhensible et facilement utilisable pour des analyses ultérieures. Il est ensuite possible de se concentrer sur le comportement des occurrences d'une protéine en particulier et d'obtenir un système de transitions pour décrire leurs changements potentiels de configuration.

Cette analyse passe à l'échelle de grands modèles. Cependant, pour ceux-ci, le temps de calcul reste trop important pour permettre une analyse interactive et sans latence pendant l'écriture même des modèles. Une formulation du calcul du plus petit point fixe abstrait sous forme de résolution de clauses de Horn pourrait donner lieu à une analyse incrémentale. Celle-ci permettrait de mettre à jour très rapidement le résultat de l'analyse lorsque des règles sont retirées ou ajoutées à un modèle. Par ailleurs, une collaboration étroite avec les modélisateurs est toujours nécessaire pour identifier des nouvelles familles de propriétés d'intérêt. Un autre axe de recherche est l'intégration de l'analyse statique dans des cycles de modélisations automatiques. En effet, les méthodes de fouille de la littérature basées sur l'intelligence artificielle et le traitement automatique des langages naturels pourront bénéficier de l'analyse statique d'une part pour évaluer le bien fondé d'une étape de raffinement de modèle et d'autre part pour orienter les méthodes automatiques dans leur recherche de nouvelles règles.

En ce qui concerne la modélisation en Kappa, il est important de considérer non pas un réseau d'interactions biomoléculaires dans son individualité, mais une famille de réseaux d'interactions pouvant représenter un système dans différents contextes cellulaires et ses évolutions potentielles. Les travaux sur la plate-forme de modélisation Kami vont dans ce sens [79, 81]. Il est aussi important de proposer des méthodes pour assister le modélisateur dans la construction de modèles, afin d'agglomérer des informations partielles sur les interactions biomoléculaires en les raffinant progressivement. Une approche inspirée des approches déductives, qui assimile le processus de modélisation à une recherche de preuves assistée par ordinateur, est très prometteuse [85, 84]. Dans ce contexte, une analyse statique le plus tôt possible dans la chaîne de modélisation doit être développée pour aider au mieux le modélisateur dans sa tâche.

Améliorer l'interactivité des outils [16, 20] et un travail sur le rendu visuel des propriétés [69] sont des enjeux cruciaux pour créer des outils utilisables pour des modélisateurs non experts en langage formel. Il est important d'intéresser un spectre plus large d'utilisateurs. D'une part, c'est une source inépuisable de défis scientifiques qui permettent l'amélioration des outils. D'autre part, c'est nécessaire pour construire un nombre satisfaisant de modèles.

Les modèles sont de plus en plus grands, que ce soit en nombre de complexes biochimiques différents ou en nombre d'instances des complexes biochimiques. Évaluer leur comportement est primordial, mais difficile. Les méthodes exactes de réduction de modèles sont utiles, mais limitées, pour ce type de modèles. Il est important de développer des méthodes numériques approchées pour les sémantiques différentielles et stochastiques des modèles qui permettront de trouver un encadrement garanti de l'évolution du nombre d'instances ou de la concentration, selon le choix de la sémantique, de motifs d'intérêt au cours du temps, sous la forme de paires de fonctions, elles-mêmes définies comme la solution d'un système différentiel ou comme les trajectoires d'un système stochastique. Des travaux préliminaires ont permis d'intégrer dans un cadre formel des méthodes de troncation de développement formel [106] ou des méthodes inspirées de la physique comme la tropicalisation [7], tout en fournissant des bornes évoluant au cours de l'exécution des modèles sur les erreurs numériques accumulées. Il devrait également être possible de définir une version quantitative de l'analyse de flot d'information entre sites des protéines, afin de négliger les petits flots d'information, au prix d'une perte de précision dans les modèles réduits. Un cadre formel pour l'exécution numériquement approchée des modèles permettra d'interfacer les sémantiques différentielles et stochastiques de Kappa pour concevoir une sémantique hybride, plus adaptée à la description des interactions entre des complexes biochimiques géants rares et des petits complexes présents en très grand nombre.

# Bibliography


[1] Wassim Abou-Jaoudé, Jérôme Feret et Denis Thieffry : Derivation of qualitative dynamical models from biochemical networks. *In* Olivier F. Roux et Jérémie Bourdon, éditeurs : *Computational Methods in Systems Biology - 13th International Conference, CMSB 2015, Nantes, France, September 16-18, 2015, Proceedings*, volume 9308 de *Lecture Notes in Computer Science*, pages 195–207. Springer, 2015.

[2] Wassim Abou-Jaoudé, Denis Thieffry et Jérôme Feret : Formal derivation of qualitative dynamical models from biochemical networks. *Biosystems*, 149:70–112, 2016.

[3] Emilie Allart, Joachim Niehren et Cristian Versari : Computing difference abstractions of metabolic networks under kinetic constraints. *In* Luca Bortolussi et Guido Sanguinetti, éditeurs : *Computational Methods in Systems Biology - 17th International Conference, CMSB 2019, Trieste, Italy, September 18-20, 2019, Proceedings*, volume 11773 de *Lecture Notes in Computer Science*, pages 266–285. Springer, 2019.

[4] Jakob L. Andersen, Christoph Flamm, Daniel Merkle et Peter F. Stadler : A software package for chemically inspired graph transformation. *In* Rachid Echahed et Mark Minas, éditeurs : *Graph Transformation - 9th International Conference, ICGT 2016, in Memory of Hartmut Ehrig, Held as Part of STAF 2016, Vienna, Austria, July 5-6, 2016, Proceedings*, volume 9761 de *Lecture Notes in Computer Science*, pages 73–88. Springer, 2016.

[5] Oana Andrei et Hélène Kirchner : A rewriting calculus for multigraphs with ports. *Electr. Notes Theor. Comput. Sci.*, 219:67–82, 2008.

[6] Nicolas Behr et Jean Krivine : Compositionality of rewriting rules with conditions. *CoRR*, abs/1904.09322, 2019.

[7] Andreea Beica, Jérôme Feret et Tatjana Petrov : Tropical abstraction of biochemical reaction networks with guarantees. *Electr. Notes Theor. Comput. Sci.*, 350:3–32, 2020.

[8] Bruno Blanchet, Patrick Cousot, Radhia Cousot, Jérôme Feret, Laurent Mauborgne, Antoine Miné, David Monniaux et Xavier Rival : A static analyzer for large safety-critical software. *In* Ron Cytron et Rajiv Gupta, éditeurs : *Proceedings of the ACM SIGPLAN 2003 Conference on Programming Language Design and Implementation 2003, San Diego, California, USA, June 9-11, 2003*, pages 196–207, 2003.

[9] M. L. Blinov, J. R. Faeder, B. Goldstein et W. S. Hlavacek : Bionetgen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*, 20(17):3289–3291, 2004.

[10] Michael L. Blinov, James R. Faeder, Byron Goldstein et William S. Hlavacek : Bionetgen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*, 20(17), 2004.

[11] Michael L. Blinov, James R. Faeder, Byron Goldstein et William S. Hlavacek : A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *BioSystems*, 83:136–151, 2006.

[12] Michael L. Blinov, James R. Faeder, Byron Goldstein et William S. Hlavacek : A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *Bio Systems*, 83, 2006.

[13] Chiara Bodei, Linda Brodo, Roberta Gori, Diana Hermith et Francesca Levi : A global occurrence counting analysis for brane calculi. *In* Moreno Falaschi, éditeur : *Logic-Based Program Synthesis and Transformation - 25th International Symposium, LOPSTR 2015, Siena, Italy, July 13-15, 2015. Revised Selected Papers*, volume 9527 de *Lecture Notes in Computer Science*, pages 179–200. Springer, 2015.

[14] Chiara Bodei, Pierpaolo Degano, Flemming Nielson et Hanne Riis Nielson : Control flow analysis for the pi-calculus. *In* Davide Sangiorgi et Robert de Simone, éditeurs : *CONCUR '98: Concurrency Theory, 9th International Conference, Nice, France, September 8-11, 1998, Proceedings*, volume 1466 de *Lecture Notes in Computer Science*, pages 84–98. Springer, 1998.

[15] Nikolay M. Borisov, Nick I. Markevich, Boris N. Kholodenko et Ernst Dieter Gilles : Signaling through receptors and scaffolds: Independent interactions reduce combinatorial complexity. *Biophysical Journal*, 89, 2005.

[16] Pierre Boutillier : The kappa simulator made interactive. *In* Luca Bortolussi et Guido Sanguinetti, éditeurs : *Computational Methods in Systems Biology - 17th International Conference, CMSB 2019, Trieste, Italy, September 18-20, 2019, Proceedings*, volume 11773 de *Lecture Notes in Computer Science*, pages 296–301. Springer, 2019.

[17] Pierre Boutillier, Ferdinanda Camporesi, Jean Coquet, Jérôme Feret, Kim Quyên Lý, Nathalie Théret et Pierre Vignet : Kasa: A static analyzer for kappa. *In* Milan Ceska et David Safránek, éditeurs : *Computational Methods in Systems Biology - 16th International Conference, CMSB 2018, Brno, Czech Republic, September 12-14, 2018, Proceedings*, volume 11095 de *Lecture Notes in Computer Science*, pages 285–291. Springer, 2018.

[18] Pierre Boutillier, Thomas Ehrhard et Jean Krivine : Incremental update for graph rewriting. *In* Hongseok Yang, éditeur : *Programming Languages and Systems - 26th European Symposium on Programming, ESOP 2017, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2017, Uppsala, Sweden, April 22-29, 2017, Proceedings*, volume 10201 de *Lecture Notes in Computer Science*, pages 201–228. Springer, 2017.

[19] Pierre Boutillier, Aurélie Faure de Pebeyre et Jérôme Feret : Proving the absence of unbounded polymers in rule-based models. *In Nine International Workshop on Static Analysis and Systems Biology (SASB'18)*, volume 350 de *ENTCS*, pages 33–56. elsevier, 2020.

[20] Pierre Boutillier, Mutaamba Maasha, Xing Li, Héctor F. Medina-Abarca, Jean Krivine, Jérôme Feret, Ioana Cristescu, Angus G. Forbes et Walter Fontana : The kappa platform for rule-based modeling. *Bioinformatics*, 34(13):i583–i592, 2018.

[21] Frances A. Brightman et David A. Fell : Differential feedback regulation of the mapk cascade underlies the quantitative differences in egf and ngf signalling in pc12 cells. *FEBS Letters*, 482(3):169–174, 2000.

[22] P. Buchholz : Exact and ordinary lumpability in finite markov chains. *Journal of Applied Probability*, 31(1):59–75, 1994.

[23] P. Buchholz : Bisimulation relations for weighted automata. *TCS*, 393(1-3):109–123, 2008.

[24] Ferdinanda Camporesi, Jérôme Feret, Heinz Koeppl et Tatjana Petrov : Combining model reductions. *In MFPSXXVI: Postproceedings of the 26th Conference on the Mathematical Foundations of Programming Semantics*, volume 265 de *Electronic Notes in Theoretical Computer Science*, pages 73–96. Elsevier Science Publishers, 2010.

[25] Ferdinanda Camporesi, Jérôme Feret et Kim Quyên Lý : Kade: A tool to compile kappa rules into (reduced) ODE models. *In* Jérôme Feret et Heinz Koeppl, éditeurs : *Computational Methods in Systems Biology - 15th International Conference, CMSB 2017, Darmstadt, Germany, September 27-29, 2017, Proceedings*, volume 10545 de *Lecture Notes in Computer Science*, pages 291–299. Springer, 2017.

[26] Luca Cardelli : Brane calculi. *In* Vincent Danos et Vincent Schächter, éditeurs : *Computational Methods in Systems Biology, International Conference, CMSB 2004, Paris, France, May 26-28, 2004, Revised Selected Papers*, volume 3082 de *Lecture Notes in Computer Science*, pages 257–278. Springer, 2004.

[27] Luca Cardelli et Andrew D. Gordon : Mobile ambients. *In* Maurice Nivat, éditeur : *Foundations of Software Science and Computation Structure, First International Conference, FoSSaCS'98, Held as Part of the European Joint Conferences on the Theory and Practice of Software, ETAPS'98, Lisbon, Portugal, March 28 - April 4, 1998, Proceedings*, volume 1378 de *Lecture Notes in Computer Science*, pages 140–155. Springer, 1998.

[28] Luca Cardelli et Andrew D. Gordon : Mobile ambients. *Theor. Comput. Sci.*, 240(1):177–213, 2000.

[29] Luca Cardelli, Mirco Tribastone, Max Tschaikowski et Andrea Vandin : Forward and backward bisimulations for chemical reaction networks. *In 26th International Conference on Concurrency Theory, CONCUR 2015, Madrid, Spain, September 1.4, 2015*, pages 226–239, 2015.

[30] Luca Cardelli, Mirco Tribastone, Max Tschaikowski et Andrea Vandin : Symbolic computation of differential equivalences. *Theor. Comput. Sci.*, 777:132–154, 2019.

[31] Federica Ciocchetta et Jane Hillston : Bio-PEPA: A framework for the modelling and analysis of biological systems. *Theoretical Computer Science*, 410(33 – 34):3065 – 3084, 2009. Concurrent Systems Biology: To Nadia Busi (1968–2007).

[32] Paul R Cohen : DARPA's big mechanism program. *Physical Biology*, 12(4):045008, jul 2015.

[33] Holger Conzelmann, Dirk Fey et Ernst D. Gilles : Exact model reduction of combinatorial reaction networks. *BMC Systems Biology*, 2:78, 2008.

[34] Holger Conzelmann, Julio Saez-Rodriguez, Thomas Sauter, Boris N. Kholodenko et Ernst D. Gilles : A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. *BMC Bioinformatics*, 7, 2006.

[35] Byron Cook, Jasmin Fisher, Elzbieta Krepska et Nir Piterman : Proving stabilization of biological systems. *In* Ranjit Jhala et David A. Schmidt, éditeurs : *Verification, Model Checking, and Abstract Interpretation - 12th International Conference, VMCAI 2011, Austin, TX, USA, January 23-25, 2011. Proceedings*, volume 6538, pages 134–149. Springer, 2011.

[36] Andrea Corradini, Tobias Heindel, Frank Hermann et Barbara König : Sesqui-pushout rewriting. *In* Andrea Corradini, Hartmut Ehrig, Ugo Montanari, Leila Ribeiro et Grzegorz Rozenberg, éditeurs : *Graph Transformations, Third International Conference, ICGT 2006, Natal, Rio Grande do Norte, Brazil, September 17-23, 2006, Proceedings*, volume 4178 de *Lecture Notes in Computer Science*, pages 30–45. Springer, 2006.

[37] Andrea Corradini, Ugo Montanari, Francesca Rossi, Hartmut Ehrig, Reiko Heckel et Michael Löwe : Algebraic approaches to graph transformation - part I: basic concepts and double pushout approach. *In* Grzegorz Rozenberg, éditeur : *Handbook of Graph Grammars and Computing by Graph Transformations, Volume 1: Foundations*, pages 163–246. World Scientific, 1997.

[38] Patrick Cousot : The calculational design of a generic abstract interpreter. *In* M. Broy et R. Steinbrüggen, éditeurs : *Calculational System Design*, pages 1–88. NATO ASI Series F. IOS Press, Amsterdam, 1999.

[39] Patrick Cousot : Constructive design of a hierarchy of semantics of a transition system by abstract interpretation. *Theoretical Computer Science*, 277(1–2):47–103, 2002.

[40] Patrick Cousot et Radhia Cousot : Abstract interpretation: A unified lattice model for static analysis of programs by construction or approximation of fixpoints. *In* Robert M. Graham, Michael A. Harrison et Ravi Sethi, éditeurs : *Conference Record of the Fourth ACM Symposium on Principles of Programming Languages, Los Angeles, California, USA, January 1977*, pages 238–252. ACM, 1977.

[41] Patrick Cousot et Radhia Cousot : Systematic design of program analysis frameworks. *In* Alfred V. Aho, Stephen N. Zilles et Barry K. Rosen, éditeurs : *Conference Record of the Sixth Annual ACM Symposium on Principles of Programming Languages, San Antonio, Texas, USA, January 1979*, pages 269–282. ACM Press, 1979.

[42] Troels Christoffer Damgaard, Espen Højsgaard et Jean Krivine : Formal cellular machinery. *Electr. Notes Theor. Comput. Sci.*, 284:55–74, 2012.

[43] Werner Damm et David Harel : LSCs: Breathing life into message sequence charts. *Formal Methods in System Design*, 19(1):45–80, 2001.

[44] V. Danos, R. Honorato-Zimmer, S. Jaramillo-Riveri et Sandro Stucki : Rigid geometric constraints for Kappa models. *In SASB'12: PostProceedings of the 3rd International Workshop on Static Analysis and Systems Biology*, volume 313 de *ENTCS*, pages 23–46. Elsevier, 2015.

[45] Vincent Danos, Jérôme Feret, Walter Fontana, Russell Harmer, Jonathan Hayman, Jean Krivine, Christopher D. Thompson-Walsh et Glynn Winskel : Graphs, rewriting and pathway reconstruction for rule-based models. *In* Deepak D'Souza, Telikepalli Kavitha et Jaikumar Radhakrishnan, éditeurs : *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2012, December 15-17, 2012, Hyderabad, India*, volume 18 de *LIPIcs*, pages 276–288. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2012.

[46] Vincent Danos, Jérôme Feret, Walter Fontana, Russell Harmer et Jean Krivine : Rule-based modelling of cellular signalling. *In* Luís Caires et Vasco Thudichum Vasconcelos, éditeurs : *CONCUR 2007 - Concurrency Theory, 18th International Conference, CONCUR 2007, Lisbon, Portugal, September 3-8, 2007, Proceedings*, volume 4703 de *Lecture Notes in Computer Science*, pages 17–41. Springer, 2007.

[47] Vincent Danos, Jérôme Feret, Walter Fontana, Russell Harmer et Jean Krivine : Abstracting the differential semantics of rule-based models: Exact and automated model reduction. *In Proceedings of the 25th Annual IEEE Symposium on Logic in Computer Science, LICS 2010, 11-14 July 2010, Edinburgh, United Kingdom*, pages 362–381. IEEE Computer Society, 2010.

[48] Vincent Danos, Jérôme Feret, Walter Fontana et Jean Krivine : Scalable simulation of cellular signaling networks. *In* Zhong Shao, éditeur : *Programming Languages and Systems, 5th Asian Symposium, APLAS 2007, Singapore, November 29-December 1, 2007, Proceedings*, volume 4807 de *Lecture Notes in Computer Science*, pages 139–157. Springer, 2007.

[49] Vincent Danos, Jérôme Feret, Walter Fontana et Jean Krivine : Scalable simulation of cellular signaling networks, invited paper. *In APLAS'07: Proceedings of the Fifth Asian Symposium on Programming Systems*, volume 4807 de *Lecture Notes in Computer Science*, pages 139–157. Springer, Berlin, Germany, 2007.

[50] Vincent Danos, Jérôme Feret, Walter Fontana et Jean Krivine : Abstract interpretation of cellular signalling networks. *In* Francesco Logozzo, Doron A. Peled et Lenore D. Zuck, éditeurs : *Verification, Model Checking, and Abstract Interpretation, 9th International Conference, VMCAI 2008, San Francisco, USA, January 7-9, 2008, Proceedings*, volume 4905 de *Lecture Notes in Computer Science*, pages 83–97. Springer, 2008.

[51] Vincent Danos et Cosimo Laneve : Graphs for core molecular biology. *In* Corrado Priami, éditeur : *Computational Methods in Systems Biology, First International Workshop, CMSB 2003, Roverto, Italy, February 24-26, 2003, Proceedings*, volume 2602 de *Lecture Notes in Computer Science*, pages 34–46. Springer, 2003.

[52] Vincent Danos et Cosimo Laneve : Formal molecular biology. *Theoretical Computer Science*, 325(1):69 – 110, 2004. Computational Systems Biology.

[53] Vincent Danos et Sylvain Pradalier : Projective brane calculus. *In* Vincent Danos et Vincent Schächter, éditeurs : *Computational Methods in Systems Biology, International Conference, CMSB 2004, Paris, France, May 26-28, 2004, Revised Selected Papers*, volume 3082 de *Lecture Notes in Computer Science*, pages 134–148. Springer, 2004.

[54] T. Ded, David Safránek, Matej Troják, Matej Klement, Jakub Salagovic et Lubos Brim : Formal biochemical space with semantics in kappa and BNGL. *Electr. Notes Theor. Comput. Sci.*, 326:27–49, 2016.

[55] Lorenzo Dematté, Corrado Priami et Alessandro Romanel : The blenx language: A tutorial. *In* Marco Bernardo, Pierpaolo Degano et Gianluigi Zavattaro, éditeurs : *Formal Methods for Computational Systems Biology: 8th International School on Formal Methods for the Design of Computer, Communication, and Software Systems, SFM 2008 Bertinoro, Italy, June 2-7, 2008 Advanced Lectures*, pages 313–365, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[56] James R. Faeder, Michael L. Blinov, Byron Goldstein et William S. Hlavacek : Rule-based modeling of biochemical networks. *Complexity*, 10(4):22–41, 2005.

[57] Manuel Fähndrich et Francesco Logozzo : Static contract checking with abstract interpretation. *In* Bernhard Beckert et Claude Marché, éditeurs : *Formal Verification of Object-Oriented Software - International Conference, FoVeOOS 2010, Paris, France, June 28-30, 2010, Revised Selected Papers*, volume 6528 de *LNCS*, pages 10–30. Springer, 2010.

[58] Martin Feinberg : Lectures on chemical reaction networks, 1979. Notes of lectures given at the Mathematics Research Centre, University of Wisconsin, in 1979.

[59] J. Feret et K. Q. Ly : Local traces: An over-approximation of the behavior of the proteins in rule-based models. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 15(4):1124–1137, July-Aug. 2018.

[60] Jérôme Feret : Confidentiality analysis of mobile systems. *In* Jens Palsberg, éditeur : *Static Analysis, 7th International Symposium, SAS 2000, Santa Barbara, CA, USA, June 29 - July 1, 2000, Proceedings*, volume 1824 de *Lecture Notes in Computer Science*, pages 135–154. Springer, 2000.

[61] Jérôme Feret : Occurrence counting analysis for the pi-calculus. *Electr. Notes Theor. Comput. Sci.*, 39(2):1–18, 2001.

[62] Jerome Feret : An algebraic approach for inferring and using symmetries in rule-based models. *In* Loïc Paulevé et Heinz Koeppl, éditeurs : *5th International Workshop on Static Analysis and Systems Biology (SASB 2014)*, volume 316 de *ENTCS*, pages 45–65. Elsevier, 2014.

[63] Jérôme Feret : An algebraic approach for inferring and using symmetries in rule-based models. *Electr. Notes Theor. Comput. Sci.*, 316:45–65, 2015.

[64] Jérôme Feret, Vincent Danos, Jean Krivine, Russ Harmer et Walter Fontana : Internal coarse-graining of molecular systems. *PNAS*, 2009.

[65] Jérôme Feret, Heinz Koeppl et Tatjana Petrov : Stochastic fragments: A framework for the exact reduction of the stochastic semantics of rule-based models. *Int. J. Software and Informatics*, 7(4):527–604, 2013.

[66] Jérôme Feret et Kim Quyên Lý : Local traces: An over-approximation of the behaviour of the proteins in rule-based models. *In* Ezio Bartocci, Pietro Liò et Nicola Paoletti, éditeurs : *Computational Methods in Systems Biology - 14th International Conference, CMSB 2016, Cambridge, UK, September 21-23, 2016, Proceedings*, volume 9859 de *Lecture Notes in Computer Science*, pages 116–131. Springer, 2016.

[67] Jérôme Feret et Kim Quyên Lý : Reachability analysis via orthogonal sets of patterns. *Electr. Notes Theor. Comput. Sci.*, 335:27–48, 2018.

[68] Maxime Folschette, Loïc Paulevé, Morgan Magnin et Olivier F. Roux : Under-approximation of reachability in multivalued asynchronous networks. *Electr. Notes Theor. Comput. Sci.*, 299:33–51, 2013.

[69] Angus Graeme Forbes, Andrew Burks, Kristine Lee, Xing Li, Pierre Boutillier, Jean Krivine et Walter Fontana : Dynamic influence networks for rule-based models. *IEEE Trans. Vis. Comput. Graph.*, 24(1):184–194, 2018.

[70] Qian Gao, Fei Liu, David Gilbert, Monika Heiner et David Tree : A multiscale approach to modelling planar cell polarity in drosophila wing using hierarchically coloured petri nets. *In Proceedings of the 9th International Conference on Computational Methods in Systems Biology*, CMSB '11, pages 209–218, New York, NY, USA, 2011. ACM.

[71] Steven Gay, François Fages, Thierry Martinez, Sylvain Soliman et Christine Solnon : On the subgraph epimorphism problem. *Discrete Applied Mathematics*, 162:214–228, 2014.

[72] Colin S Gillespie : Moment-closure approximations for mass-action models. *IET systems biology*, 3(1):52–58, 2009.

[73] Daniel T. Gillespie : Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[74] Roberta Gori et Francesca Levi : An analysis for proving temporal properties of biological systems. *In* Naoki Kobayashi, éditeur : *Programming Languages and Systems, 4th Asian Symposium, APLAS 2006, Sydney, Australia, November 8-10, 2006, Proceedings*, volume 4279 de *Lecture Notes in Computer Science*, pages 234–252. Springer, 2006.

[75] Radu Grosu, Grégory Batt, Flavio H. Fenton, James Glimm, Colas Le Guernic, Scott A. Smolka et Ezio Bartocci : From cardiac cells to genetic regulatory networks. *In* Ganesh Gopalakrishnan et Shaz Qadeer, éditeurs : *Computer Aided Verification - 23rd International Conference, CAV 2011, Snowbird, UT, USA, July 14-20, 2011. Proceedings*, volume 6806 de *Lecture Notes in Computer Science*, pages 396–411. Springer, 2011.

[76] Benjamin Gyori, John Bachman, Kartik Subramanian, Jeremy Muhlich, Lucian Galescu et Peter Sorger : From word models to executable models of signaling networks using automated assembly. *Molecular Systems Biology*, 13, 2017.

[77] Joseph Y. Halpern et Judea Pearl : Causes and explanations: A structural-model approach — part 1: Causes. *CoRR*, abs/1301.2275, 2013.

[78] Russ Harmer : Rule-based modelling and tunable resolution. *In DCM'09: Proceedings Fifth Workshop on Developments in Computational Models–Computational Models From Nature*, volume 9 de *EPTCS*, pages 65–72, 2009.

[79] Russ Harmer, Yves-Stan Le Cornec, Sébastien Légaré et Eugenia Oshurko : Bio-curation for cellular signalling: The KAMI project. *IEEE/ACM Trans. Comput. Biology Bioinform.*, 16(5):1562–1573, 2019.

[80] Russ Harmer, Vincent Danos, Jérôme Feret, Jean Krivine et Walter Fontana : Intrinsic information carriers in combinatorial dynamical systems. *Chaos*, 20, September 2010.

[81] Russ Harmer et Eugenia Oshurko : Kamistudio: An environment for biocuration of cellular signalling knowledge. *In* Luca Bortolussi et Guido Sanguinetti, éditeurs : *Computational Methods in Systems Biology - 17th International Conference, CMSB 2019, Trieste, Italy, September 18-20, 2019, Proceedings*, volume 11773 de *Lecture Notes in Computer Science*, pages 322–328. Springer, 2019.

[82] Monika Heiner et Ina Koch : Petri net based model validation in systems biology. *In* Jordi Cortadella et Wolfgang Reisig, éditeurs : *Applications and Theory of Petri Nets 2004: 25th International Conference, ICATPN 2004, Bologna, Italy, June 21–25, 2004. Proceedings*, pages 216–237. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[83] Tobias Helms, Tom Warnke, Carsten Maus et Adelinde M. Uhrmacher : Semantics and efficient simulation algorithms of an expressive multilevel modeling language. *ACM Trans. Model. Comput. Simul.*, 27(2):8:1–8:25, 2017.

[84] Adrien Husson : *Logical foundations of a modelling assistant for molecular biology.* Thèse de doctorat, Université de Paris, France, 2019.

[85] Adrien Husson et Jean Krivine : A tractable logic for molecular biology. *In* Bart Bogaerts, Esra Erdem, Paul Fodor, Andrea Formisano, Giovambattista Ianni, Daniela Inclezan, Germán Vidal, Alicia Villanueva, Marina De Vos et Fangkai Yang, éditeurs : *Proceedings 35th International Conference on Logic Programming (Technical Communications), ICLP 2019 Technical Communications, Las Cruces, NM, USA, September 20-25, 2019.*, volume 306 de *EPTCS*, pages 101–113, 2019.

[86] Mathias John, Cédric Lhoussaine, Joachim Niehren et Cristian Versari : Biochemical reaction rules with constraints. *In Programming Languages and Systems - 20th European Symposium on Programming, ESOP 2011, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2011, Saarbrücken, Germany, March 26-April 3, 2011. Proceedings*, volume 6602 de *Lecture Notes in Computer Science*, pages 338–357. Springer, 2011.

[87] Mathias John, Cedric Lhoussaine, Joachim Niehren et Cristian Versari : Biochemical reaction rules with constraints. *In ESOP'11: Proceedings of the 20th European Symposium on Programming*, volume 6602 de *Lecture Notes in Computer Science*, pages 338–357, 2011.

[88] Mathias John, Mirabelle Nebut et Joachim Niehren : Knockout prediction for reaction networks with partial kinetic information. *In* Roberto Giacobazzi, Josh Berdine et Isabella Mastroeni, éditeurs : *Verification, Model Checking, and Abstract Interpretation, 14th International Conference, VMCAI 2013, Rome, Italy, January 20-22, 2013. Proceedings*, volume 7737 de *Lecture Notes in Computer Science*, pages 355–374. Springer, 2013.

[89] Ozan Kahramanogullari et Luca Cardelli : An intuitive modelling interface for systems biology. *Int. J. Software and Informatics*, 7(4):655–674, 2013.

[90] Hannes Klarner, Alexander Bockmayr et Heike Siebert : Computing maximal and minimal trap spaces of boolean networks. *Natural Computing*, 14(4):535–544, 2015.

[91] Agnes Köhler, Jean Krivine et Jakob Vidmar : A rule-based model of base excision repair. *In* Pedro Mendes, Joseph O. Dada et Kieran Smallbone, éditeurs : *Computational Methods in Systems Biology - 12th International Conference, CMSB 2014, Manchester, UK, November 17-19, 2014, Proceedings*, volume 8859 de *Lecture Notes in Computer Science*, pages 173–195. Springer, 2014.

[92] Juraj Kolcák, David Safránek, Stefan Haar et Loïc Paulevé : Parameter space abstraction and unfolding semantics of discrete regulatory networks. *Theor. Comput. Sci.*, 765:120–144, 2019.

[93] Marta Z. Kwiatkowska, Gethin Norman et David Parker : PRISM 4.0: Verification of probabilistic real-time systems. *In* Ganesh Gopalakrishnan et Shaz Qadeer, éditeurs : *Computer Aided Verification - 23rd International Conference, CAV 2011, Snowbird, UT, USA, July 14-20, 2011. Proceedings*, volume 6806 de *Lecture Notes in Computer Science*, pages 585–591. Springer, 2011.

[94] Jonathan Laurent, Jean Yang et Walter Fontana : Counterfactual resimulation for causal analysis of rule-based models. *In* Jérôme Lang, éditeur : *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, pages 1882–1890. ijcai.org, 2018.

[95] Michael Löwe : Algebraic approach to single-pushout graph transformation. *Theor. Comput. Sci.*, 109(1&2):181–224, 1993.

[96] Antoni W. Mazurkiewicz : Traces, histories, graphs: Instances of a process monoid. *In* Michal Chytil et Václav Koubek, éditeurs : *Mathematical Foundations of Computer Science 1984, Praha, Czechoslovakia, September 3-7, 1984, Proceedings*, volume 176 de *Lecture Notes in Computer Science*, pages 115–133. Springer, 1984.

[97] D. A. McQuarrie : Stochastic approach to chemical kinetics. *Journal of Applied Probability*, 4(3):pp. 413–478, 1967.

[98] Hanne Riis Nielson et Flemming Nielson : Shape analysis for mobile ambients. *In* Mark N. Wegman et Thomas W. Reps, éditeurs : *POPL 2000, Proceedings of the 27th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, Boston, Massachusetts, USA, January 19-21, 2000*, pages 142–154. ACM, 2000.

[99] Loïc Paulevé, Morgan Magnin et Olivier F. Roux : Abstract interpretation of dynamics of biological regulatory networks. *Electr. Notes Theor. Comput. Sci.*, 272:43–56, 2011.

[100] Tatjana Petrov, Jérôme Feret et Heinz Koeppl : Reconstructing species-based dynamics from reduced stochastic rule-based models. *In* Oliver Rose et Adelinde M. Uhrmacher, éditeurs : *Winter Simulation Conference, WSC '12, Berlin, Germany, December 9-12, 2012*, pages 225:1–225:15. WSC, 2012.

[101] Brigitte Plateau : On the stochastic structure of parallelism and synchronization models for distributed algorithms. *SIGMETRICS Perform. Eval. Rev.*, 13(2):147–154, août 1985.

[102] Ovidiu Radulescu, Alexander N. Gorban, Andrei Zinovyev et Vincent Noel : Reduction of dynamical biochemical reactions networks in computational biology. *Frontiers in Genetics*, 3:131, 2012.

[103] Ovidiu Radulescu, Sergei Vakulenko et Dima Grigoriev : Model reduction of biochemical reactions networks by tropical analysis methods. *Mathematical Modelling of Natural Phenomena*, 10(3):124–138, 2015.

[104] Aviv Regev, E. M. Panina, William Silverman, Luca Cardelli et E. Y. Shapiro : Bioambients: An abstraction for biological compartments. *TCS*, 325(1):141–167, 2004.

[105] Aviv Regev, William Silverman et Ehud Shapiro : Representation and simulation of biochemical processes using the pi-calculus process algebra. *In* R. B. Altman, A. K. Dunker, L. Hunter et T. E. Klein, éditeurs : *Pacific Symposium on Biocomputing, Volume 6*, pages 459–470, Singapore, 2001.

[106] Ken Chanseau Saint-Germain et Jérôme Feret : Conservative numerical approximations of the differential semantics in biological rule- based models, 2016. Master thesis.

[107] Birgit Schoeberl, Claudia Eichler-Jonsson, Ernst D. Gilles et Gertraud Müller : Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors. *Nat Biotechnol*, 20(4):370–375, 2002.

[108] Donald Stewart : Spatial biomodelling, 2010. Master thesis, School of Informatics, University of Edinburgh.

[109] Ryan Suderman et Eric J. Deeds : Machines vs. ensembles: effective mapk signaling through heterogeneous sets of protein complexes. *PLoS Computational Biology*, 9, 2013.

[110] Alfred Tarski : A lattice-theoretical fixpoint theorem and its applications. *Pacific J. Math.*, 5(2), 1955.

# Index