Titre: Title:	Méthodes systémiques d'analyse des données de simulation de modèles de voies de signalisation cellulaire
Auteur: Author:	Simon Hardy
Date:	2007
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Hardy, S. (2007). Méthodes systémiques d'analyse des données de simulation de modèles de voies de signalisation cellulaire [Thèse de doctorat, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/7999/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/7999/
Directeurs de recherche: Advisors:	Pierre N. Robillard
Programme: Program:	Non spécifié

UNIVERSITÉ DE MONTRÉAL

MÉTHODES SYSTÉMIQUES D'ANALYSE DES DONNÉES DE SIMULATION DE MODÈLES DE VOIES DE SIGNALISATION CELLULAIRE

SIMON HARDY
DÉPARTEMENT DE GÉNIE INFORMATIQUE ET GÉNIE LOGICIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIÆ DOCTOR
(GÉNIE INFORMATIQUE)
OCTOBRE 2007

Your file Votre référence
ISBN: 978-0-494-37128-2
Our file Notre référence
ISBN: 978-0-494-37128-2

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée:

MÉTHODES SYSTÉMIQUES D'ANALYSE DES DONNÉES DE SIMULATION DE MODÈLES DE VOIES DE SIGNALISATION CELLULAIRE

présentée par: HARDY Simon

en vue de l'obtention du diplôme de: Philosophiæ Doctor

a été dûment acceptée par le jury d'examen constitué de:

Mme BOUCHENEB Hanifa, Doctorat, présidente

M. ROBILLARD Pierre N., Ph.D., membre et directeur de recherche

M. PERRIER Michel, Ph.D., membre

M. SWAIN Peter, Ph.D., membre

À ma muse créatrice,

qui m'a inspiré l'envie d'explorer de nouvelles avenues de pensée.

REMERCIEMENTS

Mes premiers remerciements vont à mon directeur de recherche, Pierre N. Robillard. Il y a un peu plus de six ans, il m'a engagé comme assistant de recherche et m'a donné le mandat de dresser un portrait de la recherche en bioinformatique au Québec. C'est grâce à lui que j'ai eu la piqûre pour les applications de l'informatique à la biologie moléculaire. Son support a été constant au cours de mon aventure doctorale dans un contexte exigeant d'interdisciplinarité. Sa confiance en ma réussite a aussi été un atout précieux pour me rendre à bon port. Par-dessus tout, il a su m'initier à la recherche et me communiquer son avidité pour de nouveaux savoirs.

Je remercie aussi Frédéric Bouchard et Stephen Michnick, respectivement professeurs de philosophie et de biochimie à l'Université de Montréal pour leur apport à ma formation et à mon projet de recherche. M. Michnick m'a invité à lire un article sur la biologie des systèmes qui fut le déclencheur des préoccupations qui ont motivé ma thèse. M. Bouchard m'a appris les rudiments de la philosophie de la biologie et m'a communiqué sa passion pour l'épistémologie des sciences.

Pour leur soutien financier durant mes études, je remercie le Conseil de recherches en sciences naturelles et génie du Canada et la Fondation de l'École Polytechnique.

Merci aux membres du jury d'évaluation de ma thèse. J'ai apprécié leurs commentaires, tant les bons que les suggestions. Ils ont souligné certaines pistes inexplorées très intéressantes.

Les deux dernières années de mon doctorat n'auraient pas été aussi agréables sans la présence de mon collègue Olivier. Merci pour ta camaraderie !

Mes derniers remerciements vont à mes proches et amis. C'est six mois après la naissance de mon fils Léo, jour pour jour, que j'ai déposé ma thèse. Un deuxième accouchement en si peu de temps n'aurait pas été possible sans leur aide inestimable. Ils sont venus garder Léo, nous aider avec les tâches ménagères et alimentaires, nous ont supportés et encouragés. Merci Esther, Alain, Marie-Ève et Victor. Merci

Céline, Colette et Geneviève. Votre aide a été salutaire. Un merci tout spécial à mes parents, Suzanne et Michel, ainsi que mes soeurs Isabelle et Fabienne et leurs conjoints Vincent et Laurent, qui nous ont aussi donné un coup de main à la maison et qui m'ont toujours encouragé dans mes études comme dans tous mes projets. Clin d'oeil à toi, mon Léo. Ta présence toute récente a déjà laissé une empreinte de titan. Il y a un peu de toi dans cette thèse. Finalement, comment exprimer toute ma reconnaissance à toi, Mahité, qui m'as accompagné fidèlement dans ce projet et qui as galéré pour me permettre de finaliser ma thèse. La réussite de cette thèse t'est aussi attribuable. Ton amour m'a porté jusqu'à la fin.

RÉSUMÉ

Un nouveau terme est de plus en plus fréquemment utilisé pour désigner une sous-discipline de la biologie : la biologie systémique. Ce nouveau champ d'étude s'intéresse aux interactions entre les composantes d'un système biologique et à l'émergence des fonctions et du comportement de ce système à partir de ces interactions. Après avoir cherché à décortiquer la cellule et ses composantes, les biologistes ont maintenant une liste exhaustive des composantes et de leurs caractéristiques. Toutefois, devant les systèmes très complexes qu'ils essayent de comprendre, les biologistes manquent de concepts et d'approches pour investiguer cette complexité. Pour mieux s'outiller et développer de nouvelles approches intégratives, les biologistes ont entamé un dialogue avec des ingénieurs, des informaticiens, des mathématiciens et des théoriciens des systèmes et du contrôle. La biologie systémique est donc un champ d'étude multidisciplinaire encore très jeune, mais qui laisse présager de grandes percées dans les sciences du vivant.

En ingénierie, la modélisation et la simulation sont des méthodes privilégiées pour l'investigation des systèmes complexes. Les avantages de la modélisation et de la simulation n'échappent pas aux biologistes, qui veulent eux aussi réaliser des études théoriques à l'aide de modèles computationnels. En conséquence, un important travail d'adaptation de différents langages de modélisation pour des applications biologiques a été accompli dans la dernière décennie et les outils informatiques de modélisation et de simulation des systèmes biologiques se sont multipliés. Les biologistes peuvent maintenant facilement modéliser et simuler. Mais pour ce qui est de l'analyse des données de simulation, peu de méthodes ont été spécifiquement conçues pour ce domaine d'application. Cette constatation a inspiré l'objectif de recherche de cette thèse. Cet objectif est de concevoir des méthodes d'analyse de données de simulation de modèles de voies de signalisation cellulaire qui permettent de dégager leur comportement systémique.

Cette thèse par articles présente trois méthodes d'analyse distinctes. L'approche présentée dans le premier article propose d'utiliser la structure du réseau biochimique, de la fonction des composantes et de leurs interrelations pour, dans un premier temps, caractériser l'activité moléculaire de modules biochimiques et ensuite, pour créer des filtres d'analyse. Ces filtres permettent de produire, à partir des données de simulation, une représentation fonctionnelle animée du système.

Le second article présente une méthode de visualisation des données de simulation des voies de signalisation cellulaire. À l'aide d'une technique d'analyse issue de la théorie des réseaux de Pétri (l'identification de P-invariants), la méthode de visualisation prend en compte le fonctionnement enzymatique pour générer une vue du système qui reflète le comportement systémique des voies de signalisation. Les méthodes de visualisation existantes, appropriées pour les voies métaboliques, donnent des résultats mitigés avec les voies de signalisation. Les problèmes détectés sont solutionnés par notre méthode.

Le troisième article présente une méthode d'analyse des modèles de voies de signalisation qui permet d'étudier leur dynamique de la propagation de signal. À l'aide de techniques d'analyse des réseaux de Pétri (l'identification de P-invariants et de T-invariants), la méthode permet d'identifier les éléments du modèle qui jouent un rôle important dans la transmission de signaux cellulaires. Elle permet aussi de décrire la dynamique de transduction et d'identifier la présence de motifs de régulation. Cette méthode transforme les données de variation de concentration moléculaire en données de transmission de signaux cellulaire, ce qui permet un accès rapide aux mécanismes systémiques de traitement d'information de la cellule.

Les trois méthodes de cette thèse ont été utilisée pour analyser les données de simulation d'une variante du modèle du réseau de régulation de l'enzyme CaMKII du neurone CA1 de l'hippocampe. Ce système biologique est majeur en neurobiochimie puisqu'il est expérimentalement associé à la plasticité synaptique, qui est la propriété

neuronale qui expliquerait les capacités d'apprentissage et de mémorisation du cerveau. Il s'agit d'un modèle complexe, avec plusieurs boucles de rétroaction, idéal pour une étude sur des méthodes d'analyse destinées à la biologie systémique.

Les résultats des deux derniers articles de cette thèse confirment la pertinence de l'utilisation du formalisme des réseaux de Pétri pour des applications de biologie systémique. Les méthodes de ces articles peuvent s'intégrer dans une méthodologie globale de modélisation-simulation-analyse basée sur ce formalisme.

ABSTRACT

A term is widely used to identify new approaches in biosciences: systems biology. This new field of study is about the interactions of the components of a biological system and how the functions and behavior of this system emerge from these interactions. After decades of a reductionist approach in molecular biology, which led to a quite exhaustive component list, biologists are now aiming at a system-level understanding of the cell. However, the actual concepts and tools of biology are inadequate to complete this new task of investigating biological complexity. To have new tools and to develop new integrative approaches, biologists started a dialogue with engineers, mathematicians, computer scientists and specialists of the system and control theory. Systems biology is a young interdisciplinary field of study, with promises of bringing our knowledge about life further.

In engineering and other sciences, modeling and simulation are widely used methods to study complex systems. Systems biology is no exception and computational biologists have also wanted to undertake theoretical studies with computational models. As a consequence, different modeling languages and formalisms have been adapted to biological applications in the last decade. Also, modeling and simulation tools have been multiplied. Biologists now can easily model and simulate biological systems. However, few tools for the analysis of the simulation data have been specifically designed for biological applications. This inspired the research goal of this thesis. This goal is to design computational methods for the analysis of the simulation data of models of signal transduction networks, that can reveal their systemic behavior.

This thesis contains three articles, each presenting a distinct analysis method. The approach presented in the first paper is a proposal to use the structure of a biochemical network, the function of its components and their interrelations to characterize the molecular activity of its modules and to create data analysis filters. These filters are used to create a functional animated representation of the system from the

simulation data.

The second paper presents a visualization method of the simulation data of signaling pathways. In this method, the p-invariant property of Petri nets is used to identify meaningful biological entities. It is then possible to create a view of the system taking into account the behavior of enzymes. Existing visualization methods have been developed mostly for metabolic network models and they are less suitable for signaling pathway models. The Petri net-based method solves this problem and highlights the switching behavior of enzymes, thereby allowing a visual exploration of the global dynamic behavior of complex biochemical model.

The third article presents a method for the analysis of the dynamics of the propagation of a signal in signaling pathway models. In this method, the p-invariant and t-invariant properties of Petri nets are used to identify key signaling elements of pathways. This method produces temporal information about signal propagation in the network, a simplified graphical representation of the network and of the signal propagation dynamics and a characterization of some signaling routes as regulation motifs. This method can be useful to deciphering the information processing capabilities of the pathway.

In each article, the presented method has been used to analyze the simulation data of variations of model of the CaMKII enzyme regulation network of the hippocampal CA1 neuron. This biological system is very important in neurobiochemistry because it is experimentally linked to synaptic plasticity, which is the neuronal property hypothesized to be responsible for memory and learning processes. This is a complex model, perfect for a study on computational analysis methods for systems biology.

The results of the two last papers of this thesis confirm the usefulness of the Petri net formalism for systems biology applications. The methods presented in these papers can be integrated into a unified methodological Petri net-based framework for the modeling, simulation and data analysis of complex biological systems.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

LISTE DES ANNEXES

LISTE DES NOTATIONS ET DES SYMBOLES

AC1	Adenylyl cyclase 1
ADN	Acide désoxyribonucléique
AMPAR	Amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptor
ARN	Acide ribonucléique
ATP	Adenosine triphosphate
CaM	Calmodulin
CaMKII	Ca^{2+}/calmodulin-dependant protein kinase II
cAMP	Cyclic adenosine monophosphate
CaN	Calcineurin
GON	Genomic Object Net software
HFPN	Hybrid functional Petri net
HOT	Highly optimized tolerance
HPN	Hybrid Petri net
I1	Inhibitory 1 protein
LTD	Long-term depression
LTP	Long-term potentiation
MAPK	Mitogen-Associated protein kinase
Ng	Neurogranin
NMDAR	*N*-Methyl-D-aspartate Receptor
ODE	Ordinary differential equation
PDE	Phosphodiesterase
PKA	Protein kinase A
PLA_2	Cytosolic phospholipase A-2
PLC	Phospholipase C

PP1	Protein phosphatase 1
PP2A	Protein phosphatase 2 A
PP2B	Protein phosphatase 2 B
RdeP	Réseaux de Pétri
RdePHF	Réseaux de Pétri hybrides fonctionnels

INTRODUCTION

Avec les percées scientifiques majeures de la biologie moléculaire des deux dernières décennies, comme le séquençage du génome humain, les biologistes ont dû se doter de nouveaux outils informatiques. Plusieurs méthodes computationnelles ont alors été adaptées ou créées par de nouveaux spécialistes, les bioinformaticiens, pour répondre à cette demande. Des techniques de criblage de données (*data mining*) ont été développées pour traiter le raz-de-marée de données collectées au laboratoire à l'aide des nouvelles technologies automatisées de la génomique et de la protéomique. Comparaison et alignement de séquences biologiques, recherche de motifs, assemblage de fragments d'ADN, cartographie génétique, structure 3D des protéines et ARN, les méthodes computationnelles se sont avérées essentielles pour accomplir ces tâches. Ceci a enfin donné aux biologistes une liste plus complète des composantes biologiques du vivant et de leurs caractéristiques.

Cette liste ne leur a toutefois pas livré les secrets de la vie. Entre autres, les biologistes se sont étonnés du nombre peu élevé de gènes dans le génome humain, soit approximativement 20 000 à 25 000. Comment si peu de gènes peut-il générer la complexité de l'espèce humaine ?

Dans leur quête de compréhension, les biologistes ont adopté une nouvelle approche, une approche systémique qui suppose que les processus et les fonctions biologiques émergent de la complexité des interactions entre les composantes. Peu habitués à ce nouveau type de questions de recherche et mal équipés conceptuellement pour faire face à la complexité des systèmes étudiés, certains biologistes se sont adjoint des scientifiques plus aguerris avec ce genre de problématique. Mathématiciens, ingénieurs, informaticiens et théoriciens de l'automatisation et du contrôle ont vu, dans les nouvelles interrogations des biologistes, un domaine de recherche où ils pouvaient

appliquer leurs connaissances. Cette nouvelle approche en biologie, dont on a commencé à parler vers l'année 2000, a été nommée biologie systémique (connue sous le nom de *systems biology* en anglais). La définition de cette jeune discipline est encore imprécise et il en existe plusieurs. Voici une définition satisfaisante, proposée par Sauer et al. (2007) :

> Rather than a reductionist viewpoint (that is, a deterministic genetic view), the pluralism of causes and effects in biological networks is better addressed by observing, through quantitative measures, multiple components simultaneously, and by rigorous data integration with mathematical models. Such a system-wide perspective (so-called systems biology) on component interactions is required so that network properties, such as a particular functional state or robustness, can be quantitatively understood and rationally manipulated.

La conception de modèles, où on représente différentes entités ainsi que leurs interrelations, est une approche privilégiée pour synthétiser la connaissance accumulée sur système. La réalisation de simulations est à la base des approches quantitatives utilisées pour comprendre le comportement global d'un système. La modélisation et la simulation sont un des axes de recherche en informatique et on en recense une variété d'applications dans tous les domaines d'étude. La biologie ne faisait pas figure d'exception, mais l'avènement de la biologie systémique a provoqué une multiplication des études sur les systèmes biologiques à l'aide de la modélisation et la simulation. C'est ainsi que différents formalismes de modélisation ont été étudiés pour leur adéquation à représenter des phénomènes biologiques.

Lors de mes travaux de recherche à la maîtrise, j'ai moi-même réalisé une de ces études où j'explorais l'utilisation des réseaux de Pétri pour la modélisation biolo-

gique. Décrits simplement, les réseaux de Pétri sont un outil de modélisation du comportement des systèmes dynamiques à événements discrets et de description des relations existantes entre des conditions et des évènements. Depuis qu'ils ont été inventés par Carl Adam Petri en 1962, les réseaux de Pétri ont principalement été appliqués aux systèmes informatiques et industriels. La théorie de ce formalisme a été l'objet de plusieurs recherches et plusieurs contributions l'ont enrichie de nouveaux concepts et ont augmenté son pouvoir de modélisation.

Mon mémoire de maîtrise était constitué de deux études. Dans la première étude, j'ai fait une recension des écrits sur les différentes utilisations de ce formalisme pour la modélisation biologique ainsi qu'une analyse identifiant les meilleures techniques selon les objectifs de modélisation (Hardy et Robillard, 2004). Dans la seconde étude, je démontrais que les réseaux de Pétri sont un formalisme qui facilite la création de modèles biologiques où l'on retrouve plusieurs niveaux d'abstractions (Hardy et Robillard, 2005).

Ma thèse de doctorat est le prolongement de ce premier projet de recherche. Elle porte sur un aspect de la modélisation et de la simulation que je n'avais jusqu'alors pas abordé : l'analyse des données de simulation. En effet, la simulation de modèles complexes génère une grande quantité de données dont il n'est pas toujours facile de dégager l'information recherchée. Cette problématique fut le point de départ de mes recherches doctorales et a servi à définir mon objectif de recherche. Cet objectif est de concevoir des méthodes d'analyse de données de simulation de modèles de voies de signalisation cellulaire qui permettent de dégager leur comportement systémique.

J'ai atteint cet objectif à l'aide de trois études distinctes, sous la forme d'articles. Chaque étude propose une approche originale pouvant s'insérer parfaitement dans une démarche de biologie systémique. Une première étude se veut une réflexion sur l'utilisation du concept de fonctionnalité dans l'analyse de données de simulation

d'un modèle biochimique (Hardy et Robillard, 2006a). L'approche présentée propose d'identifier les unités fonctionnelles d'un système, de spécifier mathématiquement le fonctionnement attendu de ces unités pour ensuite générer des représentations animées de leur dynamique. Cette approche est valide pour tout formalisme de modélisation qui permet la réalisation d'études quantitatives.

Les deux autres études présentées dans cette thèse décrivent des méthodes d'analyse des données de simulation basées sur le formalisme des réseaux de Pétri. Ce formalisme propose différentes techniques pour analyser les propriétés des systèmes et pour valider les modèles, dont l'analyse des invariants. On peut établir des relations entre les invariants des réseaux de Pétri et des caractéristiques particulières ayant une signification biologique. Les deux études exploitent ces relations. La seconde étude utilise les invariants dans une méthode de visualisation des données de simulation (Hardy et Robillard, 2007b). La troisième étude utilise les invariants pour décortiquer le comportement dynamique de la propagation d'un signal dans un réseau de transduction cellulaire (Hardy et Robillard, 2007a).

Le premier chapitre de cette thèse porte sur les réseaux de Pétri. S'y trouvent une courte introduction sur ce formalisme, une recension des écrits sur ses applications biologiques de 2004 à aujourd'hui et la présentation d'un ajout à la théorie que je propose, les réseaux de Pétri peinturés. Le chapitre 2 présente des concepts d'ingénierie et de la théorie des systèmes qui servent de fondement conceptuel à la biologie systémique ainsi qu'une recension non-exhaustive d'études qui utilisent ces concepts. Le chapitre 3 présente l'organisation générale du travail et la cohérence des articles par rapport aux objectifs de recherche. Les chapitres 4, 5 et 6 présentes les trois articles. Un dernier chapitre propose une discussion sur des aspects méthodologiques de la simulation avec les réseaux de Pétri et aussi deux réflexions philosophiques sur des sujets relatifs à la biologie systémique et aux idées de cette thèse.

CHAPITRE 1

LES RÉSEAUX DE PÉTRI, UN OUTIL DE MODÉLISATION POLYVALENT

Les réseaux de Pétri sont un outil de modélisation principalement utilisé en informatique et en automatique pour représenter des systèmes ayant des propriétés de parallélisme, de concurrence, de synchronisme et de partage de ressources. Les réseaux de Pétri ont une double nature : une nature graphique qui facilite la modélisation et la visualisation et une nature mathématique qui les dote d'intéressantes propriétés très utiles pour les analyser. De plus, cet outil de modélisation est suffisamment général pour permettre la modélisation de phénomènes très variés.

Ce chapitre se divise en trois sections. La première section est une introduction à la théorie des réseaux de Pétri, en ne s'attardant que sur les pans qui sont essentiels à la compréhension de cette thèse. Elle est entre autres inspirée d'un livre sur les réseaux de Pétri discrets, continus et hybrides de David et Alla (2004) que nous invitons le lecteur à consulter pour une introduction complète. La deuxième section est une mise à jour de la recension des travaux de recherche portant sur l'utilisation des réseaux de Pétri pour la modélisation et la simulation de systèmes biologiques. La précédente revue recensait les écrits jusqu'en 2004 (Hardy et Robillard, 2004). La troisième section présente un ajout que nous avons fait à la théorie de ce formalisme : les réseaux de Pétri peinturés.

1.1 Introduction aux réseaux de Pétri

Cet outil de modélisation a été inventé par Carl Adam Petri en 1962 et publié dans sa thèse de doctorat (Petri, 1962). Depuis, la théorie des réseaux de Pétri n'a cessé de s'enrichir et celle-ci est toujours l'objet de recherches théoriques et appliquées.

1.1.1 Concepts fondamentaux

1.1.1.1 Places, transitions et arcs

Un réseau de Pétri (RdeP) est un graphe bipartite où les deux types de noeuds sont les places et les transitions. Le RdeP traditionnel se nomme aussi un réseau place-transition. Les places sont représentées par des cercles et les transitions, par un trait (ou une boîte rectangulaire, selon les auteurs). Les places et les transitions sont reliées par des arcs. Ces arcs sont dirigés et ils ne peuvent connecter que deux noeuds de types différents. Dans un réseau place-transition, les arcs ont chacun un poids, qui est un nombre entier strictement positif et dont la valeur par défaut est un. De façon générale, dans un réseau de Pétri, les places représentent les éléments passifs d'un modèle alors que les transitions représentent les éléments actifs.

La figure 1.1 présente un réseau de Pétri qui modélise l'algorithme d'exclusion mutuelle de Peterson. Ce réseau contient 7 places, 6 transitions et 16 arcs. L'algorithme de Peterson gère l'accès à une ressource qui est partagée entre différents processus et qui ne peut être accédée ou utilisée que par un seul processus à la fois. Les places P_1, P_2 et P_3 représentent les différents états d'un premier type processus, les places P_4, P_5 et P_6 représentent les différents états d'un second type processus et la place P_7 représente la ressource que les deux processus se partagent. Les places P_3 et P_6

représentent les états où les processus accèdent ou utilisent la ressource.

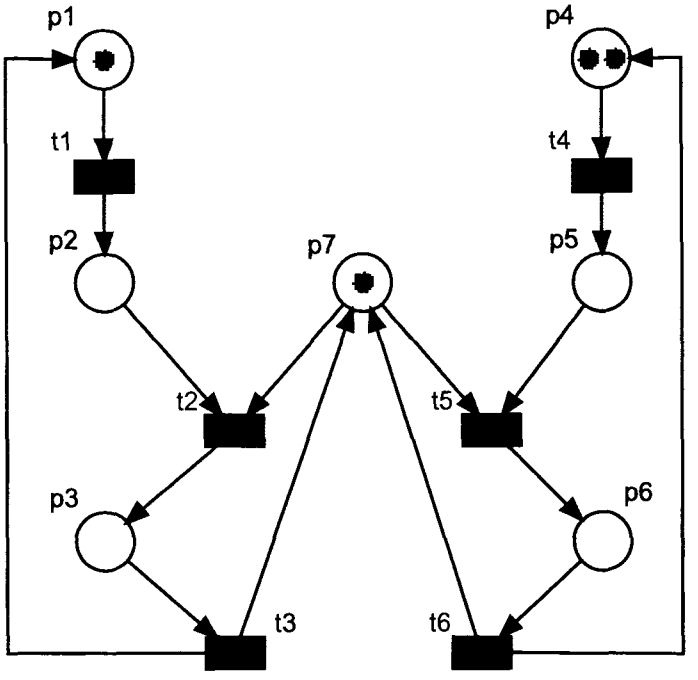


Figure 1.1: Réseau de Pétri de l'algorithme de Peterson

Les relations entre les places et les transitions peuvent être qualifiées d'entrée ou de sortie. Par exemple, dans la figure 1.1, les places P_2 et P_7 sont les places d'entrée de la transition T_2. De même, la transition T_5 est une des transitions de sortie de la place P_7.

1.1.1.2 Marquage

Pour qu'un réseau de Pétri soit pleinement fonctionnel, celui-ci doit être marqué. Un RdeP est marqué lorsque ses places contiennent un nombre positif de jetons. Le nombre de jetons de la place P_i est associé à la variable $m(P_i)$ ou m_i. Dans l'exemple de la figure 1.1, nous avons $m_1 = m_7 = 1$, $m_2 = m_3 = m_5 = m_6 = 0$ et $m_4 = 2$. La distribution de jetons dans les places est appelée le marquage d'un RdeP.

Le marquage m d'un RdeP est défini par un vecteur du marquage de ses places, tel que $\mathbf{m} = (m_1, m_2, m_3, m_4, m_5, m_6, m_7)$. Le marquage du RdeP de la figure 1.1 est $\mathbf{m} = (1, 0, 0, 2, 0, 0, 1)$. Ce marquage signifie qu'il y a un processus du premier type, deux autres processus du second type et une seule ressource à partager entre ces trois instanciations de processus. Le marquage détermine l'état dans lequel le système décrit par le RdeP se trouve. Le RdeP évolue à mesure que son marquage change. Ceci se produit lorsqu'une transition est franchie.

1.1.1.3 Franchissement d'une transition

Une transition est franchie – on dit aussi qu'elle est tirée – si chaque place d'entrée de cette transition contient un nombre de jetons égal ou supérieur au poids des arcs reliant les places d'entrée à la transition. Si cette condition de franchissement est remplie, on dit de la transition qu'elle est tirable. Dans le RdeP de la figure 1.1, les transitions T_1 et T_4 sont tirables.

Lors du franchissement d'une transition, on retire des jetons des places d'entrée de la transition en nombre équivalent au poids des arcs les reliant et on ajoute des jetons aux places de sortie de la transition, toujours selon le poids des arcs. La figure 1.2 illustre cet événement. À la figure 1.2(a), on peut voir le marquage du RdeP de la figure 1.1 après le franchissement de la transition T_1. Lors de ce franchissement, un jeton a été retiré de la place P_1 et un autre a été ajouté à la place P_2. À partir de ce marquage, les transitions T_2 et T_4 sont tirables. Toujours à partir du marquage de la figure 1.1, on peut voir le marquage du RdeP après le franchissement de la transition T_4 à la figure 1.2(b).

Dans certains cas, le franchissement d'une transition peut mener à un marquage où aucune transition n'est tirable. Un réseau de Pétri dans cet état est dit en blocage.

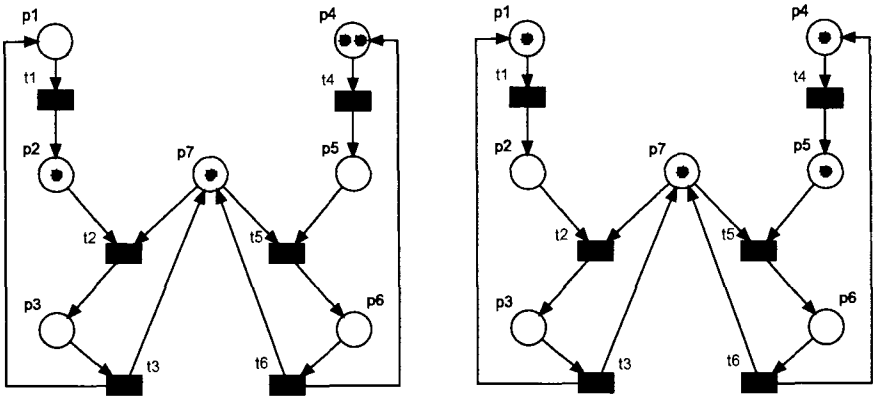


Figure 1.2: Franchissement de transition

Maintenant que la notion de jeton est introduite, remarquez que la structure du RdeP de l'exemple, à partir du marquage initial de la figure 1.1, assure qu'un seul processus à la fois est dans un état d'utilisation de la ressource. Si l'instanciation du premier processus est dans l'état représenté par la place P_3, ceci signifie qu'il utilise la ressource partagée et donc il n'y a aucun jeton à la place P_7. Ceci a pour conséquence que les conditions de franchissement de la transition T_5 ne sont pas remplies, ce qui rend cette transition infranchissable et empêche donc l'utilisation de la ressource par une des instanciations du deuxième processus tant que la ressource est utilisée par l'autre processus. En d'autres mots, il est impossible qu'il y ait simultanément un jeton à la place P_3 et un autre à la place P_6.

1.1.1.4 Réseaux de Pétri autonomes et non-autonomes

Les réseaux place-transition sont entièrement non-déterministes. Il n'y a aucune contrainte de modélisation pour fixer l'évolution du réseau. Il n'y pas de priorité pour certains changements d'état et l'ordre de franchissement des transitions n'est

pas fixé ni prévisible. On dit de ces réseaux qu'ils sont autonomes. Nous verrons que ceci n'est pas le cas pour tous les types de réseau de Pétri. Certaines extensions de la théorie permettent le franchissement de transition uniquement lors d'événements externes ou selon des contraintes temporelles. De tels RdeP sont non-autonomes. Par exemple, dans les RdeP synchrones, qui sont un type de réseaux non-autonomes, on peut spécifier une durée entre le moment où une transition devient tirable et le moment de sa transition.

1.1.2 Propriétés des réseaux de Pétri

1.1.2.1 Notations et définitions

Pour introduire certaines notations usuelles et certains concepts, ayons comme point de départ le marquage du RdeP de la figure 1.1, nommé marquage initial $\mathbf{m_0}$. Ce marquage, comme le marquage du RdeP à tout moment donné, est représenté par un vecteur colonne où la composant i est le marquage de la place P_i. Pour ne pas alourdir le texte, la forme transposée du vecteur entre parenthèses sera plus souvent utilisée. Par exemple :

$$\mathbf{m_0} = (1,0,0,2,0,0,1) = \begin{bmatrix} 1 & 0 & 0 & 2 & 0 & 0 & 1 \end{bmatrix}^T = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Du marquage initial $\mathbf{m_0}$, nous avons déjà vu que les transition T_1 et T_4 sont tirables. Si

la transition T_1 est franchie, le nouveau marquage $\mathbf{m_1}$ du réseau est $(0, 1, 0, 2, 0, 0, 1)$. Cet événement, dont le résultat est montré à la figure 1.2(a), est décrit par :

$$\mathbf{m_0} \xrightarrow{T_1} \mathbf{m_1}.$$

Dans le cas où la transition T_4 est franchie, le nouveau marquage $\mathbf{m_2}$ du réseau est $(1, 0, 0, 1, 1, 0, 1)$. Cet événement, dont le résultat est montré à la figure 1.2(b), est décrit par :

$$\mathbf{m_0} \xrightarrow{T_4} \mathbf{m_2}.$$

Du marquage $\mathbf{m_1}$, les transitions T_2 et T_4 sont tirables. Si la transition T_2 est franchie, on obtient le marquage $\mathbf{m_3}$ alors que si la transition T_4 est franchie, on obtient le marquage $\mathbf{m_4}$. Ces événements sont décrits par :

$$\mathbf{m_1} \xrightarrow{T_2} \mathbf{m_3} = (0, 0, 1, 2, 0, 0, 0),$$

$$\mathbf{m_1} \xrightarrow{T_4} \mathbf{m_4} = (0, 1, 0, 1, 1, 0, 1).$$

En poursuivant l'exploration des marquages qu'il est possible d'atteindre par le franchissement de transitions, on peut revenir à des marquages connus ou découvrir de nouveaux états. Par exemple, le franchissement de la transition T_1 à partir du marquage $\mathbf{m_2}$ produit aussi le marquage $\mathbf{m_4}$. Toujours à partir du marquage $\mathbf{m_2}$, le franchissement de la transition T_4 génère le nouveau marquage $\mathbf{m_5}$ et le franchissement de la transition T_5 génère le nouveau marquage $\mathbf{m_6}$. Ces événements sont décrits par :

$$\mathbf{m_2} \xrightarrow{T_1} \mathbf{m_4},$$

$$\mathbf{m_2} \xrightarrow{T_4} \mathbf{m_5} = (1, 0, 0, 2, 0, 0, 1),$$

$\mathbf{m_2} \xrightarrow{T_5} \mathbf{m_6} = (1,0,0,1,0,1,0)$.

L'exploration des marquages atteignables du RdeP de l'algorithme de Peterson se poursuit ainsi :

$\mathbf{m_3} \xrightarrow{T_3} \mathbf{m_0}$,

$\mathbf{m_3} \xrightarrow{T_4} \mathbf{m_7} = (0,0,1,1,1,0,0)$,

$\mathbf{m_4} \xrightarrow{T_2} \mathbf{m_7}$,

$\mathbf{m_4} \xrightarrow{T_4} \mathbf{m_8} = (0,1,0,0,2,0,1)$,

$\mathbf{m_4} \xrightarrow{T_5} \mathbf{m_9} = (0,1,0,1,0,1,0)$,

$\mathbf{m_5} \xrightarrow{T_2} \mathbf{m_2}$,

$\mathbf{m_5} \xrightarrow{T_5} \mathbf{m_6}$,

$\mathbf{m_6} \xrightarrow{T_1} \mathbf{m_9}$,

$\mathbf{m_6} \xrightarrow{T_4} \mathbf{m_{10}} = (1,0,0,0,1,1,0)$,

$\mathbf{m_6} \xrightarrow{T_5} \mathbf{m_0}$,

$\mathbf{m_7} \xrightarrow{T_3} \mathbf{m_2}$,

$\mathbf{m_7} \xrightarrow{T_4} \mathbf{m_{11}} = (0,0,1,0,2,0,0)$,

$\mathbf{m_8} \xrightarrow{T_2} \mathbf{m_{11}}$,

$\mathbf{m_8} \xrightarrow{T_5} \mathbf{m_{10}}$,

$\mathbf{m_9} \xrightarrow{T_6} \mathbf{m_1}$,

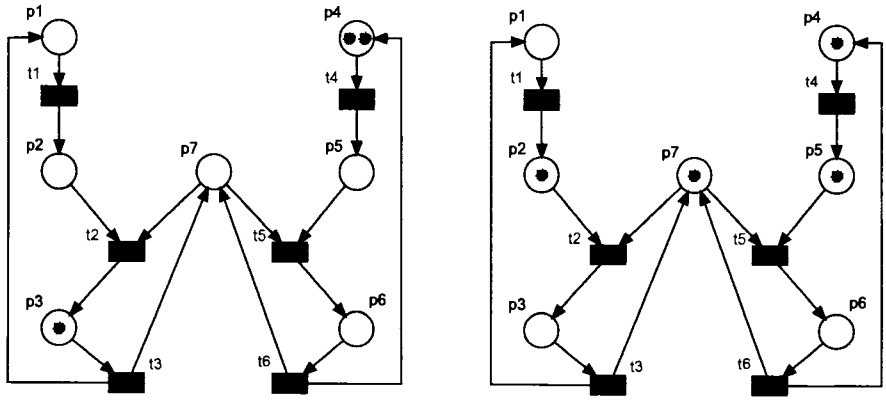
$\mathbf{m_{10}} \xrightarrow{T_1} \mathbf{m_{12}} = (0,1,0,0,1,1,0)$,

$\mathbf{m_{10}} \xrightarrow{T_5} \mathbf{m_2}$,

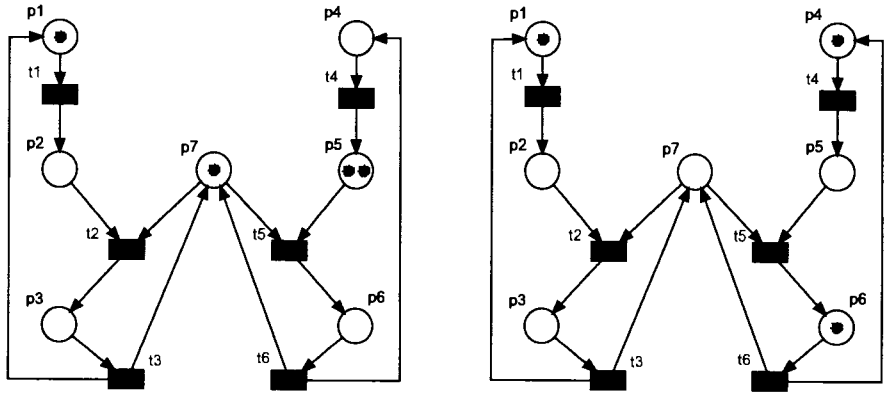
$\mathbf{m_{11}} \xrightarrow{T_3} \mathbf{m_5}$,

$\mathbf{m_{12}} \xrightarrow{T_6} \mathbf{m_4}$.

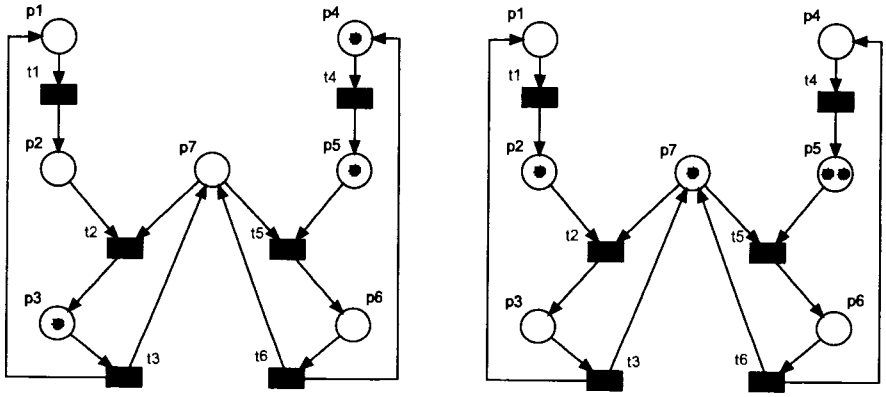
Dans le RdeP de l'algorithme de Peterson, il existe un nombre fini de marquages atteignables. Pour un RdeP, $\mathcal{M}(\mathbf{m_0})$ est l'ensemble des marquages atteignables à partir du marquage $\mathbf{m_0}$ par une séquence finie de franchissements. Dans notre exemple, $\mathcal{M}(\mathbf{m_0}) = \{\mathbf{m_0}, \mathbf{m_1}, \mathbf{m_2}, \mathbf{m_3}, \mathbf{m_4}, \mathbf{m_5}, \mathbf{m_6}, \mathbf{m_7}, \mathbf{m_8}, \mathbf{m_9}, \mathbf{m_{10}}, \mathbf{m_{11}}, \mathbf{m_{12}}\}$. Les figures 1.1, 1.2 et 1.3 illustrent tous les marquages atteignables à partir de $\mathbf{m_0}$. En examinant $\mathcal{M}(\mathbf{m_0})$, on peut constater que la propriété souhaité du modèle de l'algorithme de Peterson est en effet respecté. Dans aucun des marquages accessibles à partir d'un



(a) m_3 (b) m_4

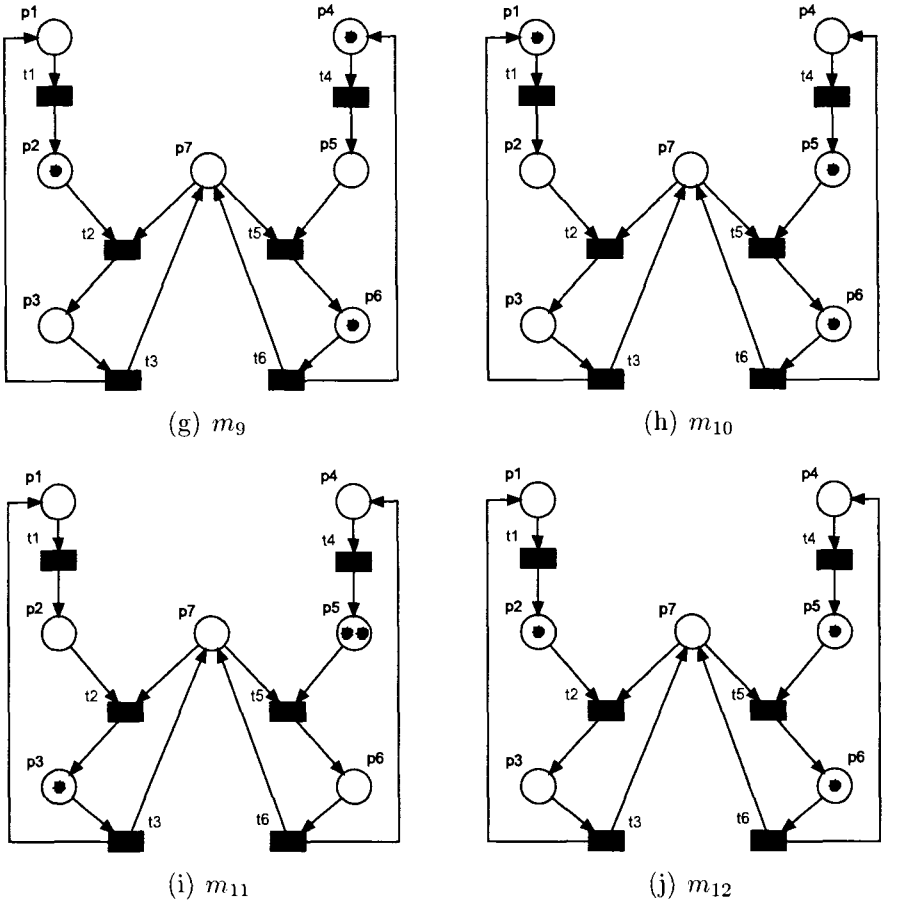


(c) m_5 (d) m_6



(e) m_7 (f) m_8

Figure 1.3: Marquages atteignables



(g) m_9 (h) m_{10}

(i) m_{11} (j) m_{12}

Figure 1.3: Marquages atteignables (suite)

état cohérent, il n'y a des jetons simultanément dans les places P_3 et P_6.

Une succession de franchissements de transition est appelée séquence de franchissement. Dans notre exemple, si à partir de $\mathbf{m_0}$ les transitions T_1 et ensuite T_4 sont franchies, le marquage obtenu est $\mathbf{m_4}$. T_1T_4 est une séquence de franchissement qui est décrite ainsi :

$$\mathbf{m_0} \xrightarrow{T_1T_4} \mathbf{m_4}.$$

Avant de poursuivre avec l'introduction de la représentation en algèbre linéaire des

réseaux de Pétri, il est utile de fournir une définition plus formelle des réseaux de Pétri.

Définition 1.1. Un réseau de Pétri marqué $\mathcal{Q}$ est défini par le quintuple (P, T, Pre, $Post$, $\mathbf{m}$) où :

$P = \{P_1, P_2, \ldots, P_n\}$, est un ensemble fini de places où $n > 0$;

$T = \{T_1, T_2, \ldots, T_m\}$, est un ensemble fini de transitions où $m > 0$;

$P \cap T = \varnothing$, c'est-à-dire que les ensembles P et T sont disjoints ;

$Pre : P \times T \rightarrow \mathbb{N}$, est la matrice d'incidence avant ;

$Post : P \times T \rightarrow \mathbb{N}$, est la matrice d'incidence arrière ;

$\mathbf{m} : P \rightarrow \mathbb{N}$, est le marquage du réseau, soit un vecteur de n composantes $(m_1, m_2, \ldots, m_n)$ où $m(P_i)$ est le nombre de jetons de la place P_i. $\mathbf{m_0}$ est le marquage initial.

Pre(P_i, T_j) est le poids de l'arc $P_i \rightarrow T_j$, cette valeur est un nombre naturel. Post(P_i, T_j) est le poids de l'arc $T_j \rightarrow P_i$, cette valeur est aussi un nombre naturel. Nous pouvons maintenant établir la notation suivante :

${}^\circ T_j = \{P_i \in P | \text{Pre}(P_i, T_j) > 0\}$ = ensemble des places d'entrée de T_j ;

$T_j^\circ = \{P_i \in P | \text{Post}(P_i, T_j) > 0\}$ = ensemble des places de sortie de T_j ;

${}^\circ P_i = \{T_j \in T | \text{Post}(P_i, T_j) > 0\}$ = ensemble des transitions d'entrée de P_i ;

$P_i^\circ = \{T_j \in T | \text{Pre}(P_i, T_j) > 0\}$ = ensemble des transitions de sortie de P_i.

Définition 1.2. La transition T_j est tirable au marquage $\mathbf{m_k}$ si

$m_k(P_i) \geq \text{Pre}(P_i, T_j)$ pour toutes $P_i \in {}^\circ T_j$.

1.1.2.2 Représentation en algèbre linéaire

Un réseau place-transition peut être représenté à l'aide de matrices. Il en est de même pour le franchissement de transition.

La matrice d'incidence avant est ainsi spécifiée :

$$\mathbf{W}^- = [w_{ij}^-], \text{ où } w_{ij}^- = \text{Pre}(P_i, T_j)\,;$$

et la matrice d'incidence arrière est ainsi spécifiée :

$$\mathbf{W}^+ = [w_{ij}^+], \text{ où } w_{ij}^+ = \text{Post}(P_i, T_j)\,;$$

Pour le RdeP de la figure 1.1, nous avons :

$$\mathbf{W}^+ = \begin{array}{c} \begin{matrix} T_1 & T_2 & T_3 & T_4 & T_5 & T_6 \end{matrix} \\ \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{array}\right] \end{array} \begin{matrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \\ P_6 \\ P_7 \end{matrix} \quad \text{et } \mathbf{W}^- = \begin{array}{c} \begin{matrix} T_1 & T_2 & T_3 & T_4 & T_5 & T_6 \end{matrix} \\ \left[\begin{array}{cccccc} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \end{array} \begin{matrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \\ P_6 \\ P_7 \end{matrix}$$

La matrice suivante est la matrice d'incidence d'un RdeP :

$$\mathbf{W} = \mathbf{W}^+ - \mathbf{W}^- = [w_{ij}].$$

Pour notre exemple, nous obtenons la matrice d'incidence suivante :

$$\mathbf{W} = \begin{array}{c} \begin{matrix} T_1 & T_2 & T_3 & T_4 & T_5 & T_6 \end{matrix} \\ \left[\begin{matrix} -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & -1 & 1 & 0 & -1 & 1 \end{matrix}\right] \end{array} \begin{matrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \\ P_6 \\ P_7 \end{matrix}$$

Les valeurs d'une colonne de cette matrice informent des modifications à apporter au réseau lors du franchissement de la transition associée à cette colonne. Par exemple, la colonne de la transition T_1 indique que lors du franchissement de cette transition, il faut retirer un jeton de la place P_1 et en ajouter un à la place P_2. La matrice d'incidence est indépendante du marquage d'un réseau, elle ne donne donc aucune indication sur la tirabilité des transitions.

Soit la séquence de franchissement S, qui peut être exécutée à partir du marquage $\mathbf{m_i}$. Ceci s'écrit $\mathbf{m_i} \xrightarrow{S}$. Par exemple, au marquage $\mathbf{m_{12}}$, illustré à la figure 1.3(j), une séquence de franchissement possible est $\mathbf{m_{12}} \xrightarrow{T_6}$. La séquence de franchissement $S_1 = T_6$ contient une fois la transition T_6 et aucune fois les autres transitions. Le vecteur caractéristique d'une séquence S, qui s'écrit $\mathbf{s}$, est un vecteur où la composante j correspond au nombre de franchissements de la transition T_j dans la séquence S. Le vecteur caractéristique de la séquence S_1 de notre exemple est $\mathbf{s_1} = (0, 0, 0, 0, 0, 1)$.

À l'aide de la matrice d'incidence et d'un vecteur caractéristique, nous pouvons définir l'équation fondamentale des réseaux de Pétri. Cette équation, pour une séquence de franchissement S tel que $\mathbf{m_i} \xrightarrow{S} \mathbf{m_k}$, est la suivante :

$$\mathbf{m_k} = \mathbf{m_i} + \mathbf{W} \cdot \mathbf{s}.$$

Dans notre exemple, cette équation devient :

$$\mathbf{m_4} = \mathbf{m_{12}} + \mathbf{W} \cdot \mathbf{s_1}$$

La solution de cette équation est :

$$\mathbf{m_4} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & -1 & 1 & 0 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

L'équation fondamentale signifie que l'effet d'une séquence de franchissement est entièrement déterminé par la matrice d'incidence et par le vecteur caractéristique de la séquence. Pour cela, le vecteur caractéristique doit correspondre à une séquence de franchissement possible pour le réseau.

1.1.2.3 Invariants

L'équation fondamentale permet de mettre en évidence des quantités invariantes dans un réseau de Pétri. En supposant qu'un réseau n'atteint jamais un état de blocage, le nombre de franchissements de transition, qui font évoluer le réseau de marquage en marquage, est illimité. Dans l'ensemble des marquages accessibles, des propriétés qui ne changent pas malgré les franchissements de transition sont dites invariantes. De plus, d'autres propriétés invariantes peuvent aussi être décelées dans les séquences de franchissements. Ces deux types d'invariant sont appelés invariant de marquage et invariant de franchissement et on leur associe respectivement des composantes

conservatives et des composantes répétitives.

Dans les marquages accessibles du réseau de Pétri modélisant l'algorithme de Peterson de la figure 1.3, nous pouvons voir qu'il n'y a toujours qu'un seul jeton dans les places P_1, P_2 et P_3 et que deux jetons dans les places P_4, P_5 et P_6. En tout temps, $m_1 + m_2 + m_3 = 1$ et $m_4 + m_5 + m_6 = 2$. Ces propriétés invariantes ont pour signification qu'il y a en tout temps une seule instanciation du processus 1 et deux instanciations du processus 2. Ce sont des invariants de marquage. De façon générale, un invariant de marquage est une somme pondérée des jetons de places d'un réseau de Pétri qui demeure constante peu importe le franchissement de transition.

Définition 1.3. Soit R un réseau de Pétri et P l'ensemble de ses places. Il y a un invariant de marquage dans R si pour un sous-ensemble de places $P' = P_1, P_2, ..., P_s$ inclus dans P et un vecteur de pondération $(q_1, q_2, ..., q_s)$ dont toutes les composantes sont positives nous avons :

$$q_1 \cdot m(P_1) + q_2 \cdot m(P_2) + ... + q_s \cdot m(P_s) = \text{constante, pour tout } \mathbf{m} \in \mathcal{M}(\mathbf{m_0}).$$

L'ensemble de places P' est une composante conservative du RdeP R. Les composantes conservatives sont constituées indépendamment du marquage initial d'un RdeP, ce sont des propriétés structurales. Toutefois, la valeur de la constante de l'invariant de marquage dépend du marquage initial.

Certaines séquences de franchissement du réseau de Pétri modélisant l'algorithme de Peterson forment la base de séquences répétitives. C'est le cas des séquences $T_1T_2T_3$ et $T_4T_5T_6$. En effet, nous avons $\mathbf{m_0} \xrightarrow{T_1T_2T_3} \mathbf{m_0}$ et $\mathbf{m_0} \xrightarrow{T_4T_5T_6} \mathbf{m_0}$. Ces séquences induisent un retour à l'état initial et elles pourraient être répétées infiniment.

Définition 1.4. Soit R un réseau de Pétri et T l'ensemble de ses transitions. Il y a

un invariant de franchissement dans R si pour une séquence de franchissements S_k composée du sous-ensemble de transition T' inclus dans T, nous avons :

$$\mathbf{m_0} \xrightarrow{S_k} \mathbf{m_0}.$$

L'ensemble de transitions T' est une composante répétitive du RdeP R.

L'équation fondamentale des RdeP est liée aux invariants, ce qui permet de dégager deux propriétés.

Soit un vecteur pondéré de places $\mathbf{x} = (q_1, q_2, ..., q_n)$ où q_i est une valeur entière de pondération associée à la place P_i. Soit $P(\mathbf{x})$ l'ensemble des places dont le poids est non-nul, ce qui en fait un sous-ensemble de P.

Propriété 1.1. L'ensemble de places B est une composante conservative si et seulement si un vecteur de pondération $\mathbf{x}$ existe tel que

$$P(\mathbf{x}) = B \text{ et } \mathbf{x^T} \cdot \mathbf{W} = \mathbf{0}.$$

Le vecteur $\mathbf{x}$ est un P-invariant.

On peut démontrer ceci à l'aide de l'équation fondamentale où pour une séquence de franchissement S à partir de $\mathbf{m_0}$, le marquage résultant est donné par $\mathbf{m_k} = \mathbf{m_0} + \mathbf{W} \cdot \mathbf{s}$. En multpliant les deux termes de l'équation par $\mathbf{x^T}$, on obtient $\mathbf{x^T} \cdot \mathbf{m_k} = \mathbf{x^T} \cdot \mathbf{m_0} + \mathbf{x^T} \cdot \mathbf{W} \cdot \mathbf{s}$. Enfin, si $\mathbf{x^T} \cdot \mathbf{W} = \mathbf{0}$, nous avons

$$\mathbf{x^T} \cdot \mathbf{m_k} = \mathbf{x^T} \cdot \mathbf{m_0}$$

peu importe la séquence S et donc pour tout marquage $\mathbf{m_k} \in \mathcal{M}(\mathbf{m_0})$. Finalement, $\mathbf{x^T} \cdot \mathbf{m_k}$ est une somme pondérée dont le résultat est l'invariant de marquage.

Dans le modèle de l'algorithme de Peterson, voici ce que donne ces équations dans

le cas de la composante conservative composée des places P_1, P_2 et P_3 où le vecteur de pondération $\mathbf{x_1}$ est (1, 1, 1, 0, 0, 0, 0) :

$$\mathbf{x^T \cdot W} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & -1 & 1 & 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Ceci prouve que l'ensemble de places $P(\mathbf{x_1}) = \{P_1, P_2, P_3\}$ est bien une composante de conservation. Avec le marquage initial $\mathbf{m_0} = (1, 0, 0, 2, 0, 0, 1)$, on obtient la constante de l'invariant de marquage avec l'équation suivante :

$$\mathbf{x^T \cdot m_0} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \\ 0 \\ 0 \\ 1 \end{bmatrix} = 1$$

Ainsi, cet invariant de marquage est $m_1 + m_2 + m_3 = 1$.

Un raisonnement similaire est possible pour les invariants de franchissement.

Propriété 1.2. Soit D, un ensemble de transitions. L'ensemble D est une composante répétitive si et seulement si une séquence de franchissement S, dont le vecteur caractéristique est $\mathbf{s}$, existe tel que :

$T(\mathbf{s}) = D$ et $\mathbf{W} \cdot \mathbf{s} = \mathbf{0}$.

Le vecteur $\mathbf{s}$ est un T-invariant.

Dans le modèle de l'algorithme de Peterson, le vecteur caractéristique de la séquence $S_1 = T_1T_2T_3$ $\mathbf{y_1} = (1, 1, 1, 0, 0, 0)$ est un T-invariant en raison du résultat suivant :

$$\mathbf{W} \cdot \mathbf{y_1} = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & -1 & 1 & 0 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Pour trouver tous les P-invariants d'un réseau de Pétri, l'algorithme 1.1 a été conçu. Celui-ci peut aussi être adapté pour trouver les T-invariants. Cet algorithme permet de trouver les P-invariants de support minimaux. Un P-invariant $\mathbf{x_1}$ est un P-invariant de support minimal s'il n'y a aucun P-invariant $\mathbf{x_2}$ dans le même RdeP tel que $P(\mathbf{x_2}) \subset P(\mathbf{x_1})$.

Algorithme 1.1. Recherche des P-invariants de support minimaux.

Étape 1. Soit $\mathbf{A}$ une matrice unitaire de dimension n (n étant le nombre de places) et $\mathbf{B} = \mathbf{W}$ (la matrice d'incidence). Construire la matrice $[\mathbf{A}|\mathbf{B}]$.

Étape 2. Pour chaque index j de la transition T_j.

Étape 2.1. Ajouter à la matrice $[\mathbf{A}|\mathbf{B}]$ autant de lignes i qu'il y a de combinaisons linéaire avec des coefficents entiers positifs composées de deux lignes, tel que l'élément(i,j) soit nul.

Étape 2.2. Éliminer de la matrice $[\mathbf{A}|\mathbf{B}]$ toutes les lignes k dont l'élément (k,j) n'est pas nul.

Étape 3. Soit $l_A \cdot l_B$ une ligne l de la matrice $[\mathbf{A}|\mathbf{B}]$ et $P(l_A)$ le support de l_A (c'est-à-dire l'ensemble de places pour lesquels le poids n'est pas nul). Si deux lignes, p et q, de la matrice tel que $P(p_A) \supseteq P(q_A)$, alors la ligne p est retirée de la matrice.

Étape 4. Les P-invariants de support minimaux correspondent aux lignes non-nulles de $\mathbf{A}$.

□

Voici le déroulement de l'exécution de l'algorithme 1.1 lorsqu'il est utilisé pour identifier les invariants de support minimaux du modèle de l'algorithme de Peterson. La matrice qui suit est la matrice construite par l'algorithme.

Étape 1. La matrice de 7 lignes (P_1 à P_7) et de 13 colonnes est construite.

Étape 2. *Étape 2.1.* pour T_1 : la ligne $P_1 + P_2$ est ajoutée, il s'agit de la somme des deux premières lignes.

Étape 2.2. pour T_1 : les lignes P_1 et P_2 sont retirées de la matrice.

Étape 2.1. pour T_2 : les lignes $P_3 + P_7$ et $P_3 + (P_1 + P_2)$ sont ajoutées.

Étape 2.2. pour T_2 : les lignes P_3, P_7 et $P_1 + P_2$ sont retirées de la matrice.

Étapes 2.1. et 2.2 pour T_3 : rien à faire.

Étape 2.1. pour T_4 : la ligne $P_4 + P_5$ est ajoutée.

Étape 2.2. pour T_4 : les lignes P_4 et P_5 sont retirées de la matrice.

Étape 2.1. pour T_5 : les lignes $P_6 + (P_3 + P_7)$ et $P_6 + (P_4 + P_5)$ sont ajoutées.

Étape 2.2. pour T_5 : les lignes P_6, $P_3 + P_7$ et $P_4 + P_5$ sont retirées de la matrice.

Étapes 2.1. et 2.2 pour T_6 : rien à faire.

Étape 3. Aucune ligne ne peut être retirée.

Étape 4. Il reste trois lignes à la matrice. Il y a donc trois composantes conservatives dans le modèle, soit $\{P_1, P_2, P_3\}$, $\{P_4, P_5, P_6\}$ et $\{P_3, P_6, P_7\}$. Les P-invariants de support minimaux sont $\mathbf{x_1} = (1,1,1,0,0,0,0)$, $\mathbf{x_2} = (0,0,0,1,1,1,0)$ et $\mathbf{x_3} = (0,0,1,0,0,1,1)$.

	A							B						
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	T_1	T_2	T_3	T_4	T_5	T_6	
	1	0	0	0	0	0	0	−1	0	1	0	0	0	P_1
	0	1	0	0	0	0	0	1	−1	0	0	0	0	P_2
	0	0	1	0	0	0	0	0	1	−1	0	0	0	P_3
	0	0	0	1	0	0	0	0	0	0	−1	0	1	P_4
	0	0	0	0	1	0	0	0	0	0	1	−1	0	P_5
	0	0	0	0	0	1	0	0	0	0	0	1	−1	P_6
	0	0	0	0	0	0	1	0	−1	1	0	−1	1	P_7
	1	1	0	0	0	0	0	0	−1	1	0	0	0	$P_1 + P_2$
	0	0	1	0	0	0	1	0	0	0	0	−1	1	$P_3 + P_7$
	1	1	1	0	0	0	0	0	0	0	0	0	0	$P_3 + (P_1 + P_2)$
	0	0	0	1	1	0	0	0	0	0	0	−1	1	$P_4 + P_5$
	0	0	1	0	0	1	1	0	0	0	0	0	0	$P_6 + (P_3 + P_7)$
	0	0	0	1	1	1	0	0	0	0	0	0	0	$P_6 + (P_4 + P_5)$

La recherche des P-invariants du modèle de Peterson a permis de retrouver les deux invariants de marquage précédemment identifiés, mais aussi un troisième, dont la composante conservative est $\{P_3, P_6, P_7\}$. Cet invariant correspond à la ressource que les processus 1 et 2 se partagent et sa signification est que la quantité de ressource est invariante. Dans notre exemple, cette quantité est 1, et l'invariant permet de

démontrer que la ressource ne peut être que dans un des trois états suivants à la fois : libre (place P_7), utilisée par le processus 1 (place P_3) ou utilisée par le processus 2 (place P_6).

L'algorithme 1.1 peut facilement être modifié pour la recherche des T-invariants. Il suffit de rechercher les P-invariants de la matrice $\mathbf{W^T}$. En effectant cette transformation, l'algorithme trouve les T-invariants suivants : $\mathbf{y_1} = (1, 1, 1, 0, 0, 0)$ et $\mathbf{y_2} = (0, 0, 0, 1, 1, 1)$. À partir du marquage $\mathbf{m_0}$, ces T-invariants correspondent aux composantes répétitives $T_1T_2T_3$ et $T_4T_5T_6$ que nous avions informellement identifiées précédemment.

1.1.2.4 Autres propriétés

Les réseaux de Pétri ont plusieurs autres propriétés que nous ne présenterons pas puisqu'elles ne sont pas utilisées dans cette thèse. Le lecteur devrait toutefois savoir que la théorie des RdeP permet d'analyser plusieurs propriétés d'un modèle. Par exemple, un RdeP est dit borné si le marquage de chacune de ses places n'est jamais supérieure à une certaine valeur. Aussi, une transition d'un RdeP est dite vivace si cette transition apparaît dans toutes les séquences de franchissement possibles à partir des marquages atteignables.

1.1.3 Extensions à la théorie originale

Certaines limites au pouvoir de modélisation des réseaux place-transition sont rapidement apparues. Afin d'augmenter le nombre d'applications possibles, la théorie des réseaux de Pétri a été étendue et enrichie. On appelle extensions les nouveaux types de réseau de Pétri qui ont été développés. Les extensions ont de nouvelles règles

de fonctionnement qui ont été ajoutées à celles que nous avons déjà définies. Les sous-sections 1.1.3.1 à 1.1.3.3 présentent trois extensions à la théorie des réseaux de Pétri : les réseaux de Pétri colorés, les réseaux de Pétri stochastiques et les réseaux de Pétri hybrides fonctionels. Ceci n'est pas une présentation exhaustive de toutes les extensions existantes. Nous présentons ces extensions-ci puisqu'elles ont déjà été utilisées pour modéliser des systèmes biologiques.

1.1.3.1 Réseaux de Pétri colorés

Pour modéliser adéquatement certains modèles, il fut nécessaire de distinguer différentes catégories de jetons. Ce qui fut fait en étiquetant les jetons d'un RdeP de différentes couleurs. Les conditions de franchissement d'une transition d'un RdeP coloré peuvent varier selon la couleur des jetons. Cette extension à la théorie des RdeP est ainsi définie :

Définition 1.5. Un réseau de Pétri coloré R est défini par le n-tuple (P, T, Pre, $Post$, C, $\mathbf{m}$) où :

(P, T, Pre, $Post$, $\mathbf{m}$) est un réseau de Pétri tel que défini par la définition 1.1 où les jetons de $\mathbf{m}$ sont identifiés par une couleur ;

$C = \{C_1, C_2, \ldots\}$ est un ensemble de couleurs. Les matrices d'incidence Pre et $Post$ sont des fonctions qui tiennent compte de la couleur des jetons.

Dans le modèle de l'algorithme de Peterson illustré à la figure 1.1, les deux jetons des instanciations du processus deux ne peuvent être distingués l'un de l'autre. La préservation de l'identité de chacune des instanciations est possible en coloriant les jetons, ce qui permet de savoir exactement dans quel état chaque instanciation du processus se trouve. Conceptuellement, les RdeP colorés sont des superpositions de

réseaux place-transition. Une intéressante conséquence de cette propriété est la réduction de la taille de certains modèles puisque la superposition réduit le nombre de places et de transitions.

1.1.3.2 Réseaux de Pétri stochastiques

Les réseaux de Pétri stochastiques sont une sous-catégorie de réseaux non-autonomes : les réseaux de Pétri temporisés. Ces derniers ont été développés afin d'introduire une dimension temporelle à la théorie originale. En effet, dans les réseaux place-transition, le franchissement de transition est un événement discret instantané qui a une position relative à d'autres franchissements puisqu'il fait partie d'une séquence, mais qui n'est pas situé sur une ligne de temps absolue. Les réseaux temporisés introduisent une abstraction temporelle par la notion d'horloge. Il devient possible de spécifier des délais pour le franchissement de transition. Autrement dit, les transitions sont tirées à des temps données. Les temps de franchissement peuvent être définis de façon statique (délai constant) ou de façon dynamique (délai variable dépendant du marquage).

Dans les RdeP stochastiques, les délais de franchissement sont des valeurs stochastiques. Le franchissement d'une transition se produira lorsqu'un temps d, qui est une valeur aléatoire, se sera écoulé. Cette extension à la théorie des RdeP est ainsi définie :

Définition 1.6. Un réseau de Pétri stochastique R est défini par le n-tuple (P, T, Pre, $Post$, F, λ, $\mathbf{m}$) où :

(P, T, Pre, $Post$, $\mathbf{m}$) est un réseau de Pétri tel que défini par la définition 1.1 ;

$F = \{F_{T_i},\ T_i \in T \mid F_{T_i} : [0, \infty) \rightarrow [0, 1]\}$ est un ensemble de fonctions de

répartition des délais de franchissement des transitions de R. Elles ont une moyenne égale à un et elles sont indépendantes du marquage ;

$\lambda = \{\lambda_{T_i}, T_i \in T \mid \lambda_{T_i} : \mathbb{N} \rightarrow \mathbb{R}^+\}$ est l'ensemble des taux de franchissment. Ce taux est un nombre réel positif qui est utilisé pour calculer la fonction de répartition de la transition T_i

De la définition 1.6, on peut comprendre que le délai d d'une transition suit une fonction de répartition ayant un taux spécifique. Dans le cas où toutes les fonctions de répartition d'un modèle suivent une loi exponentielle, on peut transformer le modèle en chaîne de Markov.

1.1.3.3 Réseaux de Pétri hybrides fonctionnels

Les réseaux de Pétri hybrides fonctionnels (RdePHF) sont une extension qui a été conçue spécifiquement pour la modélisation de systèmes de biologie moléculaire (Matsuno et al., 2003c). Cette extension combine les règles de fonctionnement d'autres extensions existantes. Avant de définir formellement les RdePHF, voyons les autres extensions dont ils empruntent certaines caractéristiques, soit les extensions hybride, à arc inhibiteur et fonctionnelle.

La nature discrète des réseaux de Pétri (quantités discrètes et événements discrets) est une autre limite à la modélisation de certains systèmes qui est devenue indésirable. On a alors imaginé des RdeP continus formés de places contenant des quantités continues et des transitions à flux continu. Ce type de réseau est parfait pour modéliser des flux, que ce soit le flux d'un liquide ou la production continue d'une machine. Un modèle de RdeP continu a son équivalent en équations différentielles. Toutefois, tout flux peut être subitement interrompu par un événement considéré discret, par

exemple la fermeture d'une conduite ou le bris d'une machine. Une extension à la théorie des RdeP ayant une nature mixte, discrète et continue, permet de modéliser de tels phénomènes. Ce sont les réseaux de Pétri hybrides. Les RdeP hybrides peuvent être autonomes ou non. Dans ce dernier cas, on parle aussi de RdeP hybrides temporisés. Les places continues d'un RdeP hybride ne peuvent pas contenir une quantité de jetons indivisibles comme les places discrètes. On parle alors de la marque d'une place continue, qui est un nombre réel.

Comme nous l'avons vu à la définition 1.2, les conditions de franchissement des transitions d'un réseau place-transition ne peuvent qu'être des énoncés de la forme $m_k(P_i) \geq \text{Pre}(P_i, T_j)$ pour toutes $P_i \in {}^\circ T_j$. Ce qui signifie que le marquage de toutes les places d'entrée d'une transition doit être égal ou supérieur au poids des arcs reliant les places d'entrée à la transition. Par définition, le poids d'un arc reliant une place à une transition est égal ou supérieur à 1. Pour modéliser des situations où une condition de franchissement est un test négatif, un test de seuil, c'est-à-dire qu'une transition devient tirable seulement si le marquage d'une place est inférieur au seuil s, on a créé les arcs inhibiteurs. Cette condition de franchissement est représentée ainsi : si le poids de l'arc inhibiteur $P_i \rightarrow T_j$ est s, alors la transition T_j est tirable si $m(P_i) < s$. Les RdeP qui incluent ce type d'arc sont les RdeP à arc inhibiteur. Lorsqu'une transition est franchie, les places d'entrée reliées à cette transition par un arc inhibiteur n'ont aucun jeton de retirer. Un arc inhibiteur sert exclusivement à des tests de seuil et non à la consommation de jetons par une transition. Il existe aussi les arcs de test, qui ont un comportement similaire. Un arc de test permettra le franchissement d'une transition sans consommation de jetons si la condition suivante est remplie : la transition T_j est tirable si $m(P_i) > s$.

Toutes les extensions que nous avons vues jusqu'ici ont une structure statique. Ceci empêche la modélisation de systèmes où certains paramètres s'adaptent dynami-

quement. Dans les termes de la théorie des réseaux de Pétri, cette adaptation est dépendante du marquage. Par exemple, on pourrait ajouter à la définition d'un réseau place-transition que le poids des arcs est fonction du marquage du réseau. Ainsi, après chaque franchissement de transition, le poids des arcs est mis à jour selon le nouveau marquage. Un réseau qui s'adapte dynamiquement est un réseau fonctionnel.

En combinant les règles de fonctionnement des RdeP hybrides, à arc inhibiteur et fonctionnels, nous obtenons les RdePHF.

Définition 1.7. Un réseau de Pétri hybride fonctionnel R est défini par le n-tuple $(P, T, h, A, b, V, d, \mathbf{m})$ où :

$P = \{P_1, P_2, \ldots, P_n\}$, est un ensemble fini de places où $n > 0$;

$T = \{T_1, T_2, \ldots, T_m\}$, est un ensemble fini de transitions où $m > 0$;

$P \cap T = \varnothing$, c'est-à-dire que les ensembles P et T sont disjoints ;

$h : P \cup T \rightarrow \{\texttt{discret}, \texttt{continu}\}$, est la fonction hybride qui indique pour chaque place $P_i \in P$ si c'est une place discrète, $h(P_i) = \texttt{discret}$, ou une place continue, $h(P_i) = \texttt{continu}$. La fonction hybride s'applique de la même façon à chaque transition $T_i \in T$. Ceci permet de diviser P en deux sous-ensembles (P^D, l'ensemble des places discrètes, et P^C, l'ensemble des places continues). T est aussi divisible en T^D et T^C ;

$A = PT \cup TP$ où $PT = P \times T$ et $TP = T \times P$. A est un ensemble fini d'arcs ;

$b : A \rightarrow \{\texttt{normal}, \texttt{test}, \texttt{inhibiteur}\}$, est la fonction de type d'arc qui indique pour chaque arc $A_i \in A$ si c'est un arc normal, $b(A_i) = \texttt{normal}$, un arc de test, $b(A_i) = \texttt{test}$, ou un arc inhibiteur, $b(A_i) = \texttt{inhibiteur}$. Les restrictions suivantes s'appliquent au triplet place, transition et arc de R : si $b(A_i) = \texttt{normal} \wedge A_i \in$

$P^D \times T \Rightarrow A_i \in P^D \times T^D$ et si $b(A_i) = \texttt{normal} \wedge A_i \in T \times P^D \Rightarrow A_i \in T^D \times P^D$. Ces restrictions signifient que si une place d'entrée ou une place de sortie d'une transition reliées par un arc normal est discrète, la transition doit nécessairement être discrète aussi ;

$V = \{f_{A_i^D}(m_1, ..., m_n) \cup g_{A_i^C}(m_1, ..., m_n),\ A_i^D \in P \times T^D \cup T^D \times P,\ A_i^C \in P \times T^C \cup T^C \times P \mid f : \mathcal{M} \to \mathbb{N},\ g : \mathcal{M} \to \mathbb{R}^+\}$, est un ensemble de fonctions assignées aux arcs de R qui ont le marquage $\mathcal{M}$ du réseau comme paramètres. Lorsqu'un arc est relié à une transition discrète, son poids est un nombre naturel. Lorsqu'un arc est relié à une transition continue, sa vitesse est un nombre réel positif ;

$d : T^D \to \mathbb{R}^+$, est l'application appelé délai. Pour une transition discrète T_i, $d(T_i) : \mathcal{M} \to \mathbb{R}^+$ est la fonction de délai de T_i ;

$\mathbf{m} : P \to \mathbb{N}$, est le marquage du réseau, soit un vecteur de n composantes $(m_1, m_2, \ldots, m_n)$ où $m(P_i)$ est le nombre de jetons (la marque) de la place P_i discrète (continue). $\mathbf{m_0}$ est le marquage initial.

Les RdePHF ont un grand pouvoir de modélisation, très utile pour la simulation. Toutefois, certaines propriétés des réseaux place-transition n'ont pas d'équivalent pour les RdePHF, ce qui affaiblit les possibilités d'analyse qualitative des systèmes modélisés avec cette extension. Il existe aussi les RdePHF avec extension, qui ajoute au RdePHF la capacité de modéliser d'autres types de données comme des booléens et des chaînes de caractères. Pour une définition formelle de cette extension, voir Nagasaki et al. (2004).

1.2 Les réseaux de Pétri et leur application aux systèmes biologiques de 2004 à aujourd'hui

Lors de travaux précédents sur l'utilisation des réseaux de Pétri pour la modélisation et la simulation de systèmes de biologie moléculaire, nous avons complété une recension des écrits sur le sujet (Hardy et Robillard, 2004). Cette revue de la littérature couvrait les publications à partir du premier article suggérant l'application des réseaux de Pétri à la biologie (Reddy et al., 1993) jusqu'aux parutions du début de l'année 2004. Depuis, plusieurs autres articles ont été publiés, suggérant de nouvelles applications de la théorie de réseaux de Pétri tant pour l'analyse qualitative que quantitative de systèmes biologiques. Cette section présente ces nouveaux articles. Nous verrons dans un premier temps d'autres revues de littérature parues au cours de cette période. Ensuite, nous recenserons des études dont l'objectif est l'analyse qualitative à l'aide des réseaux de Pétri et finalement les études quantitatives de systèmes biologiques.

1.2.1 Recensions des écrits

Quelques mois avant que notre première revue de la littérature sur l'utilisation des réseaux de Pétri pour l'étude de systèmes biologiques soit publiée, une autre très courte revue de littérature paraissait (Pinney et al., 2003). Cet article entreprend de façon beaucoup plus succincte de présenter les différentes avenues offertes aux biologistes qui désirent utiliser les réseaux de Pétri pour la modélisation et la simulation de systèmes biologiques, soit l'analyse structurale et l'analyse quantitative. Suivant de peu la publication de notre article, une troisième recension des écrits fut publiée (Peleg et al., 2005). Cet article propose une revue des différents formalismes et outils basés sur les réseaux de Pétri. Une particularité intéressante de cet article est

de comparer les formalismes et les outils selon leurs capacités mathématiques ainsi que leur aptitude à répondre à certaines questions pertinentes à l'étude de systèmes biologiques. Cette approche en fait un article très pratique pour les biologistes. Les auteurs comparent cinq outils : Möbius (Deavours et al., 2002), TimeNET (German et al., 1995), Design/CPN (Jensen, 1992), Genomic Object Net (Nagasaki et al., 2003; Doi et al., 2003b) et Woflan (Verbeek et van der Aalst, 2000).

Plus récemment, une quatrième recension des écrits a été publiée (Matsuno et al., 2006b). Similairement aux autres revues, cet article présente les différents formalismes de réseau de Pétri appliqués à la biologie et discute de leur utilisation. Encore une fois, la distinction entre modélisation qualitative et modélisation quantitative est utilisée pour présenter l'ensemble des études publiées. Nous utiliserons, nous aussi, cette catégorisation pour recenser les études sur le sujet.

1.2.2 Études qualitatives de voies biologiques

Reddy et al. (1993) ont suggéré pour la première fois de représenter une voie métabolique à l'aide d'un réseau de Pétri. Leur travail a constitué la fondation théorique de plusieurs autres projets de modélisation qui ont suivi. Dans leur article, les places et les transitions du modèle sont respectivement associées aux métabolites et aux réactions biochimiques. Le poids des arcs sont les coefficients stoechiométriques des réactions entre substrats et produits. Le nombre de jetons indique le nombre de molécules des différents métabolites. Ces règles ont servi aux études subséquentes de systèmes biologiques qui, jusqu'en 2004, ont porté essentiellement sur les voies métaboliques (Hofestädt, 1994; Küffner et al., 2000; Voss et al., 2003). Zevedei-Oancea et Schuster (2003) ont discuté d'autres rapprochements possibles entre les voies métaboliques et la théorie des réseaux de Pétri. Entre autres, ils ont exposé la correspondance entre les relations de conservation et les flux élémentaires des voies métaboliques et les

P-invariants et les T-invariants des réseaux place-transition.

Depuis 2004, les voies de signalisation cellulaire sont maintenant l'objet d'étude principal des nouvelles publications. Ce déplacement dans le type de systèmes de biologie moléculaire étudiés indique une augmentation du niveau de complexité des systèmes modélisés avec les RdeP. Les voies métaboliques sont constituées de mécanismes simples, une suite uniforme de réactions catalytiques. Les voies de signalisation cellulaire sont plus complexes. Elles sont formées par différents types de réactions tels la formation de complexes moléculaires, des réactions catalytiques et le transport de molécules. La combinaison de ces réactions permet la propagation d'un signal moléculaire qui provoque le changement d'état de molécules d'un état inactif à un état actif.

Voyons maintenant les différentes études qualitatives de voies et de systèmes biologiques basées sur les réseaux de Pétri menées de 2004 à aujourd'hui.

1.2.2.1 Analyse de voies métaboliques

Les derniers articles portant spécifiquement sur l'analyse de voies métaboliques à l'aide de RdeP sont ceux de Heiner et Koch (2004) et Koch et al. (2005). Ces deux articles constituent en fait la poursuite du même projet où les auteurs proposent une méthode pour la modélisation et la validation de voies métaboliques. Cette méthode est basée sur les RdeP colorés et constitue une excellente synthèse des propositions antérieures d'applications de la théorie de RdeP aux voies métaboliques. L'étape de modélisation de cette méthode reprend les règles de modélisation de Reddy et al. (1993). L'étape de validation propose différentes analyses (analyse structurale, analyse des invariants et analyse d'accessibilité pour vérifier la cohérence du modèle et déduire certaines de ses propriétés structurales et dynamiques. Cette méthode a été

appliquée à la voie du catabolisme du sucrose dans le tubercule de patate (Koch et al., 2005).

1.2.2.2 Analyse de voies de signalisation cellulaire

On a tenté d'adapter la méthode de modélisation et de validation de voies métaboliques aux voies de signalisation cellulaire (Heiner et al., 2004). Pour modéliser les réactions enzymatiques de ce type de voies, on a fait appel aux arcs de test. Pour effectuer la validation du modèle, les arcs de test sont convertis en arcs unidirectionnels afin de déterminer les T-invariants et ainsi d'offrir une description des flux dans la voie de signalisation. Cette méthode a été appliquée à l'apoptose cellulaire.

Lee et al. (2004) ont cherché à définir un nouveau cadre conceptuel pour la modélisation de voies de signalisation en se basant sur le formalisme des RdeP. Entre autres, ils ont suggéré des représentations utilisant les réseaux de Pétri pour des réactions types d'une voie de signalisation (transformation chimique, translocation, association, dissociation, activation enzymatique et cascade de signalisation). Ils ont appliqué leur cadre conceptuel à la voie de signalisation induite par l'interleukine-1β et le facteur-α de nécrose tumorale.

En analysant les différentes utilisations faites des RdeP pour la modélisation des voies de signalisation, Takai-Igarashi (2005) a remarqué des incohérences. Elle ne remet pas en cause l'utilité de ces approches pour la modélisation et l'analyse des complexes voies de signalisation, mais elle entrevoit des difficultés futures pour la réutilisation, le partage et la modification des modèles. Elle proposa une standardisation des représentations par RdeP des processus de signalisation cellulaire qui repose sur le Cell Signaling Network Ontology (Takai-Igarashi et Mizoguchi, 2004).

En parallèle, deux articles sont parus en 2006 qui témoignent, à mon avis, d'une maturation de ce champ de recherche. Sackmann et al. (2006) et Li et al. (2006) explorent des applications pointues de techniques d'analyse basées sur les RdeP et tentent de les arrimer aux concepts des voies de signalisation.

D'une part, Li et al. (2006) poursuivent le travail de Lee et al. (2004) et proposent des modèles de RdeP pour d'autres réactions types des voies de signalisation (entre autres, phosphorylation, autophosphorylation, homodimérisation ainsi que l'ouverture et la fermeture de canaux). D'autre part, ils introduisent un nouveau concept, la *composante d'activation de transduction*, qui permet de décrire le processus d'activation enzymatique des réactions des voies de signalisation. Ils montrent la correspondance entre ce concept et les T-invariants des modèles de RdeP. Ils suggèrent un algorithme qui identifie les composantes d'activation de transduction et qui établit les liens entre celles-ci. Cet algorithme nécessite une identification des places qui modélisent une enzyme. Cette méthode permet de décomposer une voie de signalisation en blocs d'activation qui agissent séquentiellement les uns sur les autres pour faire passer des enzymes d'un état inactif à un état actif et ainsi faire transiter un signal par la voie. Ils démontrent leur méthode sur un modèle des voies de signalisation stimulées par l'interleukine-3 et ils obtiennent comme résultat l'architecture d'activation de la voie.

L'approche de Sackmann et al. (2006) atteint un objectif similaire de décomposition du modèle d'une voie de signalisation en unités fonctionnelles, qui sont les plus petits regroupements possibles ayant une signification biologique. Cette approche, au lieu d'utiliser une identification explicite des places représentant une enzyme, repose sur les P-invariants du RdeP. En effet, une enzyme est une entité constante du modèle puisqu'elle est toujours dans au moins un état. L'action d'une enzyme est modélisée par un arc de lecture (*read arc*, c'est-à-dire un arc double, à la fois d'entrée

et de sortie, qui entraîne une consommation et une production de jetons nulles). La présence d'arcs de lecture complique l'identification de T-invariants à partir de la matrice d'incidence. Les auteurs de cet article proposent donc le concept de T-invariant faisables, des T-invariants réalisables à partir du marquage d'un réseau. Les T-invariants faisables ont tous une signification biologique. Les auteurs définissent un second concept, les ensembles maximaux de transitions communes, qui regroupent des transitions qui apparaissent toujours ensemble dans les T-invariants. Ces ensembles représentent des réactions qui sont toujours exécutées ensemble pour accomplir une fonction biologique. Les conclusions de cette approche servent principalement à valider les modèles complexes. La voie de signalisation de la phéromone de reproduction de *Saccharomyces cerevisiae* sert d'exemple d'application de cette approche.

1.2.2.3 Analyse du système corporel d'homéostasie du fer

L'approche d'analyse que nous venons de recenser fut reprise pour le système d'homéostasie corporel du fer. Dans une première étude, un modèle de RdeP a été développé (Formanowicz et al., 2007). Selon les auteurs, il s'agit du premier modèle formel de ce système qui n'est toujours pas complètement élucidé. Dans une seconde étude, ce modèle a été analysé en y identifiant les T-invariants faisables et les ensembles maximaux de transitions communes (Sackmann et al., 2007). Ce modèle comporte 85 T-invariants et 14 ensembles-MTC qui ont tous une signification biologique. Les auteurs utilisent aussi une nouvelle technique, le regroupement des T-invariants faisables en grappes selon la méthode UPGMA (Grafahrend-Belau, article à paraître). Les grappes identifiées correspondent aux voies biologiques du système.

1.2.3 Études quantitatives

Jusqu'à présent, nous avons recensé les études qui utilisent le cadre théorique des réseaux de Pétri pour analyser qualitativement les systèmes biologiques. L'avantage principal d'une telle approche est de permettre la validation d'un modèle. Ces modèles qualitatifs deviendront réellement utiles que s'ils permettent par la suite une étude quantitative, réelle génératrice de nouvelles hypothèses à être testées expérimentalement puisque les études quantitatives permet d'investiguer le comportement dynamique des systèmes.

Comme nous l'avons vu à la section 1.1.3.3, une extension particulière de la théorie des RdeP a été spécifiquement conçue pour réaliser des études quantitatives de systèmes biologiques. Il s'agit des réseaux de Pétri hybrides fonctionnels. Beaucoup d'études ont été menées à l'aide de ce formalisme dédié à la simulation. Quelques autres études quantitatives ont aussi été conduites à l'aide d'autres extensions de RdeP. Nous les répertorions ici.

1.2.3.1 Analyse quantitative de voies biochimiques avec les réseaux de Pétri hybrides et leurs extensions dérivées

La modélisation du mécanisme de commutation génétique du phage-λ avec les RdeP hybrides a démontré pour la première fois que ce formalisme avait un intéressant potentiel pour la modélisation et la simulation des voies biochimiques (Matsuno et al., 2000). Depuis, plusieurs types de modèles ont été développés, simulés et analysés.

- Réseaux de régulation génétique
  - Commutateur génétique du phage λ (Matsuno et al., 2000)
  - Rythmes circadiens de la mouche à fruit (Matsuno et al., 2003c)

  - Rythmes circadiens de la souris (Matsuno et al., 2006a)
  - Régulation de gènes du cancer (p53, MDM2 et p19ARF) (Doi et al., 2006)
- Voies de signalisation cellulaire
  - Apoptose induite par la protéine Fas (Matsuno et al., 2003c)
  - Voie de signalisation Notch-Delta chez *Drosophila* (Matsuno et al., 2003b)
  - Voie de la drogue chimiothérapique gemcitabine (Peleg et al., 2005)
  - Maladie d'Huntington (Nagasaki et al., 2005)
  - Rôle d'interleukine-6 dans le développement des cellules souches hématopoïétiques (Troncale et al., 2006)
  - Effet de l'inhibiteur de la kinase Raf sur la voie de signalisation de la kinase ERK (Gilbert et Heiner, 2006)
  - Voies de signalisation des kinases Akt et MAPK (Koh et al., 2006)
- Voies métaboliques
  - Cycle de l'urée et sa régulation (Chen et Hofestädt, 2003)
  - Opéron *lac* et la voie de la glycolyse (Matsuno et al., 2003a)
- Réseaux protéiques des cycles cellulaires
  - Processus de division cellulaire de *Xenopus* (Matsui et al., 2004)
  - Cycle de fission cellulaire de la levure (Fujita et al., 2004)

Pour la plupart des études citées précédemment, l'approche est similaire. On procède à la modélisation d'un système biochimique en structurant les molécules et leurs interrelations. On spécifie les paramètres dynamiques (concentrations, paramètres cinétiques) pour ensuite simuler le modèle. Les données de simulation sont des séries de données temporelles, principalement la variation dans le temps des concentrations de molécules. Une particularité de l'étude de Koh et al. (2006) est de proposer une approche de décomposition pour l'estimation des paramètres dynamique d'un modèle de RdePHF qui tient compte de sa structure. Cette approche diminue de façon importante le temps d'exécution de l'algorithme.

1.2.3.2 Analyse quantitative de voies biochimiques avec d'autres extensions

Depuis 2004, il y a eu peu de tentatives d'étudier quantitativement un système biochimique avec des RdeP d'un autre type que les RdePHF. Lee et al. (2006) ont simulé la voie de signalisation du facteur de croissance épidermique à l'aide de réseaux de Pétri exécutables et de l'outil Design/CPN. Popova-Zeugmann et al. (2005) ont utilisé les réseaux de Pétri temporisés pour transformer un modèle qualitatif à l'état stable en un modèle quantitatif à l'aide des T-invariants. Elles présentent des techniques structurales qui permettent de déterminer la réalisabilité temporelle d'une séquence de transitions et de calculer les plus court et long délais de la séquence. Inspirés par cette dernière approche discrète, Li et al. (2007) ont aussi utilisé les réseaux de Pétri temporisés pour étudier quantitativement la voie de signalisation de l'apoptose cellulaire induite par la protéine Fas. Selon ces derniers, une approche discrète est une étape de modélisation quantitative qui devrait précéder une approche hybride comme les RdePHF. Ils justifient cette affirmation en mentionnant que les modèles hybrides nécessitent un important effort de "tunning" des paramètres par essai-erreur. La simulation des modèles hybrides est aussi très coûteuse en ressources computationnelles. Avec les modèles discrets, on obtient des résultats moins précis, mais qui donnent des indications très pertinentes à une modélisation hybrides. Pour ces trois raisons, Li et al. (2007) suggèrent une première modélisation quantitative discrète avec les RdeP temporisés avant de réaliser une modélisation hybride.

1.3 Les réseaux de Pétri peinturés

Pour visualiser les données produites lors de la simulation d'un RdeP, la plupart des simulateurs supportent la visualisation par jeu de jetons qui permet de voir la

modification de la distribution des jetons dans le réseau. À l'aide d'une animation, on a l'impression que les jetons voyagent au rythme de leur production et de leur consommation. D'autres simulateurs présentent le marquage sous forme numérique directement sur le réseau ou encore, ils sont équipés d'une fonctionnalité permettant la production de graphiques de données bidimensionnels où des courbes représentent la variation de la marque des places dans le temps. Quand les modèles atteignent un degré élevé de complexité, ces méthodes de visualisation des données de simulation ne suffisent plus à donner rapidement une idée du comportement du modèle simulé. Le simple coup d'oeil n'informe plus vraiment et une analyse plus approfondie devient nécessaire.

Afin de conserver la possibilité d'explorer visuellement les données de simulation d'un modèle de RdeP complexe, nous avons imaginé un ajout simple à la théorie que nous avons appelé les RdeP peinturés. Nous avons présenté cette idée pour la première fois dans l'article de conférence référencé ici (Hardy et Robillard, 2006b), qui se trouve à l'annexe II. Les réseaux de Pétri peinturés consistent en l'application de teintes sur chacune des places du modèle. Une règle de peinturage, qui a comme principal paramètre d'entrée la marque d'une place, sert à déterminer la teinte à appliquer aux places. Différentes règles de peinturage peuvent être créées afin de générer des représentations qui révèlent différents aspects des données. À la sous-section 5.3.4, nous présentons deux règles différentes de peinturage.

Les teintes des places du modèle sont calculées à chaque pas de simulation, ce qui permet ensuite de produire une représentation animée du graphe du modèle de RdeP. Dans cette représentation animée, les teintes changent pour refléter l'évolution du marquage du réseau. Les RdeP peinturés permettent donc d'apprécier visuellement le comportement dynamique du modèle. Nous croyons que cet ajout présente synthétiquement les données de simulation et complète très bien la théorie existante et

supporte l'objectif principal du formalisme de RdeP est de représenter comment un système fonctionne.

CHAPITRE 2

UNE APPROCHE SYSTÉMIQUE EN BIOLOGIE MOLÉCULAIRE : LE GÉNIE À LA RENCONTRE DE LA BIOLOGIE

2.1 Introduction

Depuis quelques années, des scientifiques observent une profonde transformation de la recherche en biologie. En 50 ans de recherche, plus précisément depuis la découverte de la structure de l'ADN en 1953 jusqu'à la fin du séquençage du génome humain en 2001, les chercheurs en biologie ont majoritairement adopté une approche de recherche réductionniste. En 1966, Francis Crick a affirmé que "L'objectif ultime de la biologie moderne est d'expliquer tous les phénomènes biologiques en termes provenant de la physique et de la chimie." Crick est cité par Van Regenmortel (2004). Cette phrase résume bien tout le courant de pensée qui a prévalu en biologie pendant la deuxième moitié du siècle dernier. C'est ainsi que tous les éléments participant aux processus biologiques : organes, cellules, organites et macromolécules ont été décortiqués, analysés, caractérisés, etc.

Mais cette démarche scientifique a certaines limites et une nouvelle démarche a émergé : la biologie systémique, appelée *systems biology* en anglais. Cette approche prône la compréhension des processus biologiques comme des systèmes hautement complexes, ayant une structure et une dynamique propres, des mécanismes de contrôle et des principes de conception (Kitano, 2002b). Par exemple, il ne s'agit plus d'associer des fonctions précises à des protéines, mais d'identifier et comprendre les réseaux d'interactions où elles sont actives.

Les partisans de la nouvelle approche systémique et holistique considèrent maintenant que la biologie est une discipline autonome qui requiert son vocabulaire propre et des concepts qui ne sont issus ni de la chimie ni de la physique (Van Regenmortel, 2004). Toutefois, les bases de cette nouvelle biologie sont difficiles à imaginer et à poser de façon consensuelle. Dans un texte d'anticipation scientifique étonnamment juste, Hartwell et al. (1999) stipulent que les systèmes biologiques sont très différents des systèmes chimiques et physiques : les systèmes physiques sont composés de copies d'éléments habituellement simples, alors que les systèmes biologiques contiennent des milliers de composantes différentes, variant en nombre de quelques unités à des millions, ayant des interactions précises et formant souvent des dispositifs plus complexes. Selon eux, plus importantes encore sont les contraintes qui distinguent la biologie de la physique et de la chimie, à savoir la survie et la reproduction des organismes vivants, et la notion concomitante de fonctionnalité.

Quelques chercheurs ont avancé, comme Hartwell, qu'une meilleure compréhension du vivant proviendrait d'une application des principes d'ingénierie à la biologie. Ceux-ci sont d'avis que les principes d'organisation des systèmes biologiques se rapprochent de ceux des systèmes technologiques hautement complexes et que plusieurs parallèles sont à établir entre les deux.

Wolkenhauer (2001) a fait remarquer que les origines de la biologie systémique ne sont pas si récentes puisque des biologistes avaient déjà émis les mêmes principes de base il y a quelques décennies. Toutefois, c'est à partir des informations biologiques sur les génomes, les protéomes et autres *omes* biologiques qui sont amassées avec les technologies d'aujourd'hui que cette discipline peut réellement prendre son envol. Les discussions qui s'amorcent entre ingénieurs et biologistes pour bâtir ce nouvel axe de recherche en biologie offrent de nouvelles opportunités. Mes activités de recherches, menées au Laboratoire de Recherche en Génie Logiciel de l'École Polytechnique dans

le cadre de la maîtrise et du doctorat et portant sur les outils de modélisation et de simulation des systèmes de biologie moléculaire, s'inscrivent dans l'optique de cette *nouvelle* biologie puisque les chercheurs, tant théoriciens qu'expérimentalistes, réclament de nouveaux outils informatiques.

Cette section est structurée en deux sous-sections distinctes. La première sous-section, intitulée *Concepts systémiques* en biologie, fait un tour d'horizon des concepts de la théorie des systèmes qui peuvent être appliqués à la biologie et qui ont été discutés dans de nombreuses publications. Ces concepts sont l'émergence, la robustesse, la fragilité, la modularité et la rétroaction. Y est aussi présentée l'architecture en noeud papillon qui est une structure architecturale particulière qui rassemble tous les concepts de la biologie systémique. La deuxième sous-section, intitulée *Méthodes d'analyse en biologie systémique*, retrace certaines études qui utilisent et illustrent les concepts de la biologie systémique. Certains résultats de ces études sont présentés. Il s'agit d'études de la robustesse, de la rétroaction et de la modularité des systèmes biologiques. Nous présenteront aussi des types particuliers d'études, à savoir de nouvelles études théoriques en théorie des systèmes et du contrôle qui sont motivées par la biologie systémique dans un cas, et dans l'autre, des études purement expérimentales dont les objectifs sont ceux de la biologie systémique.

2.2 Concepts systémiques en biologie

Un des premiers défis de la biologie systémique naissante est d'établir un cadre conceptuel. Parmi toutes les suggestions qui ont été émises dans la littérature spécialisée, la nature complexe des systèmes biologiques semble faire consensus (Weng et al., 1999). De cette complexité découle différents principes d'organisation qui pourraient constituer la structure conceptuelle et la cohérence logique qui sont actuellement re-

cherchées. Cette sous-section présente ces différents concepts et certaines études qui les supportent.

Plusieurs concepts théoriques qui sont décrits ci-bas proviennent des travaux de Csete et Doyle (2002). John C. Doyle a participé à titre d'auteur ou de collaborateur à plusieurs des publications décortiquées dans ce rapport et il a, avec Hiroaki Kitano, une influence indéniable sur le domaine. Cet ascendant exercé sur la théorie de la biologie systémique par Doyle est principalement dû au développement pionnier qu'il a accompli avec Carlson sur la théorie HOT (*Highly Optomized Tolerance*). Ses propositions sont issues de ses réflexions sur la complexité et sur ses caractéristiques qui transcendent les domaines analysés (Carlson et Doyle, 1999, 2002). Les mécanismes de cette théorie reproduisent une observation généralisée dans les systèmes complexes optimisés : la génération de distributions en loi de puissance[1]. Par systèmes complexes optimisés, on entend des systèmes comme les organismes biologiques qui sont optimisés par la sélection naturelle ou comme les technologies avancées qui sont optimisées par les méthodes de conception d'ingénierie. La théorie HOT présente quatre caractéristiques : (1) de l'efficacité, de la performance et de la robustesse pour contrer les incertitudes prévues ; (2) une hypersensibilité aux défauts de design et aux perturbations non-anticipées ; (3) des configurations non-génériques, spécialisées et structurées ; et (4) des lois de puissances. Dans un contexte biologique, nous parlerons entre autres de robustesse, de modularité, de protocole, de rétroaction et de fragilité.

[1]Simplement définie, une loi de puissance implique d'une part un très grand nombre de petites occurrences et d'autre part, une rareté des grandes occurrences, dans une proportion log-log. De telles distributions ont été observées dans la taille des villes, la magnitude des secousses sismiques, la fréquence d'utilisation des mots, pour ne nommer que quelques exemples. Plusieurs distributions de loi de puissance ont aussi été observées dans les statistiques se rapportant à Internet.

2.2.1 Émergence

À mesure que les caractéristiques des macromolécules biologiques sont déterminées, il appert que les propriétés des systèmes biologiques sont rarement le fruit d'acteurs isolés. En d'autres mots, la spécificité des activités complexes des systèmes biologiques ne provient pas de la spécificité des molécules individuelles (Van Regenmortel, 2004). Les interactions entre les composantes ainsi que les influences de l'environnement font apparaître de nouvelles propriétés, qui sont absentes des composantes isolées. Cette constatation a mené à la conceptualisation de l'émergence, qui est un complément aux approches réductionnistes. Une particularité des propriétés émergentes est qu'elles ne peuvent pas être prédites ou déduites à partir d'informations sur les comportements de bas niveau.

Les travaux de modélisation et de simulation sur les voies de signalisation cellulaire impliquées dans la régulation de l'enzyme CaMKII de Bhalla et Iyengar (1999) illustrent bien la recherche de comportements émergents dans les systèmes biologiques. En assemblant plusieurs modules de réactions biochimiques et de voies de signalisation dans un seul modèle informatique, les chercheurs ont observé des propriétés émergentes dans le grand réseau de signalisation résultant. Dans ce cas-ci, il s'agit plus précisément d'une propriété systémique de stockage d'informations à l'intérieur du système. Cette propriété est le résultat de multiples boucles de rétroaction qui supportent un comportement bistable. Ainsi, en fonction de stimuli précis, le réseau de signalisation agit comme un mécanisme binaire qui existe grâce aux différents états stables que le réseau peut atteindre et conserver.

Mesarovic et al. (2004) avancent qu'un deuxième type de relation existe dans un système complexe entre les comportements de haut et de bas niveaux : la coordination. Si l'émergence consiste en la manifestation d'un comportement distinct du

système résultant uniquement de l'interaction des sous-systèmes, alors la coordination serait la double dépendance d'un comportement systémique à l'interaction des sous-systèmes et à un processus de coordination maintenant l'harmonie entre les sous-systèmes. Les auteurs supportent cette théorie par la présence de mécanismes de régulation de l'expression génétique et ils effectuent une analyse de leurs motifs de coordination. Toutefois, je suis d'avis que le bien-fondé scientifique de l'isolation des mécanismes de l'expression génétique à un niveau organisationnel supérieur telle qu'ils le proposent reste à démontrer.

2.2.2 Robustesse

La robustesse est la capacité d'un système à maintenir ses fonctionnalités malgré des perturbations internes et externes (Alon, 2003). Un système biologique devrait être en mesure de résister au bruit occasionné par la modification des paramètres de réaction moléculaire ou par des modifications de concentration. Pour Kitano (2004), la robustesse est un principe organisationnel ubiquitaire chez tous les systèmes complexes qui doivent évoluer. La robustesse englobe de façon générale toutes les propriétés systémiques comme la modularité, l'utilisation de protocoles, la rétroaction et la fragilité.

La présence aujourd'hui de robustesse dans tous les systèmes vivants s'expliquerait par le favoritisme exercé par l'évolution (Kitano, 2004). Cette sélection procède d'un cercle vertueux où d'une part, les systèmes robustes facilitent leur propre évolution à l'aide de leurs propriétés intrinsèques et d'autre part, l'évolution privilégie des traits robustes. Ainsi, les requis pour la robustesse et la capacité d'évolution auraient des objectifs similaires qui se soutiennent mutuellement.

En adoptant un point de vue "robustesse" lorsque l'on étudie les systèmes biologiques,

on remarque que certaines pathologies sont reliées à cette propriété. Par exemple, le cancer, qui est un des pires scénarios possibles pour un système puisque la maladie est maintenue et même encouragée par les mécanismes robustes du système hôte. Les tumeurs prolifèrent et développent des métastases grâce aux mécanismes habituellement sains et essentiels de la reproduction cellulaire, ce qui contrecarre la plupart des thérapies. Un médicament anti-cancer peut réduire la masse d'une tumeur, mais celle-ci peut réapparaître après un certain temps, laissant peu d'espoir au patient d'une guérison complète. Toutefois, à l'aide d'une perspective de robustesse, de nouvelles stratégies cliniques pourraient être explorées. En comprenant les mécanismes cellulaires de robustesse, il serait peut-être possible d'induire la dormance des cellules pour stopper leur reproduction, si dévastatrice lors du cancer.

Wilhelm et al. (2004) distinguent deux types de robustesse : la robustesse structurale et la robustesse dynamique. La robustesse structurale est une propriété qui permet à un système de palier à la perte d'un ou de plusieurs de ses éléments. Ce type de robustesse s'observe par la redondance présente dans un système. La robustesse dynamique consiste en mécanismes d'adaptation ou de contrôle qui compensent pour la modification de paramètres internes ou externes.

2.2.3 Fragilité

Dans la théorie de tolérance grandement optimisée (HOT) de Carlson et Doyle (1999), la robustesse s'accompagne invariablement de fragilité, d'où l'expression *robust, yet fragile*. Ceci peut sembler contradictoire à première vue, mais ce principe se vérifie dans la nature et dans la technologie : tous systèmes complexes semblent vulnérables à des perturbations exceptionnelles. En effet, lorsque des perturbations imprévues dans le design du système surviennent, une cascade souvent catastrophique d'événements mène à la détérioration, à l'arrêt ou même à la destruction du système. Un design

optimal pour un système n'est pas une robustesse à toute épreuve, ce qui nécessiterait une complexité incroyable ainsi qu'une perte importante de performance, mais plutôt une tolérance optimale aux perturbations probables. La nature et les hommes se satisfont de ce compromis.

Une nouvelle interprétation du diabète de type 2 a été proposée où les propriétés de robustesse et de fragilité seraient en cause (Kitano, 2004). Des milliers de générations auraient permis à notre organisme de devenir particulièrement robuste à la famine et à un mode de vie très exigeant en énergie où les risques d'infection sont élevés, mais l'organisme serait vulnérable aux conditions de vie de plus en plus répandues aujourd'hui, soient la suralimentation et un mode de vie très sédentaire. Impuissant devant ces nouvelles conditions imprévues par notre système, et donc expression de sa fragilité, notre organisme développerait des pathologies reliées à cette abondance.

2.2.4 Modularité et protocoles

Le concept de modularité en biologie n'est pas nouveau. Toutefois, il est de plus en plus reconnu que l'organisation en modules fonctionnels est un principe organisationnel majeur en biologie (Hartwell et al., 1999). Un module est une entité discrète dont les fonctions sont séparables de celles d'autres modules. L'isolation chimique est responsable de cette séparation en unités, dont la cause peut être une localisation spatiale restreinte ou une spécificité chimique. Du point de vue de la robustesse des systèmes, la modularité est un mécanisme efficace pour contenir localement les perturbations et les dommages afin d'en minimiser les effets sur l'ensemble d'un système.

Les modules sont indissociables du design des systèmes complexes. Les modules ont différentes caractéristiques : (1) ils ont des interfaces identifiables entre modules ;

(2) ils peuvent être modifiés et évoluer de façon indépendante ; (3) ils facilitent la modélisation par abstraction ; (4) ils conservent une certaine identité même s'ils sont isolés ou réorganisés ; mais (5) ils ont aussi une identité supplémentaire qu'ils dérivent du reste du système avec lequel ils interagissent (Csete et Doyle, 2002).

Un exemple évident de modularité est la cellule, qui interagit avec l'environnement et les autres cellules d'un organisme pluricellulaire via une interface, tout en étant un milieu fermé. Les modules sont aussi souvent organisés en niveaux hiérarchiques : par exemple, le module cellulaire contient d'autres modules plus petits. En plus des modules physiques comme la cellule, il existe aussi des modules fonctionnels, spatiaux et temporaux, qui sont des sous-systèmes des réseaux métaboliques, de signalisation cellulaire et de régulation du développement (Kitano, 2004).

Les protocoles qui régissent les communications entre modules sont tout aussi importants à connaître que les modules eux-mêmes (Csete et Doyle, 2002). Dans cet article de Csete et Doyle référencé précédemment, une intéressante illustration des propriétés des systèmes robustes est faite à l'aide des blocs Lego. Les blocs Lego sont des modules de tailles et de fonctions diverses qui utilisent toutefois le même protocole, un mécanisme à pression, pour relier les blocs entre eux. Un système simple, mais qui permet néanmoins de créer une très grande variété de structures.

2.2.5 Rétroaction

Le contrôle par rétroaction est un mécanisme de régulation qui est essentiel aux systèmes qui doivent être robustes. Une boucle de contrôle est fermée si elle inclut un signal de rétroaction. Ce signal, qui provient de la sortie d'un dispositif, revient en entrée pour agir sur son fonctionnement. La rétroaction assure des réponses dynamiques et permet aux systèmes de diminuer l'impact des perturbations extérieures

et des variations internes. La figure 2.1 présente une boucle de rétroaction.

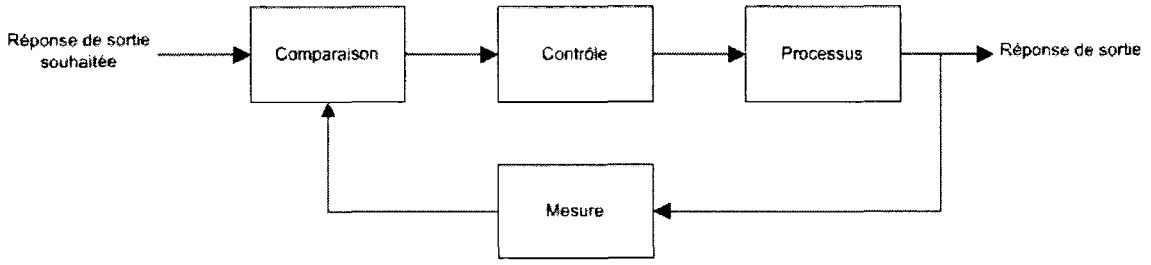


Figure 2.1: Système de contrôle en boucle fermée avec signal de rétroaction

Il existe deux types de rétroaction. La rétroaction négative cherche à renverser la direction d'un changement dans un système afin de lui permettre de s'adapter parfaitement et de le stabiliser. La rétroaction positive contribue, elle aussi, à la robustesse d'un système, mais en amplifiant certains stimuli, accélérant ainsi le passage d'un état à un autre. Un exemple classique de boucle de rétroaction est la régulation de l'opéron lactose, proposé par Jacob et Monod (1961). Les boucles de rétroaction permettent, entre autres, de maintenir un système dans un état homéostatique.

2.2.6 Architecture en noeud papillon

Un des objectifs de la biologie systémique est de trouver un cadre organisationnel en biologie. Dans la recherche de celui-ci, on a identifié deux caractéristiques architecturales qui facilitent de façon compatible l'évolution des systèmes et leur donnent davantage de robustesse : un noyau très stable de processus interfacés à diverses entrées et sorties ; ainsi que des mécanismes versatiles qui soutiennent les processus essentiels du systèmes, de façon à ce que tout nouveau processus qui a une interface adaptable à ceux-ci puisse les utiliser (cette propriété est nommée faible liaison ou *weak linkage*) (Kitano, 2004). Ce type d'architecture est appelé noeud papillon, ou *bow tie*, en raison de sa forme (voir figure 2.2, qui est inspirée de la figure 3 de (Kitano, 2004)). Cette architecture constitue présentement une des principales hypothèses de

principe d'organisation des systèmes biologiques.

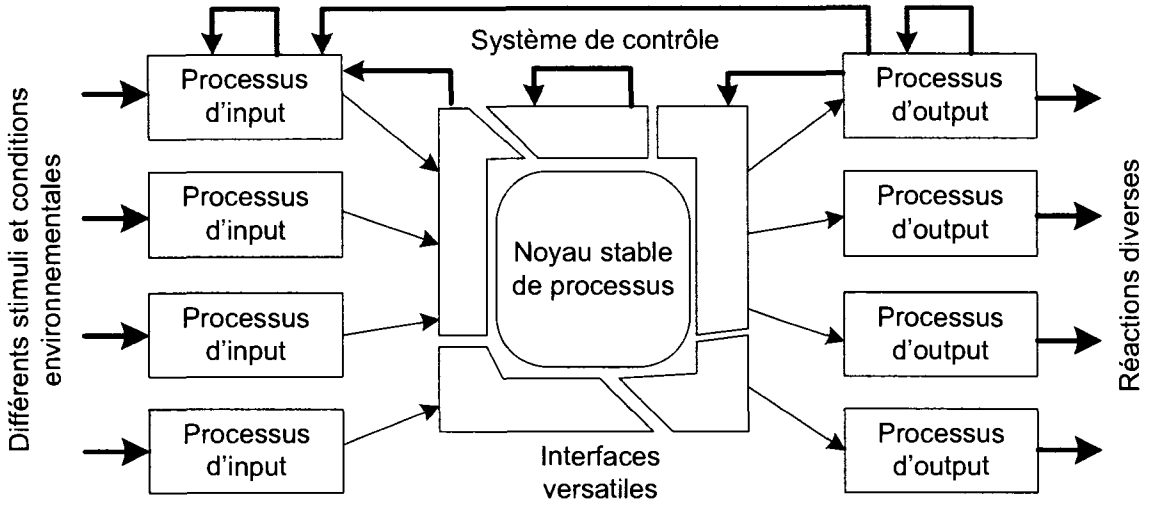


Figure 2.2: Architecture en noeud papillon

Des analyses récentes de génomes supportent l'hypothèse de cette structure, plutôt qu'une organisation *scale-free*. Dans un réseau ayant une organisation *scale-free*, certains noeuds plus importants et répartis uniformément dans le réseau sont impliqués dans de nombreuses connexions alors que les autres noeuds sont impliqués dans peu de connexions. Toutefois, à travers différentes expériences sur les réseaux métaboliques, on a plutôt remarqué qu'un noyau important de grappes de noeuds est hautement connectés et qu'il est relié à d'autres groupes moins intensément connectés. Cette observation supporte l'hypothèse de l'architecture *bow tie*.

La structure *bow tie* accommode les flots d'informations et de contrôle, tout comme les flux de matières et d'énergie. De nouveaux processus sont facilement ajoutés au noyau de l'architecture. Ces processus n'ont qu'à utiliser les processus versatiles. Le système de transcription et de traduction des gènes en protéines est un exemple de cette architecture. Quelques mécanismes (et toujours les mêmes) sont requis pour exprimer tous les gènes. Ainsi, si des mutations introduisent de nouveau gènes dans un génome, il n'est pas nécessaire pour le système de développer un nouveau méca-

nisme de transcription et de traduction. Ceci a pour résultat de donner à la cellule la possibilité d'exprimer une vaste variété de gènes en macromolécules très différentes. Le métabolisme est aussi est très bon exemple d'architecture en noeud papillon. Une multitude de sources nutritionnelles sont catabolisées en une douzaine de métabolites précurseurs qui sont ensuite synthétisés en environs 70 molécules de constructions (les acides aminés, les nucléotides, les acides gras et les sucres). Ces dernières serviront ensuite à la constitution de toutes les macromolécules de biosynthèse de l'organisme.

Cette architecture serait universelle et il en existe aussi des exemples en technologie (Csete et Doyle, 2004). Citons tout d'abord celui d'un réseau électrique. À une extrémité, ce réseau tire son énergie de plusieurs sources distinctes pouvant avoir des types de production différents. Cette énergie est livrée dans un réseau de transport dans un format unique. À l'autre extrémité, l'utilisateur s'en sert pour des usages variés qui évoluent rapidement. Internet est aussi un exemple évident de cette architecture. Entre différentes applications logiciels se trouve un système de matériel électronique et des couches de réseautique qui transmettent des paquets de données selon un protocole unique. Ces deux exemples décrivent des systèmes où un large éventail d'inputs différents est transformé par un noyau de processus et de mécanismes pour donner un tout aussi large éventail d'outputs.

2.3 Méthodes d'analyse en biologie systémique

Les échanges entre les théoriciens du contrôle et les biologistes ont eu comme premier résultat de donner naissance à différentes méthodes d'analyse des systèmes biologiques, qui sont en fait des applications de techniques existantes de la théorie du contrôle. Ces études concernent différents aspects de la biologie systémique, dont la robustesse. Par exemple, il s'agit de définir des métriques pour mesurer quantitative-

ment cette propriété. Ou bien d'étudier la sensibilité d'un système à la modification de ses variables et d'explorer l'espace de ses paramètres. D'autres études ont pour objectif la caractérisation des mécanismes de rétroaction, par exemple pour identifier les molécules responsables de certains comportements complexes ou pour identifier le rôle d'une boucle de rétroaction dans un système. Cette sous-section présente ces différentes méthodes qui utilisent les concepts de robustesse, de rétroaction et de modularité, les systèmes biologiques qui ont été analysés par celles-ci ainsi que certains résultats obtenus.

Un deuxième résultat des échanges entre théoriciens du contrôle et biologistes est la formulation de nouveaux problèmes théoriques sur le contrôle et les systèmes et qui proviennent de la biologie systémique. Nous discuterons de deux études qui s'inscrivent dans cette problématique. Finalement, nous présenterons aussi deux études expérimentales dont les objectifs sont inspirés de la biologie systémique.

2.3.1 Études de la robustesse

La robustesse des systèmes intéresse de plus en plus de chercheurs de la biologie systémique (Stelling et al., 2004b). Des problématiques de recherche en biotechnologie et en pharmaceutique ciblent directement cette propriété. Il y a donc un besoin d'outils théoriques. Des métriques de robustesse comptent parmi ces outils. Wilhelm et al. (2004) ont développé une métrique qui permet de quantifier la robustesse structurale et de comparer le caractère robuste de différents systèmes biologiques entre eux. Pour se faire, les chercheurs proposent de calculer le nombre de chaînes réactionnelles qui sont coupées par le retrait d'une enzyme par rapport au nombre total de chaînes, incluant dans leur métrique une pondération selon l'importance du produit en bout de chaîne. Ils ont appliqué leur métrique au métabolisme d'*Escherichia coli* et à celui de l'érythrocyte de l'homme pour comparer la robustesse des deux systèmes. Les résul-

tats obtenus correspondent à une réalité bien connue, le métabolisme d'*Escherichia coli* doit être en mesure de s'adapter à des conditions variées et il est donc plus robuste, alors que la cellule érythrocyte qui évolue dans des conditions beaucoup plus homéostatiques n'est pas autant robuste.

Morohashi et al. (2002) ont proposé d'utiliser la robustesse comme une mesure de la plausibilité d'un modèle. Leur hypothèse est qu'un modèle qui reflète adéquatement la réalité devrait être aussi robuste à des modifications de paramètres que le système réel. Leur méthode consiste à établir une mesure quantitative du comportement d'un système et d'étudier par des graphiques à deux et trois dimensions l'évolution de ce comportement selon des variations de paramètres. Selon eux, un système robuste conserve un comportement normal sur une plage étendue de valeurs de ses paramètres. Ils ont appliqué cette méthode à deux modèles des oscillations biochimiques du cycle cellulaire du *Xenopus*. Le premier modèle, datant de 1991, est plus simple que le modèle de 1998 et l'étude a confirmé que ce modèle avait certaines déficiences de robustesse puisqu'il réagissait mal à l'introduction de bruit dans la valeur des paramètres. Le modèle de 1998 semble plus robuste, donc plus plausible. Cette méthode a d'ailleurs permis aux auteurs d'optimiser les paramètres du modèle existant. Les auteurs sont d'avis que ce type d'étude sur la robustesse devrait faire partie de tous processus de modélisation puisqu'il offre des indications éclairantes et des indices sur les faiblesses d'un modèle pour les modeleurs lorsqu'ils sont confrontés à des choix ou lorsqu'ils doivent procéder à l'évaluation de leur modèle.

De façon générale, des analyses de sensibilité sont utilisées pour mesurer la robustesse d'un système, mais elles ont peu été appliquées aux systèmes biologiques dynamiques en raison de leur caractère complexe et non-linéaire. Deux équipes de recherche ont proposé des avenues différentes pour pallier aux difficultés d'application des analyses traditionnelles. Dans un premier temps, Stelling et al. (2004a) ont investigué sys-

tématiquement l'espace des paramètres de deux modèles des oscillateurs génétiques responsables du rythme circadien de *Drosophila* via trois types d'analyse : sensibilité d'état, sensibilité de période et sensibilité d'amplitude, chacune ayant leur propre métrique mathématique. Ils ont ainsi observé dans les systèmes plusieurs aspects de la robustesse comme l'influence des processus de régulation sur le système, les hiérarchies de contrôle et les impacts des perturbations. Dans un second temps, Chen et al. (2005) ont émis l'opinion que des analyses de sensibilité ne sont pas suffisantes en biologie puisqu'elles font abstraction de la nécessité de maintien de l'état d'équilibre du système. Ils ont utilisé la représentation par synergie et saturation de système (*S-system*) pour développer une métrique de la robustesse des réseaux biochimiques lorsqu'ils sont soumis à des variations de leurs paramètres cinétiques. Cette métrique leur a permis de calculer la limite supérieure des valeurs tolérées pour la variation d'un paramètre à l'état d'équilibre du système et d'utiliser cette limite pour calculer la robustesse d'un système. Ils ont démontré l'utilité de leur métrique en l'appliquant à quatre réseaux biochimiques : un réseau modèle de la cascade biochimique, la voie glycolitique-glycogénolytique, le cycle de l'acide tricarboxylique et le réseau de l'oscillation du cAMP.

2.3.2 Études de la rétroaction

Les études sur la rétroaction m'ont semblé les plus près de la pensée de la biologie systémique puisqu'elles offrent des pistes concrètes du fonctionnement interne des systèmes biologiques complexes. Dans une première étude, Khammash et El-Samad (2004) font une nouvelle interprétation qualitative de deux systèmes : un système physiologique, la régulation homéostatique du calcium du plasma chez les mammifères ; et un système de régulation génétique, le réseau de réponse au choc thermique. Dans le cas du système physiologique, les chercheurs ont analysé la courbe de retour à

la normale de la concentration du calcium après une perturbation. La forme de cette courbe et l'absence d'erreur dans la concentration finale suppose la présence d'un mécanisme de contrôle proportionnel-intégral, ce qui impose l'action de deux agents régulateurs (un premier proportionnel, et un deuxième intégral). L'étude de l'endocrinologie de ce système a permis d'identifier deux hormones, déjà reconnues comme des acteurs dans ce système, dont le comportement dynamique correspond à celui des agents régulateurs recherchés. Dans le cas du système de régulation génétique, l'analyse des chercheurs, encore une fois qualitative, décortique le comportement dynamique d'un système complexe qui doit être modélisé à l'aide de 31 équations différentielles ou algébriques. À première vue, une telle complexité semble superflue pour la fonction à accomplir par ce réseau génétique lors d'un choc thermique, c'est-à-dire la production de la molécule σ^{32} lors d'une augmentation de la température pour réparer les protéines dénaturées. Toutefois, en adoptant une vision systémique, les chercheurs justifient la présence de tous les acteurs de régulation, qu'ils soient impliqués dans des boucles de rétroaction positive ou négative. Ceux-ci appuient leur raisonnement de différentes données expérimentales comme la fonction connue de certaines protéines, le comportement dynamique des macromolécules à différentes températures, les mécanismes d'expression génétique et le temps d'action de la molécule σ^{32}. Dans une deuxième étude, El-Samad et al. (2005) poursuivent leur analyse de ce système de régulation génétique, mais de façon quantitative, pour discuter du rôle de chacune des boucles de rétroaction. Pour ce faire, ils utilisent le modèle mathématique de ce système composé de 31 équations. Leur démarche consiste à ajouter dans le modèle les modules de rétroaction un à un et de comparer les réponses dynamiques du système par simulation. Leur modèle de base, un système en boucle ouverte, se voit successivement ajouter des boucles de rétroaction négative et positive, ce qui permet aux auteurs de discuter du rôle de chacun de ces mécanismes dans le système.

Schmidt et Jacobsen (2004a,b) ont développé une méthode mathématique qui identifie les mécanismes de rétroaction dans les systèmes caractérisés par un comportement multistable et par des oscillations périodiques soutenues. Ils proposent une linéarisation du modèle mathématique du système autour de certains points significatifs afin d'être en mesure d'utiliser les outils d'analyse de la rétroaction linéaire provenant de la théorie du contrôle. Ces points précis, appelés racines en théorie du contrôle, correspondent à des états de transition entre les différents états stables du système. Il est possible de les déterminer par une analyse de bifurcation. Par la suite, en décomposant le système en sous-modules et en appliquant à ceux-ci différents théorèmes et transformations, on parvient à identifier les molécules ainsi que les interactions qui jouent un rôle significatif dans le comportement oscillatoire du système (en d'autres mots, les acteurs de la rétroaction qui font transiter périodiquement le système à travers différents états). Cette approche est purement qualitative puisqu'elle ne permet pas d'obtenir des résultats quantitatifs à propos de la robustesse du système ou de la sensibilité de ses paramètres. Néanmoins, cette méthode, utilisée tôt dans un processus de modélisation, sert à cibler les composantes les plus importantes dans un modèle et elle fournit des indications quant à leur fonctionnement dans un système complexe. À travers leurs deux publications, Schmidt et Jacobsen ont appliqué leur méthode à trois systèmes différents : le cycle cellulaire du *Xenopus*, le cycle oscillatoire de la glycolyse chez la levure et le cycle circadien de *Drosophila*. Dans chaque cas, les molécules et les interactions qui sont instrumentales pour générer un comportement non-linéaire complexe ont été identifiées.

2.3.3 Études de la modularité

La modularité est à la base de la méthode proposée par Saez-Rodriguez et al. (2004). Ils suggèrent qu'il est utile de décomposer les voies de signalisation en modules pour

démêler la complexité biologique. Leur méthode prône la reconnaissance de modules en ayant comme critère de délimitation l'absence de rétroaction à l'intérieur de ceux-ci. Les chercheurs justifient le choix de ce critère par le nombre d'outils qui existent pour analyser de tels modules sans rétroaction. Dans le cas de cette recherche, des réseaux de signalisation sont analysés. Suite à la formation des modules, il devient possible de calculer pour ceux-ci l'amplitude d'un signal d'output, le temps de signalisation et la durée du signal en fonction d'un signal d'entrée. Cette méthode a été appliquée à deux systèmes simples, la signalisation à deux composantes et la cascade MAPK, et à un système plus complexe, le réseau de signalisation du facteur de croissance épidermique. Leur méthode permet de caractériser le fonctionnement de signalisation des modules. L'assignation à la dernière sous-unité du réseau de signalisation du facteur de croissance épidermique de l'entière responsabilité d'une insensibilité à des stimuli, que l'on croyait auparavant distribuée dans tout ce réseau, est un exemple de propriété déterminée par cette méthode d'analyse par modularité.

2.3.4 Études théoriques

Pour les spécialistes de la théorie du contrôle, Sontag (2004) a identifié quatre catégories de possibilités de recherche en biologie systémique. Les deux premières catégories se réfèrent aux applications de la théorie du contrôle au domaine biologique : 1) le rôle du contrôle et des techniques de traitement de signaux dans la conception d'instrumentation de mesure biologique à haute précision ; 2) l'utilisation de techniques existantes de la théorie du contrôle (identification, quantification de gain, analyse de sensibilité, contrôle optimal, etc.) dans l'analyse et la solution de problèmes qui intéressent les biologistes. Les deux autres catégories de possibilités de recherche résultent de la pression de la biologie systémique sur la théorie du contrôle pour se développer davantage : 3) l'abstraction de nouvelles idées pour l'ingénierie du contrôle à partir

de la recherche biologique ; 4) la formulation de nouveaux problèmes théoriques pour la théorie du contrôle et des systèmes, provenant de la recherche en biologie systémique. C'est sur ce dernier point que Sontag poursuit son article en affirmant que la cellule est un système d'entrée/sortie semblable à d'autres qui sont étudiés en théorie du contrôle, mais que les systèmes biologiques posent des problèmes qui diffèrent fondamentalement des problèmes classiques. Il effectue même une mise en garde : trop souvent et à tort, on croit que les problèmes posés par la biologie systémique sont des problèmes standards de la théorie du contrôle. À son avis, de nombreuses nouvelles questions théoriques émergent de ce constat et il tire de ses recherches différents exemples. Parmi ceux-ci, le mélange de modélisation qualitative et quantitative, les systèmes monotones, les réponses à l'état stable, la détection du bruit, les systèmes d'identification et de réingénierie, le contrôle adaptatif et la stabilité robuste des structures.

L'article d'Angeli et Sontag (2003) est un bon exemple des recherches théoriques rendues nécessaires par la biologie systémique. Dans cet article, les auteurs se sont attardés à faire la démonstration mathématique rigoureuse que la notion de systèmes monotones peut être étendue aux systèmes avec entrées et sorties. Ceci est le premier pas permettant la compréhension des interconnections qui sont formées entre des modules monotones, notamment celles qui incluent des boucles de rétroaction. Ceci ouvre la porte à l'application de la théorie des systèmes monotones et de ses outils aux modules biologiques et à la démonstration de certaines de leurs propriétés.

2.3.5 Études expérimentales

Le développement de nouvelles approches expérimentales pour la biologie systémique pose certains problèmes, malgré les nombreuses nouvelles questions formulées par les théoriciens. Les scientifiques doivent faire appel à plusieurs technologies de collecte

d'informations qui sont coûteuses et qui sont habituellement cloisonnées dans différents domaines de la biologie. Les problèmes sont complexes et de taille importante. Ideker et al. (2001b) ont été parmi les premiers à proposer des analyses génomiques et protéomiques intégrées pour développer itérativement un modèle d'une voie cellulaire. Aderem et Smith (2004) ont approfondi cette idée dans leurs recherches sur le système immunitaire et les réactions inflammatoires en y incorporant un environnement informatique qui analyse les données génomiques et protéomiques et qui facilitent la détermination d'interactions complexes. Cet environnement informatique est tout d'abord composé du logiciel et base de données SBEAMS, qui régit la collecte, l'entreposage et l'accès aux données produites par différents types d'expérience. Ils utilisent ensuite le logiciel Cytoscape qui intègre les données, et en permet la visualisation et l'exploration. Certaines fonctionnalités de Cytoscape sont de regrouper en modules les réseaux moléculaires, de faire des analyses de génomiques comparatives et de dériver des réseaux de régulation génétique.

Certaines problématiques de recherche à plus petite échelle sont tout de même influencées par la philosophie de la biologie systémique. Mentionnons, en exemple, l'étude de Batchelor et al. (2004) dans laquelle ils ont caractérisé des circuits régulateurs d'*Escherichia coli* et déterminé les modes de contrôle de ceux-ci. Leur analyse pointe vers un système de contrôle continu de la transcription d'un gène dans un cas (alors qu'on le supposait discret jusqu'alors) et d'une boucle de rétroaction ouverte (ce qui est surprenant pour cette espèce).

2.4 Sommaire

Peu de disciplines scientifiques ont la chance de revoir leur cadre conceptuel comme c'est le cas actuellement en biologie à cause de la biologie systémique. Ceci est une op-

portunité intellectuelle très stimulante qui suscite et suscitera encore bien des débats. Les scientifiques se dotent de nouveaux concepts et de nouveaux outils pour repousser les limites de leurs connaissances du vivant. Et le rapprochement qui se produit entre la biologie et le génie est porteur de nombreuses collaborations prometteuses.

Cette section a offert un survol théorique de ce pont qui se construit actuellement entre les sciences de la vie et les sciences appliquées. À plusieurs égards, les systèmes biologiques et les systèmes technologiques complexes ont plusieurs points en commun, dont leurs principes d'organisation. Certains concepts de ce lieu commun entre biologie et génie, qui sont actuellement proposés dans la littérature, ont été présentés : l'émergence, la robustesse, la fragilité, la modularité, la rétroaction ainsi qu'une première proposition de structure architecturale, le noeud papillon. Dans la seconde sous-section, des études réalisées par différents chercheurs et qui exploitent les différents concepts de la biologie systémique ont été recensées.

Depuis environs une décennie, la biologie est en ébullition, frappant l'imaginaire populaire par de grandes réussites et occasionnant plusieurs questionnements techniques, scientifiques et éthiques dans le milieu scientifique. Les ingénieurs sont maintenant invités à partager avec leurs collègues biologistes leurs connaissances des systèmes complexes et à participer aux nouvelles avancées de la biologie. Aurons-nous le génie de relever le défi ?

CHAPITRE 3

DÉMARCHE DE L'ENSEMBLE DU TRAVAIL DE RECHERCHE ET ORGANISATION GÉNÉRALE DU DOCUMENT

L'étude de Bhalla et Iyengar (1999), portant sur les propriétés émergentes d'un réseau de voies de signalisation du neurone, est une contribution scientifique majeure pour deux raisons. Premièrement, il s'agit d'une étude de biologie systémique avant l'heure puisqu'on y étudie les propriétés d'un système complexe à l'aide d'un modèle computationnel. Deuxièmement, ces propriétés peuvent être reliées à la plasticité synaptique, une importante propriété neuronale qui serait à la base des processus de mémorisation et d'apprentissage du cerveau et dont la mécanique biochimique était encore obscure à l'époque.

Nous avons utilisé ce modèle dans notre étude portant sur la modélisation biologique à multiples niveaux d'abstraction basée sur le formalisme des réseaux de Pétri (Hardy et Robillard, 2005). La réalisation de cette étude nous a donné une compréhension approfondie du comportement dynamique de ce modèle. Toutefois, il nous a semblé que les méthodes traditionnelles de représentation et d'analyse des données de simulation ne permettaient pas de dégager facilement ce comportement et ne facilitaient donc pas l'atteinte d'une compréhension systémique. Ce sont ces lacunes auxquelles les trois articles constituant le corps de cette thèse tentent de pallier. Le modèle du réseau de régulation de l'enzyme CaMKII développé par Bhalla et Iyengar (1999) et augmenté par Kikuchi et al. (2003) a été utilisé dans les trois articles de cette thèse parce qu'il rassemble tous les attributs d'un excellent cas d'étude pour notre recherche. Il s'agit d'un modèle complexe d'un système très important en neurobiochimie. C'est aussi un réseau de signalisation constitué de plusieurs mécanismes de

régulation qui a un comportement dynamique très intéressant à analyser.

Le chapitre 4 introduit le premier article. Intitulé "A functional representation of the simulation data of biochemical models based on molecular activity", et écrit par Simon Hardy et Pierre N. Robillard, il a été soumis pour publication à la revue *In Silico Biology*. Cet article présente une approche d'analyse de données de simulation où l'activité moléculaire de modules biochimiques sert à créer une représentation fonctionnelle du modèle. Dans toute démarche d'analyse de données d'un système complexe, l'analyse se fait en partie grâce à une conceptualisation mentale. Cette conceptualisation tient compte de la structure du réseau, de la fonction des composantes et de leurs interrelations. Le scientifique qui effectue une analyse se base sur cette conceptualisation pour déchiffrer ses données et pour leur donner un sens générant de la connaissance. L'approche présentée dans cet article propose d'utiliser cette conceptualisation mentale pour créer des filtres d'analyse pour la visualisation des données. Ces filtres permettent de produire, à partir des données de simulation, une représentation fonctionnelle du système.

Le chapitre 5 introduit le second article. Intitulé "Petri net-based visualization of signal transduction pathway simulations", et écrit par Simon Hardy et Pierre N. Robillard, il a été soumis pour publication à la revue *IET Systems Biology*. Cet article présente une méthode de visualisation des données de simulation des voies de signalisation cellulaire. À l'aide d'une technique d'analyse des réseaux de Pétri (l'identification de P-invariants), la méthode de visualisation prend en compte le fonctionnement enzymatique pour générer une vue du système plus fidèle à son comportement systémique. Les méthodes de visualisation existantes, appropriées pour les voies métaboliques, donnent des résultats mitigés avec les voies de signalisation. Les problèmes détectés sont solutionnés par notre méthode.

Le chapitre 6 introduit le troisième article. Intitulé "Petri net-based method for the

analysis of the dynamics of signal propagation in signaling pathways", et écrit par Simon Hardy et Pierre N. Robillard, il a été soumis pour publication à la revue *Bioinformatics*. Cet article présente une méthode d'analyse des modèles de voies de signalisation qui permet d'étudier leur dynamique de propagation de signal. À l'aide de techniques d'analyse des réseaux de Pétri (l'identification de P-invariants et de T-invariants), la méthode permet d'identifier les éléments du modèle qui jouent un rôle important dans la transmission de signaux cellulaires, de décrire la dynamique de transduction et d'identifier la présence de motifs de régulation. Cette méthode transforme les données de variation de concentration moléculaire en données de transmission de signaux cellulaire, ce qui permet un accès rapide aux mécanismes systémiques de traitement d'information de la cellule.

Les trois articles atteignent indépendamment l'objectif de recherche de cette thèse, qui est de concevoir des méthodes d'analyse de données de simulation de modèles de voies de signalisation cellulaire qui permettent de dégager leur comportement systémique. Les deux premières méthodes accordent une place de choix aux techniques de visualisation. Mais ce sont les méthodes des deux derniers articles, puisqu'elles sont basées sur les réseaux de Pétri, qui sont complémentaires et qui peuvent parfaitement s'intégrer dans une approche unique.

Par ailleurs, deux articles de conférences auxquels l'auteur a principalement contribué ont été inclus en annexes à cette thèse. L'annexe I est constituée de l'article intitulé "Painted Petri net and functional abstraction to visualize dynamic modeling" où les réseaux de Pétri peinturés sont décrits pour la première fois. L'annexe II est constituée de l'article intitulé "Visualization of the simulation data of biochemical network models : a painted Petri net approach", qui est une version préliminaire du chapitre 5 de cette présente thèse.

CHAPITRE 4

FUNCTIONAL ABSTRACTION AND SPECTRAL REPRESENTATION TO VISUALIZE THE SYSTEM DYNAMICS AND THE INFORMATION FLUX IN A BIOCHEMICAL MODEL

Abstract

Interpreting the simulation data[1] of a complex biochemical model to understand its dynamic behavior is a difficult task. Traditional data representations display simulation outputs as concentration plots. To study the dynamic behavior of a model from these plots, it is necessary to have in mind the topology of the modeled system, know the function of the individual elements of the system and be able to describe their activity. Only with this mental image of the model can the dynamic behavior be deciphered. In this paper, we suggest exploiting this knowledge to create a preprocessing filter for the simulation data. This data filter is based on the concept of molecular activity and transforms the simulation data from a concentration perspective to a molecular activity perspective. This is done in two steps: identify the functional groups of the system, and mathematically describe the molecular activity of these groups. In this paper, we demonstrate this new data representation approach with a complex model of the signal transduction system of long-term potentiation in the hippocampal post-synapse, a model exhibiting a bistable behavior. To facilitate viewing of the resulting data matrix, the preprocessed data are displayed with known visualization techniques, followed by the production of an animated and a spectral

[1]Cet article a été soumis à la revue *In Silico Biology*. Voir la référence (Hardy et Robillard, 2006a)

functional representation. One advantage of the functional data filter is that, once created, it can be applied to a large number of simulation runs while at the same time performing parametric and structural modifications on the model in order to quickly explore the impacts on the model's behavior.

Keywords: biochemical modeling, simulation data, system dynamics, function, visualization.

4.1 Introduction

The simulation of biochemical models based on kinetic reactions generates data time series of concentration. To interpret these time series, the data are usually displayed on Cartesian graphs where time is the x-axis and concentration the y-axis. These plotted data are used to study the dynamic behavior of simulated models by combining sets of graphs. This simple type of representation is convenient for the study of simple models, but using it to study the dynamic behavior of complex models is a difficult task. Biochemical systems are highly organized and composed of many heterogeneous processes, formed by the interaction of the activities of different molecules. The signaling pathways of the cell are typical cases of such processes, where signals are transmitted by the perturbation of enzymatic activities and the actions of molecular messengers. An example of enzymatic activity is phosphorylation: some proteins are turned "on" and "off" by the addition and removal of a phosphate group, catalyzed by a kinase. Thus, it is possible to link a structural modification of a molecule to its role in a biochemical process.

Researchers develop models of signaling pathway networks on the basis of experimentally established relationships between molecules, their function and the property of a biological system under study. Analyses of the simulation data of these models re-

quire a deep understanding of the underlying molecular or biochemical processes. To make sense of the simulation data, they are processed and filtered according to the researcher's understanding of the activities of the system components. Our analysis will focus mostly on some relevant dynamic behaviors of model components.

The original approach presented in this paper enables researchers to explore a new functional interpretation of a signaling pathway based on their existing knowledge of molecular activities by using simulation data. This functional interpretation transforms the data from a molecular concentration perspective to a molecular activity perspective, a transformation which is achieved in two steps. In the first step, the molecular substances of the model are grouped into functional units, a functional unit being defined as a group of molecules participating in a given activity performed by the signaling network. In the second step, the activity of each functional unit is mathematically formalized by an equation. This equation defines an indicator of the degree of molecular activity as a function of the concentration of the molecules of a functional unit. The parameters of the equation are empirically defined by the researchers based on their understanding of the molecular processes at the functional unit level.

This new functional interpretation of the simulation data provides a unique dynamic perspective when combined with powerful visualization techniques. Colors are used to represent the degree of molecular activity of the functional units of a model. The mapping of these colors on a graph displaying the topology of the model creates a highly informative animated graphical representation of the dynamic behavior of the model.

This paper presents the use of this functional interpretation approach combined with visualization techniques based on the simulation data of the model of the signal transduction system of long-term potentiation (LTP) in the hippocampal post-synapse.

Bhalla et Iyengar (1999) were the first to model this network. Their model was later modified by Kikuchi et al. (2003). This system is interesting because of its property of bistability, which, it is hypothesized, is linked to the neuronal memory processes. Representation of the data from a functional perspective clearly illustrates this systemic property, as well as the dynamic behavior of the system and its molecular components.

4.2 Previous work

In many different scientific fields, making sense of large quantities of complex data can be perilous without the help of data processing and visualization techniques. This is the case for genome, transcriptome and microarray data, protein interaction maps and metabolic pathways – molecular biology is no exception. Biologists use data analysis and visualization techniques to both explore their data and present their results. In computational biology, the simulation of complex biochemical models can also generate a substantial amount of data. The interpretation of this data can be a slow and inefficient process using the traditional means of data representation: two-dimensional plots of concentration-time series data. Another problem with the presentation of the simulation data of biochemical models with raw numbers and plots is the loss of the network topology, which is information that is crucial to building a mental representation of the dynamic behavior of a model.

Prior to the creation of the molecular activity perspective developed in this paper, software tools had been developed to try to solve these problems. MetVis (Qeli et al., 2003), SimWiz (Rost et Kummer, 2004) and BioPathwize[2] make it possible to visualize the simulation data of metabolic networks and signaling pathways in a

[2]http://bioanalyticsgroup.com

more comprehensive way. These tools show a manually or automatically generated network graph to which the simulation data are mapped. This creates an animated representation of the data. Either the edges or the nodes of the network graph change in color or volume to show the evolution of the concentration of the substances making up the biochemical network. The tools MetVis and SimWiz are also available in a three-dimensional version (Qeli et al., 2004a; Wegner, 2005).

The advantages of the animated views produced by these tools are numerous. They take advantage of the benefits of visual data exploration, one of which is to integrate the human perceptual abilities into data interpretation, a useful asset in the exploratory steps of data analysis (Keim, 2002). They also make full use of the capacity of the human mind to detect structures and patterns in images, such as synchronicities, global changes of state and oscillations (Müller et Schumann, 2003). The developers of MetVis and SimWiz stress the usefulness of the topology of the biochemical networks for their animated graphical representations. To map the simulation data to the model structure gives a good impression of the dynamic behavior of the system. Thus, a wealth of information becomes accessible to the user.

Despite these advantages, the animated representations generated by the existing visualization software tools of the biochemical network simulation data show only the variation in concentration of the network substances. In short, they provide a single point of view without any data processing, which is the concentration perspective. The same is true for concentration plots. These tools fail, however, to fully benefit from another data analysis method: the visualization of data with multiple views (Roberts, 1998). By displaying data in multiple ways, the user may interpret the data through different perspectives, hasten its understanding and avoid possible misinterpretations. Visualization systems usually follow a four-step pipeline dataflow model: the data are filtered to create a subset of data, which is then mapped to a

representation which can be displayed. To produce multiple views from the same data set, three modifications of the visualization dataflow are possible: 1) modification of the data filter; 2) modification of the data mapping; and 3) modification of the display. A modification of the data mapping or the display, as proposed by the tools MetVis and SimWiz with the transformation of concentration plots into a topological view, preserves the same data filter: concentration variation. The functional perspective presented in this paper is based on a modification of the data filter, which is the first step in the pipeline dataflow model.

4.3 The long-term potentiation signal transduction model

Much research has been conducted on the synaptic plasticity of the neuron. This property enables the neuron to undergo a lasting alteration to the efficiency of its neurotransmission signaling process. When there is an enduring increase in the amplitude of post-synaptic exciter potentials, synaptic plasticity is called long-term potentiation (LTP). LTP was first described by Bliss et Lomo (1973). Experimentally, it is caused by series of short, high-frequency electrical stimulations, also described as tetanic stimulations, to a nerve cell synapse which strengthen, or potentiate, that synapse for minutes or hours. Research on LTP (and its opposite, long-term depression or LTD) is in part motivated by the assumption that synaptic plasticity forms the cellular basis of learning and memory. (Blitzer, 2005) is an introductory paper on LTP, and (Ajay et Bhalla, 2006) is a review of the theoretical and experimental research on synaptic plasticity. Many kinetic simulations of LTP at the molecular level, as well as models based on phenomenological facts, have been developed to theoretically study this complex property. As reported in a previous paper, we transformed these two kinds of model into a single, Petri net-based representation (Hardy et Robillard, 2005).

Bhalla et Iyengar (1999) developed a complex LTP model at the molecular level composed of many simple networks combined to form post-synaptic signaling pathways. A remarkable aspect of their work is that they used a systemic approach to unravel long-term potentiation. They demonstrated that their model was characterized by the property of bistability, as a result of the combination of multiple networks, since none of them exhibited this behavior by itself. They based their work on the experimentally observed fact that bistability is an enduring biochemical modification linked to LTP. New experimental evidence has led Kikuchi et al. (2003) to add some reactions to this model, involving the dynamic inactivation of the protein phosphatase 2A (PP2A). This addition was significant because PP2A affects important molecules in synaptic plasticity, and thus has an impact on the bistability of the system. Adding the dynamic modeling of PP2A made the LTP model more robust and facilitated LTP induction in the system. The block diagram of the modified model of the signal transduction system of LTP is shown in Figure 4.1. In this figure, rounded rectangles represent enzymes, circles represent messenger molecules, rectangles represent receptors and dotted, rounded rectangles represent a reaction module. Regular arrows represent activation and circle-ended arrows represent inhibition. Each block in this diagram corresponds to a set of chemical reactions between several substances. Consequently, the blocks are simplifications of more complex networks.

In the remainder of this paper, the simple interconnected networks of the model are referred to as *modules*. The modules of the LTP model are represented in two different ways in Figure 4.1: either a module is composed of a number of blocks (in which case, the module is represented by a dotted, rounded rectangle) or a module is a single block outside a dotted rounded rectangle (like PKA, CaM and CaMKII). One module is a set of chemical reactions involving approximately five to fifteen substances, and each reaction has constants, all of which have been experimentally obtained and documented. Those parameters are compiled in the DOQCS database

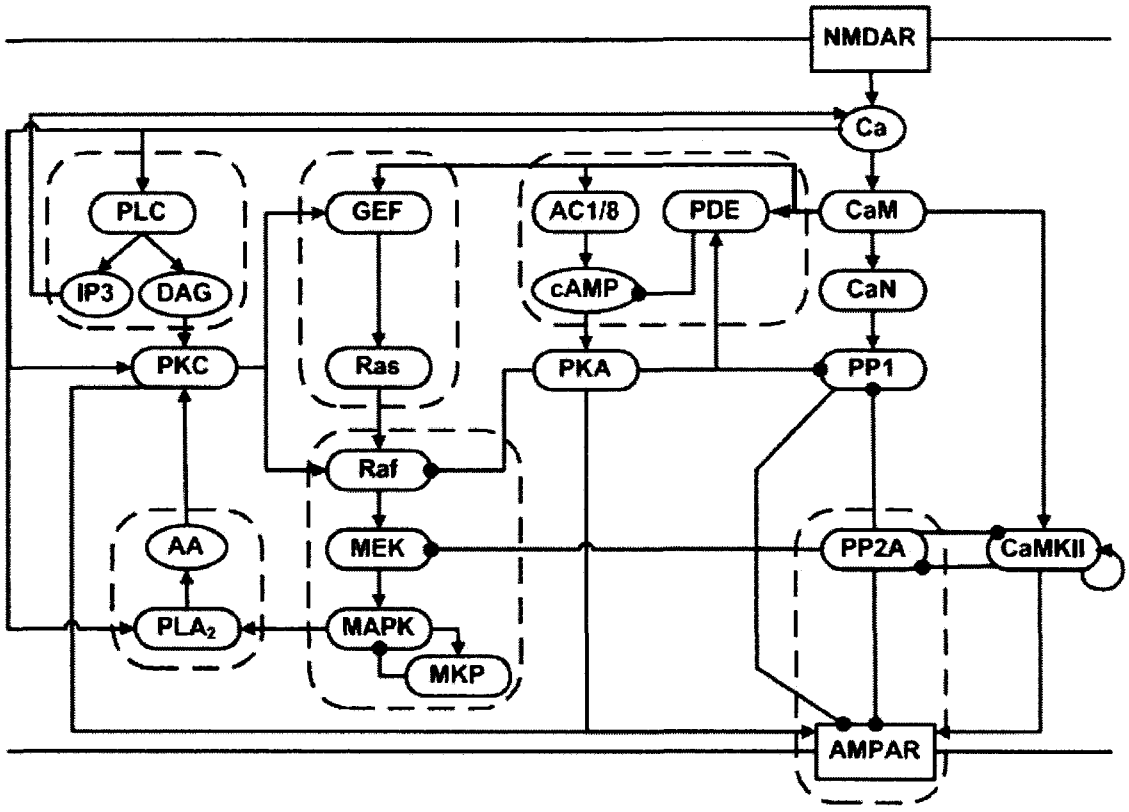


Figure 4.1: Block diagram of the long-term potentiation signal transduction system in the hippocampal post-synapse

(Sivakumaran et al., 2003). The model contains approximately 200 substances and 350 reactions. The reader is referred to Kikuchi et al. (2003) for more information about the model.

The model was simulated with the software E-CELL (Tomita et al., 1999). Model bistability from a normal steady state to a potentiated steady state was observed by inducing a tetanic inflow of calcium ions to the model. The entry point of the model is calcium, and the increase in its concentration activates several enzymes, which in turn activate various molecules. The activity of the model's interconnected positive and negative feedback loops, combined with the appropriate stimulus, leads to a potentiated steady state.

4.4 Creation of a molecular activity perspective of the simulation data of the LTP model

The first step in the transformation of the simulation data from a concentration perspective to a molecular activity perspective is the identification of the functional units of the LTP model. Grouping the substances of the LTP model into functional units was performed by interpreting the reaction equations of the modules and the topology of the model, with the experience gained with the model and according to engineering reasoning. A key factor in the identification of the functional units of the LTP model is to group together the different configurations of a molecular species (for example, the calmodulin (CaM) functional unit is composed of the following different molecular configurations of the CaM molecules: CaM, CaM.Ca_2, CaM.Ca_3, CaM.Ca_4, CaM.Ng). However, the functional units are specific to the biological process under study and cannot be derived from formal analysis of the model. This is discussed in more detail later in this paper.

Figure 4.2 shows examples of functional units. On the left of this figure is the reaction scheme of the PKA module. On the right is the same reaction scheme represented in a hybrid functional Petri net notation, where double circles are substances, rectangles are reactions and arrows indicate the direction of the reaction (see (Hardy et Robillard, 2004) for more information about Petri nets and molecular biology). The PKA module of the LTP model is divided into the following three functional units: 1) the R_2C_2 complex unit, where the two R complex subunits first bond successively to four cAMP molecules, after which the complex releases two PKA enzymes; 2) the PKA inhibitor unit, in which inhibitor molecules are either in a free or a coupled configuration with inhibited PKA enzymes; and 3) the PKA enzyme unit. These units represent the three main activities performed by the molecules of the PKA module, which are, respectively, the entry point of the module, where a molecular

complex needs input molecules to produce an important enzyme; the inner control mechanism; and the output molecule that will interact with other modules. One of the advantages of forming functional units is to reduce the number of numerical values of interest. In the PKA module, ten concentration values are expressed with the activity level value of three functional units. For the entire LTP model, 200 concentration values are expressed with an activity level value of 34 functional units.

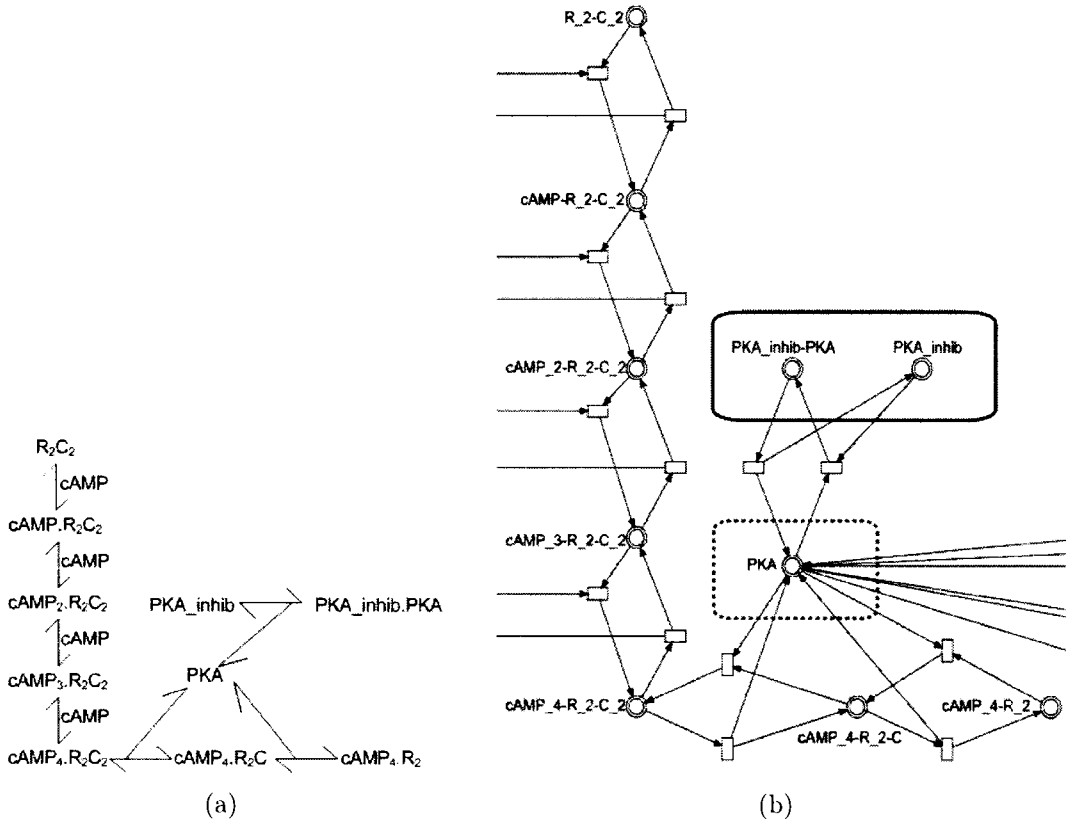


Figure 4.2: Reaction equations and functional units of the protein kinase A module of the LTP model

The second step in the transformation of the simulation data is to mathematically describe the activity of each functional unit by an equation. The continuous concentrations of the substances of the functional unit are converted into discrete states

indicating the level of activity of the functional unit. This conversion is performed with equations using concentrations as inputs to compute a value on a discrete scale from 0 to 10 (0 being the lowest level of activity for a functional unit, 10 being the highest). For the LTP model, each functional unit has a state equation designed to produce an activity level value that is consistent with the unit behavior. Equations 4.1, 4.2 and 4.3 are the equations of the R_2C_2 complex unit, the PKA inhibitor unit and the PKA enzyme unit of the PKA module. The equations are used to determine the activity level of the three functional units at every simulation time step.

$$\lceil state_{R_2C_2} \rceil = \frac{10}{50}\left(\frac{0}{6}[R_2C_2] + \frac{1}{6}[cAMP.R_2C_2] + \frac{2}{6}[cAMP_2.R_2C_2] + \frac{3}{6}[cAMP_3.R_2C_2] + \frac{4}{6}[cAMP_4.R_2C_2] + \frac{5}{6}[cAMP_4.R_2C] + \frac{6}{6}[cAMP_4.R_2]\right) \tag{4.1}$$

$$\lceil state_{PKA_{inhib}} \rceil = 20(0.5 - [PKA_{inhib}]) \tag{4.2}$$

$$\lceil state_{PKA} \rceil = 100[PKA] \tag{4.3}$$

For the purposes of this paper and to explain the functional representation, we provide a general description of the design of the first state equation. Equation 4.1 uses the concentration of the seven molecular configurations of the R_2C_2 complex. The function of this unit is to produce PKA enzymes when cAMP molecules are available in sufficient quantity. In its least active configuration (R_2C_2), the R_2C_2 complex is not bound to any cAMP molecule. In its most active configuration ($cAMP_4.R_2$), it is bound to four cAMP molecules and has released two PKA enzymes. The other five molecular configurations can be positioned relative to this activity scale. According to the distribution of the concentration of the R_2C_2 complex in the seven configurations, the equation gives a value from 0 to 10, associating a discrete value to the level of the activity of the functional unit. If all the complexes are in the R_2C_2

configuration, the level of activity is 0. If all the complexes are in the $cAMP_4.R_2$ configuration, the level of activity is 10. If the complexes are distributed among the seven configurations, the level of activity will be between 0 and 10. The first constant of equation 4.1, $\frac{10}{50}$, is a normalization constant, 50 being the total concentration for every R_2C_2 configuration, and 10 being the number of possible activity levels.

4.5 Visualization of the simulation data from a functional perspective

The 34 state equations of the LTP model each provide an integer from 0 to 10 for every time step of a simulation run. An easy way to visualize this data matrix is to associate the integer values with colors. For the LTP model, we used the colors of the light spectrum. Violet is associated with the state 0, low activity, and red is associated with the state 10, high activity. The color association process is complete when the color of every functional unit for every simulation step is determined. Subsequently, the colored functional data are displayed in two different types of visual representation. The first representation is an animated, global view of the model, where the colored functional data are mapped to the topology of the model. The result is that the variation in the activity of all the model components can be seen simultaneously. This view is made up of a succession of images of the colored functional units of the model for various time steps, in other words an animation. The second representation is a collection of spectra of the functional units, showing their changing level of activity through time. A spectrum is a timeline, where the color changes represent the activity variation for a single functional unit.

The next two subsections present the animated and spectral representations of two simulation runs of the LTP model. The first simulation run is based on normal concentrations and parameters. For the second simulation run, the value of the Kcat

parameter of the inactivation reaction of PP2A by CaMKII.CaM is modified. The normal value of the parameter is 5 s^{-1} and the modified value is 0.5 s^{-1}. These values were extracted from the work of Kikuchi et al. (2003). The modification is known to affect the model's bistability. In both simulation runs, a tetanic calcium stimulation is induced at 120 seconds. The concentration of the 200 substances of the model is recorded at every second. The simulation duration is 1,000 seconds. The simulation outputs, a set of the concentration data of the 200 substances for each simulation run, were transformed to create a colored functional perspective.

4.5.1 Animated functional representation of the simulation data of the LTP model

The animated functional representation of the simulation data of the LTP model is a mapping of the colored functional data to a graphical representation of the topology of the signaling network. The graphical representation of the topology is a graph where the functional units are displayed as nodes. Regular and rounded arrows link the functional units to show the activation and inhibition relationships. The nodes are colored according to their level of activity computed from the simulation data. A change in color reflects a change in activity level[3]. Four frames of the animation of the first simulation run are shown in Figure 4.3. Two frames of the animation of the second simulation with the modified parameter are shown in Figure 4.4. Subfigure 4.3(a) shows the system from the functional perspective of both simulation runs at 90 s (normal steady state), subfigure 4.3(b) shows the system at 124 s (a few seconds after the tetanic stimulation), subfigures 4.3(c) and 4.4(a) show the system at 200 s (in transition to the final steady state) and subfigures 4.3(d) and 4.4(b) show the

[3]The animated representations of the first and second simulation runs, SpectroNet 1 and SpectroNet2 respectively, can be downloaded at http://www.polymtl.ca/rgl/Downloads.php.

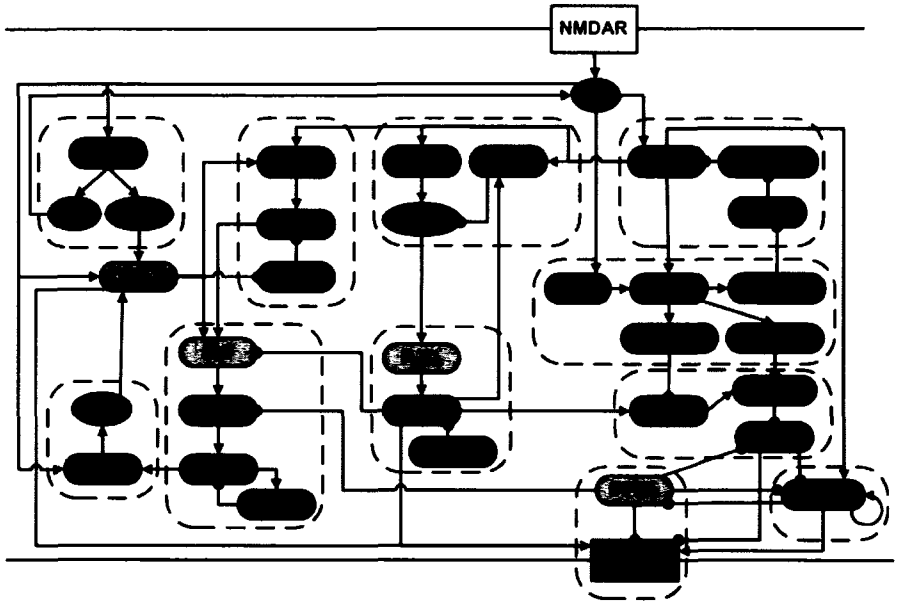
system at 1,000 s (almost final steady state).

An initial visual exploration of subfigure 4.3(a) shows mainly cold colors. Before the introduction of any calcium input, the system is in a non-potentiated state. Only the PP1 functional unit is active, as its red color indicates. The PP1 functional unit has a unique behavior in the LTP model, in that it is the only unit with a reverse behavior: at the normal steady state, the PP1 functional unit is highly active, while at the potentiation steady state it is inactive. The function of this unit is inhibitory; it limits the activation of the CaMKII and PP2A units. Part of the system bistability property is triggered by a persistent diminution in the PP1 unit inhibition activity, in order to let the CaMKII and PP2A units be persistently more active.

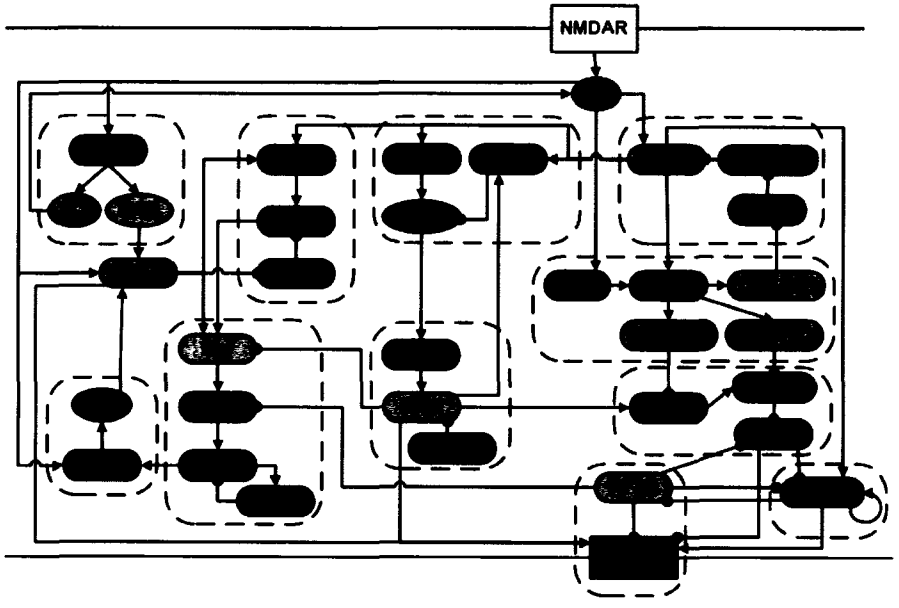
In subfigure 4.3(b), 4 seconds after the induction of the calcium stimulation, the cyan color of the calcium functional unit (Ca) denotes the higher concentration of this ion due to the tetanic calcium stimulation. The modules directly connected to the Ca unit, such as the PLC, PLA_2, CaM and CaN modules, react rapidly. The change in color of the first-degree neighbors of the Ca unit shows an activity increase, which has also been communicated to some of the Ca unit's second-degree neighbors, the GEF and cAMP modules.

Subfigure 4.3(c) shows the state of the system 80 seconds after the Ca stimulation, which occurs after 120 seconds of simulation. Two groups of functional units behave in distinctive ways. The first-degree neighbors of the Ca unit are returning to their initial state. However, the modules at the end of the signaling network, such as the MAPK, PP1, CaMKII and PP2A modules, which are slower to react to the Ca stimulation, are showing an increase in their activity.

This transitional state leads to the final state of the system, 880 seconds after the Ca stimulation and shown in subfigure 4.3(d). First-degree neighbors of the Ca unit are

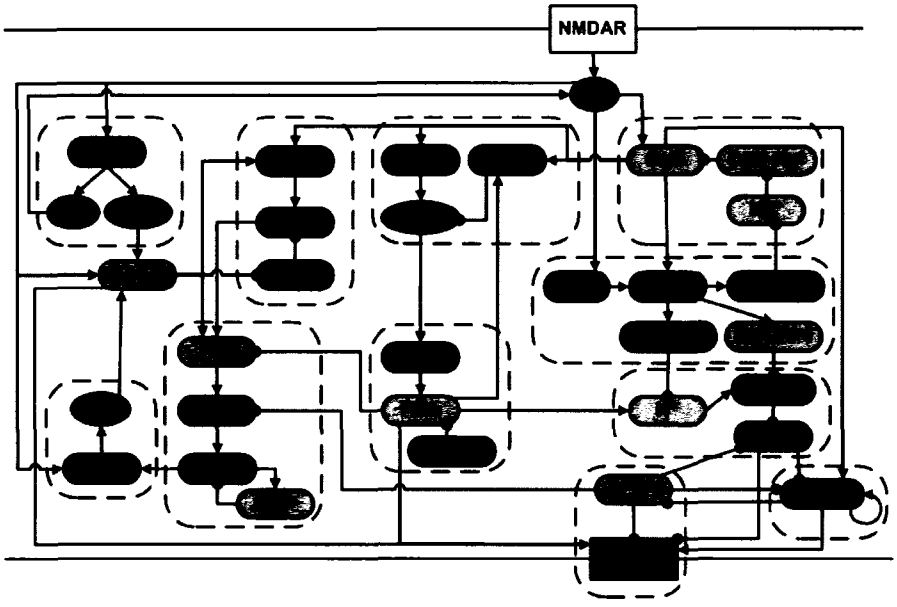


(a) 90s

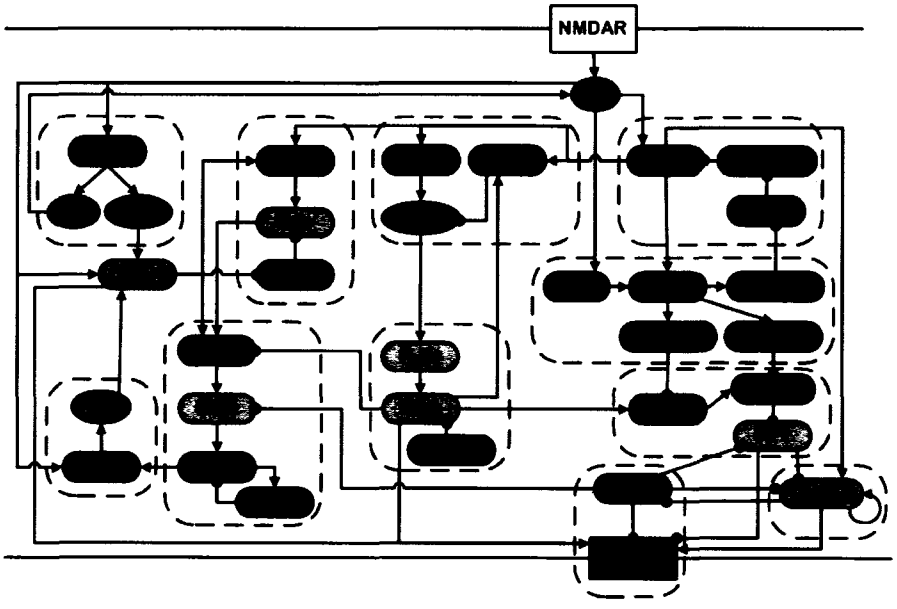


(b) 124s

Figure 4.3: Four snapshots of the animated representation of the LTP model simulation in normal conditions at various moments



(c) 200s



(d) 1000s

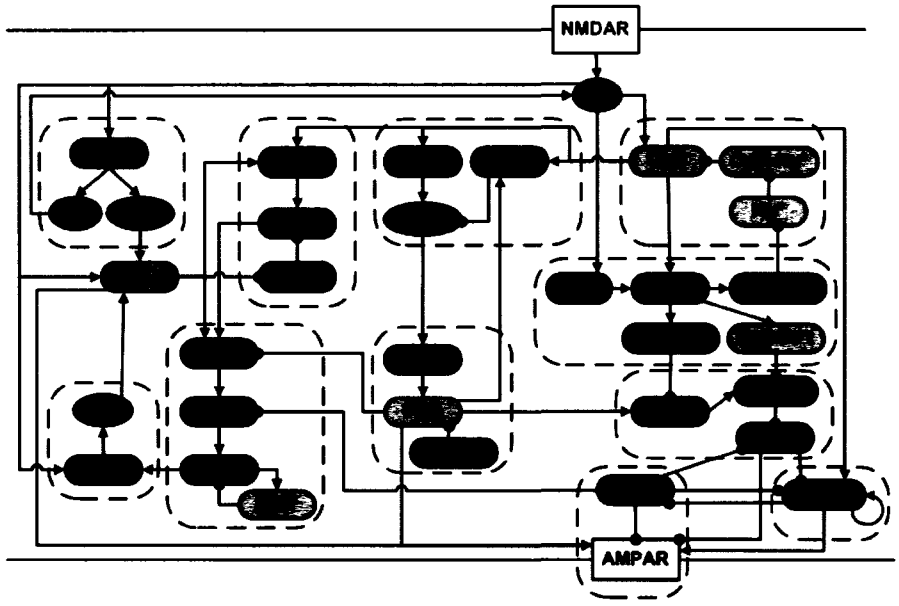
Figure 4.3: Four snapshots of the animated representation of the LTP model simulation in normal conditions at various moments (continued)

back in their initial state. The final colors of the modules at the end of the signaling pathway, however, which are located at the bottom of the frame, indicate a persistent activity in the system's final state. This is the potentiated state. Subfigures 4.3(a) and 4.3(d) are visual representations of the bistability of the LTP model. According to Kikuchi et al. (2003), a particularly important functional unit is the AMPAR unit (at the bottom center of the model). At the beginning of the simulation, this receptor is in a depressed, or inactive, state (blue), and, at the end of the simulation, this receptor is in a potentiated, or active, state (orange).

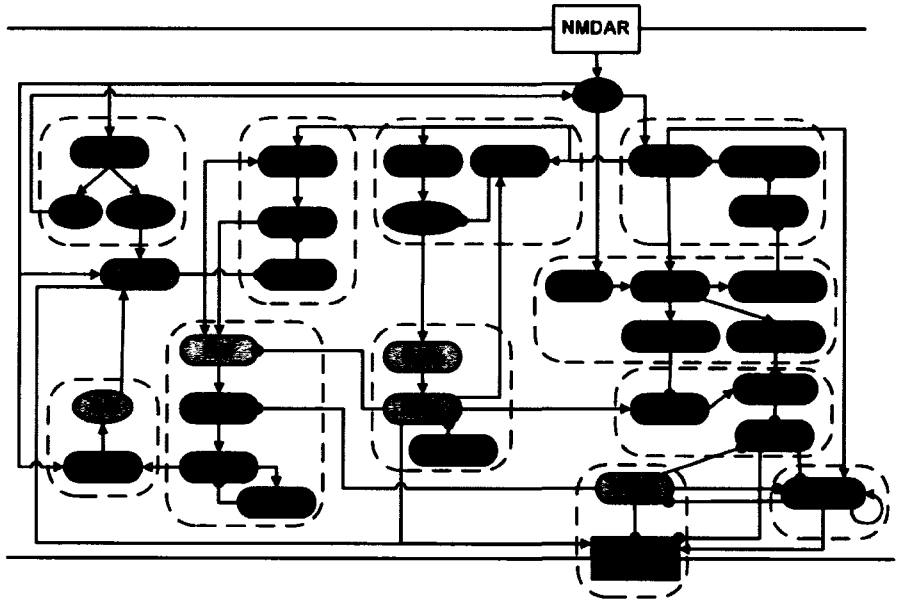
The visual exploration of Figure 4.4 shows a different behavior of the LTP model. The frames of the simulation of the modified model at 90 s and 124 s are not shown because they are identical to the same frames of the normal model. This suggests that the initial state and the first reactions to the Ca stimulation are similar in the two versions of the model. The impact of the modified parameter appears later in the simulation. Subfigure 4.4(a) shows a different transitional state: the modules at the end of the signaling pathway are less active than they were in subfigure 4.3(c). The last frame of the animation, shown in Subfigure 4.4(b), displays a final state that is similar to the initial state. Thus, this version of the LTP model is not bistable.

4.5.2 Spectral functional representation of the simulation data of the LTP model

The animated representation is made up of a sequence of images of the network topology showing the activity level of every functional unit of the model as time progresses. It enables analysis of the system behavior and provides information on the interactions between modules. The spectral representation is the set of colored timelines of the level of activity of the functional units. It provides a view of the complete simulation period. Figure 4.5 shows the spectral representation of the



(a) 200s



(b) 1000s

Figure 4.4: Two snapshots of the animation of the LTP model simulation with the parameter of the inactivation reaction of PP2A by CaMKII modified

functional units for the simulation data of the LTP model. For each functional unit, the spectrum on the left is the visualization of the simulation outputs of the normal LTP model, and the spectrum on the right is the visualization of the simulation outputs of the modified LTP model. The Ca spectrum indicates the calcium stimulus at 120s. The activity-level scale is at the left (violet = inactivity, red = activity) and the time proceeds from top to bottom, starting at 90 s and ending at 1,000 s.

Figure 4.5 shows the spectral representation of the functional units of the modules exhibiting a modification in their level of activity. A spectral representation can display more information in less space than a traditional representation using charts and plots. The spectra of the CaM and the CaN functional units for the two simulations confirm the behavior of the first-degree neighbors of the Ca unit suggested by the animations. They react rapidly to the increase in the calcium concentration, and their activity pattern is, for the most part, unaffected by modification of the PP2A reaction parameter. Because their final state is the same as their initial state, these units do not exhibit bistability. This observation is also true for the functional units of the PLC and cAMP modules. The spectra of the MAPK, CaMKII, AMPAR and PP2A functional units illustrate the behavior of the units at the end of the signaling network. Their activation is slower than the activation of the Ca first-degree neighbors, and it occurs after a delay. The activity pattern of the spectra of these units, produced from the simulation of the normal LTP model, shows the bistable property of this network: from an inactive, initial state, the spectra indicate that the functional units stabilize to a different and more active state. On the spectra of the modified LTP model, we can see a different activation pattern, one without bistability. These three functional units, after a transitory change in their activity level, eventually return to their normal steady state after the Ca stimulation. As already explained, the PP1 functional unit exhibits a reverse activity relative to the other functional units of the system. The MAPK, CaMKII, AMPAR and PP2A

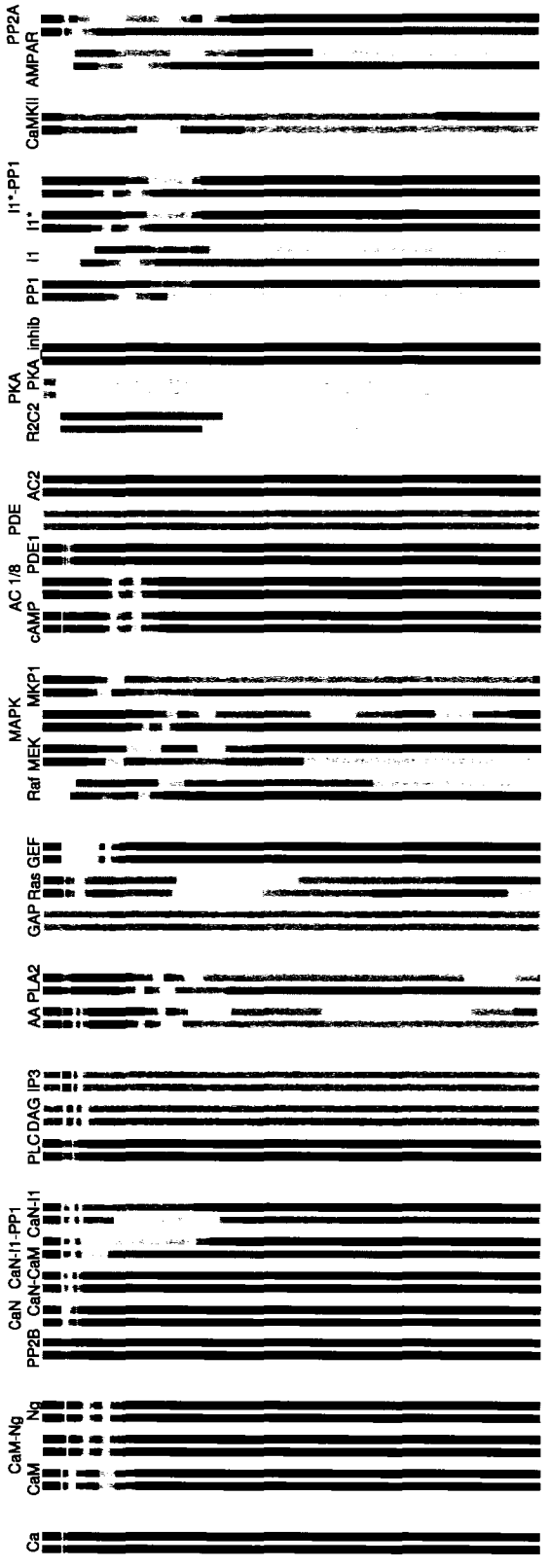


Figure 4.5: Spectral representation of the functional units of the LTP model. For each functional unit, the spectrum on the left is the visualization of the simulation outputs of the normal LTP model, and the spectrum on the right is the visualization of the simulation outputs of the modified LTP model.

units change from a low level of activity (cold colors) to a higher level of activity (warmer colors) as the PP1 unit undergoes the opposite variation.

The spectra of the PLA2 and AA functional units are examples of a combined behavior involving the behavior of a Ca first-degree neighbor and that of an end-of-signal unit. An initial activity change to the stimulus induction is noticeable immediately after this, followed by a return to its original, inactive state. A second inactivation occurs later, resulting in a persistent activation of the unit in the normal model.

4.6 Concluding remarks

In this paper, we explored an innovative perspective of the simulation data: a new data filter to present and analyze the data, based on the concept of molecular activity. This data filter is an approach to creating data subsets: functional units. In two steps – identification of the functional units and mathematical description of the molecular activity of the functional units – the simulation data are transformed, with a functional data filter, into data on the molecular activity of the model. This perspective of the molecular activity of the simulation data of a biochemical model formed the basis for two views of the simulation data of the model: a spectral functional view displaying the behavior of individual functional units, and an animated functional view of all the functional units of the model. This last type of data representation integrates the system structure and its dynamics into a single view. One of the goals of these views is to represent many numerical simulation outputs in such a way as to more easily access a system's overall systemic behaviors. This is achieved by identifying the functional units of a system and by visualizing the interactions between them. As demonstrated in this paper, this approach can replace numerous traditional concentration graphs and will help enable the rapid observation of the

impacts of structural or parametrical modifications of the model. This is an asset for the computational biologist developing a biochemical model. With the molecular activity approach, the computational biologist can create a data-preprocessing filter and afterwards easily explore the parameter space of the model and study the resulting behavioral modifications. This filter, built with the knowledge he or she usually uses to interpret the raw simulation data, is implemented only once and can then be applied to the data of any number of subsequent simulation runs. In this way, the modeler will have rapid access to the dynamic behavior of the different instantiations of a model.

A limitation of the simulation data representation approach based on molecular activity is that it is only efficient for qualitative analyses. The level of molecular activity is assigned on a relative basis that is specific to the characteristics of each functional unit. The result is a qualitative indication of the activity level of a functional unit. Despite this limitation, a qualitative glance at the dynamic behavior of a model is useful in the initial exploratory steps of simulation data analysis, while concentration graphs remain useful for detailed quantitative analyses. Taken together, concentration graphs and a data representation based on molecular activity are complementary approaches constituting two distinct perspectives, and it is advantageous to use them together.

At the moment, our approach and its two steps cannot be implemented in a systematic method for two reasons. First, the composition of functional units is intended to be customizable in order to fit research hypotheses and goals. The composition of functional units can be adapted to fit either micro- or macrosystem dynamics, thus allowing the representation of a model behavior at different levels of abstraction. The composition can also be focused on some specific molecular activities and leave some other activities out. In the LTP model, this is exemplified with the calcineurin mod-

ule (CaN). If the criterion for the composition of the functional units of this module was to be solely based on the recognition of a single molecular species and its various configurations, only one functional unit would be identified for this module. However, this protein phosphatase has three distinct regulatory effects of dephosphorylation on neurogranin (Ng) and inhibitory-1 protein (I1). To distinguish among these three different molecular activities, three functional units were identified, one for each dephosphorylation activity. The second reason why this data representation approach based on molecular activity cannot be implemented in a systematic method is that, as we have already mentioned, it does not involve a specific modeling language. Each modeling language will require its own algorithm. Depending on whether the model is specified with ordinary differential equations, is a Petri net model or has a stochastic nature, the algorithms will be different because of the particularities of the languages. However, the properties of a modeling language can be useful for the development of such an algorithm. Part of our current work involves the use of the invariants of a Petri net, which are structural properties, to implement an algorithm.

Acknowledgment

This work was supported in part by grant A-0141 from the Natural Sciences and Engineering Research Council of Canada and an NSERC Postgraduate Scholarship.

CHAPITRE 5

PETRI NET-BASED VISUALIZATION OF SIGNAL TRANSDUCTION PATHWAY SIMULATIONS

Abstract

Biological simulation[1] is a growing field, and attempts are being made to model and simulate ever more complex systems. Software tools for the visualization of simulation data have been developed, mostly for metabolic network models. In this paper, we present a visualization approach based on the Petri net formalism that is more suitable for signaling pathway models. The marking invariant property of Petri nets is used to identify meaningful biological entities and to generate, with the painting feature of Petri nets, animated representations of the simulation data in which the switching behavior of enzymes is highlighted, thereby allowing a visual exploration of the global dynamic behavior of complex biochemical models.

5.1 Introduction

The increasing complexity of computational biochemical models raises some problems for the visualization of their simulation data. The traditional means of representation, two-dimensional plots of time-series data, is more suitable than raw numbers, but they become cumbersome for large and complex models. Another problem with concentration plots is the loss of the structure of the biochemical model. The topol-

[1]Cet article a été soumis à la revue *IET Systems Biology*. Voir la référence (Hardy et Robillard, 2007b)

ogy of a network is an important source of information for analyzing a model's simulation outputs. Without a proper visualization technique, the topology has to be memorized by the user. These problems have been addressed by developing a local relative visualization approach, in which the graph structure of a biochemical model is constructed from its ordinary differential equations. The nodes of the graph represent the molecular substances of the model and the arcs represent the chemical reactions. The simulation data of the model are then mapped to the model structure: a graphical parameter of the nodes, such as their color or shape, changes to show the local concentration variations, resulting in an animated display of the simulation data of the model. The software tools MetVis (Qeli et al., 2003), SimWiz (Rost et Kummer, 2004) implement this local relative method. To enhance the visual rendering, the tools MetVis and SimWiz also have a three-dimensional graph generator (Qeli et al., 2004a; Wegner, 2005). With this local relative visualization approach, the simulation data of metabolic networks can be visualized in a more comprehensive way. One advantage to such a view is to make full use of the capacity of the human mind to detect structures and patterns in images, such as synchronicities, global changes of state and oscillations (Müller et Schumann, 2003). Also, it is possible to integrate the human being's perceptual abilities in data interpretation, a useful asset during the exploratory steps of data analysis (Keim, 2002). The combination of an animated representation of the topology of biochemical networks and simulation data allows a wealth of information to be accessible to the user and provides information on the system's dynamic behavior.

However, we detect two problems with this approach. First, a graphical representation of the model has to be generated with a graph layout algorithm, since the formalisms of the models that are visualized with these tools are not graphical in nature. For example, SimWiz uses an extended version of the Becker-Rojas algorithm (Becker et Rojas, 2001). Since there is no standard for biochemical network

graphical representation, each tool has its own format. A consequence of this is that users have to familiarize themselves with a new graphical representation once they have mathematically specified the model. The second problem appears when the local relative approach is used with models of signaling pathways. This approach is appropriate for visualizing the simulation data of metabolic pathways where metabolites are produced, consumed and degraded. However, it is less suitable for signal transduction networks, where the main biological activity for transmitting a signal is the change of state of enzymes. The activity in signaling pathways is normally the result of the activation and deactivation of enzymes through conformational changes as they turn "on" and "off", rather than the production and consumption of metabolites. In other words, enzymes can change from one state to another. Thus, the sum of the concentrations of the various states of an enzyme is modeled as a constant quantity.

This article presents a Petri net-based approach for the visualization of the simulation data of signal transduction networks which solves the graphical representation issue and the specific problem of the visualization of the switching activity of signaling transduction networks. The Petri net is a modeling language designed to study systems with causal concurrent properties. This formalism was used for the first time for biological modeling by Reddy et al. (1993). All the studies that followed their pioneering work correspond to either the qualitative or quantitative analysis of biological systems. To qualitatively analyze the biological properties of metabolic pathways, Petri-net properties such as invariants, liveness, boundedness and reachability have been used (Voss et al., 2003; Zevedei-Oancea et Schuster, 2003). More recently, concepts of Petri net theory have been used to identify functional units (Sackmann et al., 2006) or transduction activation components (Li et al., 2006) in signal transduction networks. These applications of Petri net concepts serve mainly to validate biological models. To quantitatively analyze biological systems, the Petri

net's high-level extensions that support simulation have been used (Hofestadt et Thelen, 1998; Matsuno et al., 2003c). The Notch-Delta signaling pathway in *Drosophila* (Matsuno et al., 2003b), the role of interleukin-6 in the fate of hematopoietic stem cells (Koh et al., 2006) and the Akt and MAPK pathways and their crosstalk (Troncale et al., 2006) are examples of signaling networks which have been quantitatively studied with the simulation of Petri net models. As well, review articles have been written on Petri net modeling approaches in biology (Hardy et Robillard, 2004; Pinney et al., 2003; Matsuno et al., 2006b). There are two main advantages to a Petri net-based approach for biological modeling. First, the theoretical elements of Petri nets with a mathematical basis are useful as a preliminary analysis tool for biological pathways. Second, researchers can easily model a biological system with Petri nets, partly due to the graphical nature of the formalism, and then study it with the simulation capabilities of Petri net tools.

In Petri net models of signaling pathways, the conservation relations known as marking invariants (or p-invariants) can be used to identify the sets of places associated with the different states of the enzymes of the model. Because of this relationship, marking invariants have been used for the validation of biochemical models, since they can help to detect inconsistencies (Zevedei-Oancea et Schuster, 2003; Voss et al., 2003). In this article, we propose to use p-invariants in a new visualization approach using painted Petri nets (Hardy et Robillard, 2006b). To the best of our knowledge, the visualization of biochemical models based on Petri net theory and presented in this paper is a new development in this field. In this study, Petri nets are not used for qualitatively or quantitatively analyzing biological systems, but rather for the visualization of the simulation data of signaling pathway models. We demonstrate that the Petri net's structural invariant property provides a basis for the automatic identification of meaningful groups of molecules and we show that the painted Petri net can be used to enhance the existing visualization method. Section 5.2 provides

a short introduction to the Petri net. Section 5.3 presents the new Petri net-based visualization approach and exemplifies it with the modeling of the calmodulin kinase pathway (CaMKII) of the complex signaling network of the hippocampal CA1 neuron. Section 5.4 compares the display produced by this new approach to another display produced by the local relative approach. Sections 5.5 and 5.6 discuss our results and the limitations of the visualization approach. We conclude by discussing future research directions.

5.2 Petri net concepts

This section introduces the basic concepts of Petri net theory that are necessary in the context of biological modeling, and which constitute a subset of the whole theory. Then, it describes a new Petri net feature designed to visualize the simulation data of Petri net models.

5.2.1 An introduction to Petri nets

Carl Adam Petri invented the Petri net formalism in 1962 to model the systems of concurrent processes (Petri, 1962). It is a modeling language depicting systems as directed bipartite graphs, i.e. directed arcs linking nodes from two disjoint sets. The first kind of node is the *place*. Places are the passive elements of the model, like entities and states. The second kind of node is the *transition*. Transitions are the active elements of the model, like processes and events. Directed *arcs* represent the causal relation between places and transitions, and a value is attached to them so that the modeled relations are weighted. In the graphical representation of a Petri net, places are displayed as circles and transitions as rectangles. For example, in a Petri net model of a biochemical system, places and transitions most often

represent molecular substances and chemical reactions respectively. Also, in a Petri net model of a metabolic pathway, the weights of the net's arcs might represent the stoichiometric coefficients.

The dynamic elements of a discrete Petri net are called *tokens*. Places contain an integer number of tokens, called *marks*, and transitions withdraw tokens from or add tokens to places. This process of withdrawing and adding occurs when a transition is fired. The firing of a transition can happen only when the transition preconditions, specified by the directed arcs whose destination is this transition, are met. Preconditions are fulfilled when the places at the source of the arcs, called *input places*, contain a number of tokens greater than or equal to the weight of the arcs. If the firing occurs, the transition postconditions, specified by the directed arcs whose source is this transition and destinations are *output places*, are met. The firing has two consequences: tokens are removed from input places, as stated by the preconditions, and tokens are added to output places in a number corresponding to the weights of the output arcs. There is no implicit conservation of tokens, which means that, for example, one transition can withdraw two tokens from the source place and add five tokens to the target place based on the weight of the transition. The state of a Petri net model is given by the token distribution in its places. The token distribution is called the marking. Most of the time, a firing modifies the marking, thus changing the state of the model. In the example of the biochemical Petri net model, the tokens are the molecules. The molecules change their forms (tokens moving from place to place) as they form molecular complexes, undergo chemical modification, etc. (transition firing). In this kind of Petri net model, the marking indicates the distribution of the molecules between the different molecular substances. The initial marking describes the state of the model before any transition firing occurs. A reachable marking is a marking that is attainable from the initial marking by a finite number of firings.

The conceptual framework of Petri nets is usually used to understand "how" a system works. The mathematical properties of some Petri net extensions, like invariants, serve to achieve a qualitative and structural analysis of a system and a validation of a model. The simulation capabilities of other Petri net extensions serve to realize a quantitative analysis of a system and a performance analysis of a model.

5.2.2 Invariants

Invariants are structural properties of Petri nets. In biochemical models, they usually have a mass conservation meaning. This relationship is detailed in subsection 5.3.3. This subsection presents the theoretical background on Petri net invariants.

The structure of a Petri net model, i.e. the arrangement of places, transitions and arcs, can be expressed in linear algebra by a matrix. This is the incidence matrix W. One dimension of the incidence matrix is the number of places of the model, and the other dimension is the number of transitions. Each element w_{ij} of the matrix indicates the token change at place i after firing transition j. From the incidence matrix, it is possible to determine the structural properties of Petri nets. Some of these properties are called invariants. Among all the reachable markings of a model, some quantities do not change, even when transitions are fired. This first type of invariant is the marking invariant (p-invariant). Every p-invariant of a Petri net model is a positive vector $\mathbf{x}$ that is a solution of the following equation:

$$\mathbf{x} \cdot \mathbf{W} = \mathbf{0} \tag{5.1}$$

A p-invariant characterizes a conservation component of the model, which is a set of places over which the weighted sum of the tokens is constant for every reachable marking.

Among all possible firing sequences of a model, some repetitions are possible. This second type of invariant is a firing invariant (t-invariant). Every t-invariant of a Petri net model is a positive vector $\mathbf{y}$ that is a solution of the following equation:

$$\mathbf{W} \cdot \mathbf{y} = \mathbf{0} \tag{5.2}$$

A t-invariant characterizes a repetitive component of the model, which is a firing sequence composed of several transitions causing a return to the model's initial state. In other words, the transition firings of a repetitive component together have a null effect on the marking of the model.

The Petri net formalism described so far is the original form of this formalism, known as the place-transition net. This modeling language can represent discrete, non-deterministic, asynchronous systems. Other types of Petri nets have been developed to enhance the modeling capabilities of the formalism. These other types are called extensions. For example, continuous and hybrid Petri nets are extensions to the original theory enabling the modeling of continuous quantities and processes (David et Alla, 2004). Ordinary differential equations have their equivalent in continuous Petri net models using continuous places and transitions. The marking of a continuous place is represented by a real number and a continuous transition has a speed. The hybrid Petri net combines elements of a discrete and a continuous nature.

5.2.3 Painted Petri nets

To visualize the simulation data of Petri net biochemical models, we suggest using the painting feature of Petri nets. By painting a Petri net model, it is possible to display the simulation data on the net's graphical structure. The painting feature is described below.

Several extensions of Petri net theory, like the stochastic, synchronous and hybrid extensions, have been used to complete the quantitative analysis of different systems. The choice of the appropriate extensions depends of the nature of the systems modeled. The simulation of a Petri net model, whatever the extensions, generates raw numbers of data on the model in the simulation environment. In the case of a discrete Petri net model, the token game display is usually used: tokens move throughout the model to show the evolution of the marking. Other Petri net tools can also generate 2-dimensional histograms or plots. Unfortunately, the problems related to the visualization of the simulation data of complex models are the same for complex Petri net models. To create a more intuitive and meaningful representation of the simulation data, the Petri net painting feature can be used (Hardy et Robillard, 2006b).

The aim of the Petri net painting feature is to visually indicate the evolution of the place contents of a Petri net model during the model's simulation. To achieve this goal, places are painted according to their token content. The painting approach can be applied on a Petri net model of any kind of extension used for simulation without modifying the properties of the extension on which paints are applied. A painted Petri net should not be confused with a colored Petri net (Jensen, 1992). In colored Petri nets, coloration is an abstract concept used to distinguish the different data token types of a model. In painted Petri nets, tints are used to visually indicate the temporal variation of the modeled entities represented by places. The painting rules use the mark of the places to generate values from 0 to 1 for each place. This numerical value is then associated with a tint from a color spectrum (0 being blue at one end and 1 being red at the other end, for example). A paint value of a place is computed for each time step of the simulation. The result is an animated graphical representation of the painted Petri net model which combines the topology of the model and the simulation data.

5.3 Method

The following section presents the Petri net-based visualization approach for the simulation data from signal transduction models. First, we explain how to represent kinetics models with Petri net elements, and we illustrate this modeling process with the signaling pathway of the CaMKII. We then identify the marking invariants of the example and define an adaptation of the marking invariant concept which enables the identification of other meaningful entities in signaling transduction models. Finally, we specify two painting rules which are used in the next section to compare the local relative visualization approach with our approach based on the marking invariant of a Petri net model.

5.3.1 Petri net-based modeling of biochemical reactions

Modeling systems with ordinary differential equations (ODE) is the quantitative method most commonly used to represent signal transduction pathways (Eungdamrong et Iyengar, 2004). For that reason, we show here the way to transform this type of model into a continuous Petri net model with marking-dependent speeds.

A biochemical model specified with ODEs is partly composed of chemical reactions like the following:

$$A + B \underset{k_b}{\overset{k_f}{\rightleftharpoons}} C + D \tag{5.3}$$

In this chemical reaction, there are four molecular substances (substrates A and B, and products C and D) and two chemical processes (forward and backward reactions). The variation in concentration of substrate A is expressed by the following:

$$\frac{d[A]}{dt} = k_b[C][D] - k_f[A][B] \tag{5.4}$$

Four continuous places and two continuous transitions can represent the chemical reaction of equation 5.3 when they are structured as in Fig. 5.1. In this small Petri net model, one can see that the substances modeled by places A and B react together through transition t_1 at a speed corresponding to the product of the concentration of these substances (marks x_1 and x_2) and the kinetic parameter k_f to produce the substances modeled by places C and D.

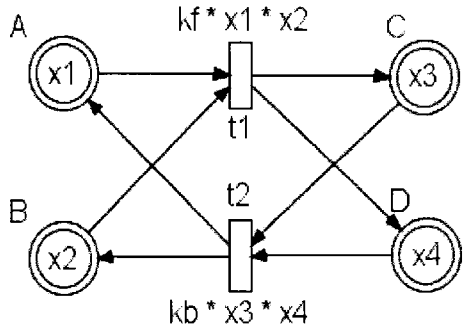


Figure 5.1: Petri net representation of a chemical reaction.

Another important building block of signaling transduction pathways is the enzymatic reaction. These reactions can be expressed either by the bidirectional reaction of the formation of a complex composed of an enzyme and a substrate, followed by an irreversible reaction creating the product from the complex, or simply by the Michaelis-Menten formulation. The Petri net representation of these two forms of enzymatic reaction is given in (Koh et al., 2006).

Following these modeling guidelines, it is possible to create a Petri net model for any chemical kinetics model where the concentration is only a function of time. It is important to note that Petri net modeling is not limited to continuous deterministic systems with known kinetic parameters. For example, a method has been developed to systematically build discrete Petri net models of signal transduction pathways that reflect their qualitative biological behavior from experimental observations, without any precise kinetic parameters (Sackmann et al., 2006), and stochastic Petri nets

have been used to model the stochastic process of *ColE1* plasmid replication, since this process involves low concentrations of biological compounds (Goss et Peccoud, 1998, 1999). Also, hybrid functional Petri nets are a modeling formalism designed specifically for biological hybrid modeling (Matsuno et al., 2003c).

5.3.2 Petri net model of the calmodulin kinase pathway

With the previous modeling guidelines, we transform an ODE model of the CaMKII pathway into a continuous Petri net model. In the next sections, we identify the invariants of this model and we use a painted version of the model to compare two visualization approaches of simulation data.

The CaMKII pathway, shown in Fig. 5.2, is one module among several others which makes up a larger signaling network. The ODE model of this pathway comes from a study on the bistability of a complex signaling network of the hippocampal CA1 neuron (Bhalla et Iyengar, 1999). The kinetic parameters and the initial concentration of the pathway substances were experimentally obtained and can now be found in the Database of Quantitative Cellular Signaling (Sivakumaran et al., 2003). In Fig. 5.2, reversible reactions are drawn as bidirectional arrows, and enzymatic reactions are represented by arrows with two bends. For clarity, a reactant can appear multiple times and in the two depicted enzymatic reactions. $\text{CaM.CaMKII}_{activeforms}$ stands for all phosphorylated forms of CaMKII. After Fig. 1I of (Bhalla et Iyengar, 1999). The CaMKII pathway is composed of an autophosphorylation cycle in which each phosphorylated form of this enzyme contributes to the activation of the enzyme by phosphorylation at Thr^{286}. The inflow of calcium ions is first signaled to the CaMKII pathway by increased binding with calmodulin (CaM.Ca_4). The CaM.CaMKII complex is then able to activate itself by autophosphorylation. Conversely, when the protein phosphatase 1 (PP1) is active, it is able to dephosphorylate CaMKII, thus

deactivating the kinase. Another regulation effect of the calcium inflow, also modeled in the complete model of the neuron, is to delay the deactivation of CaMKII by PP1. The source of this effect is the production of cyclic adenosine monophosphate (cAMP) caused by CaM activation. The elevation of cAMP levels causes an activation of PKA, which in turn phosphorylates the Inhibitor-1 protein, leading to the inhibition of PP1.

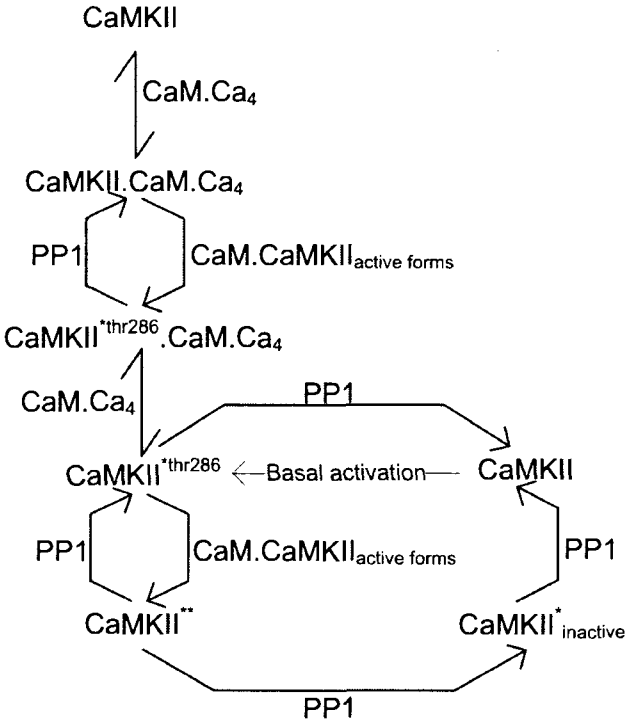


Figure 5.2: Reaction schemes of the calmodulin kinase pathway.

We create a continuous Petri net model from this ODE model by assigning a place to each substance and a transition to each reaction. The initial marking of the model corresponds to the initial concentrations, and the speed of the transitions corresponds to the kinetic equation of the reactions. The resulting Petri net model of the CaMKII pathway is shown in Fig. 5.3. The specifications of the places of the Petri net model are given in Table 5.1, and the specifications of its transitions and arcs are given in Table 5.2.

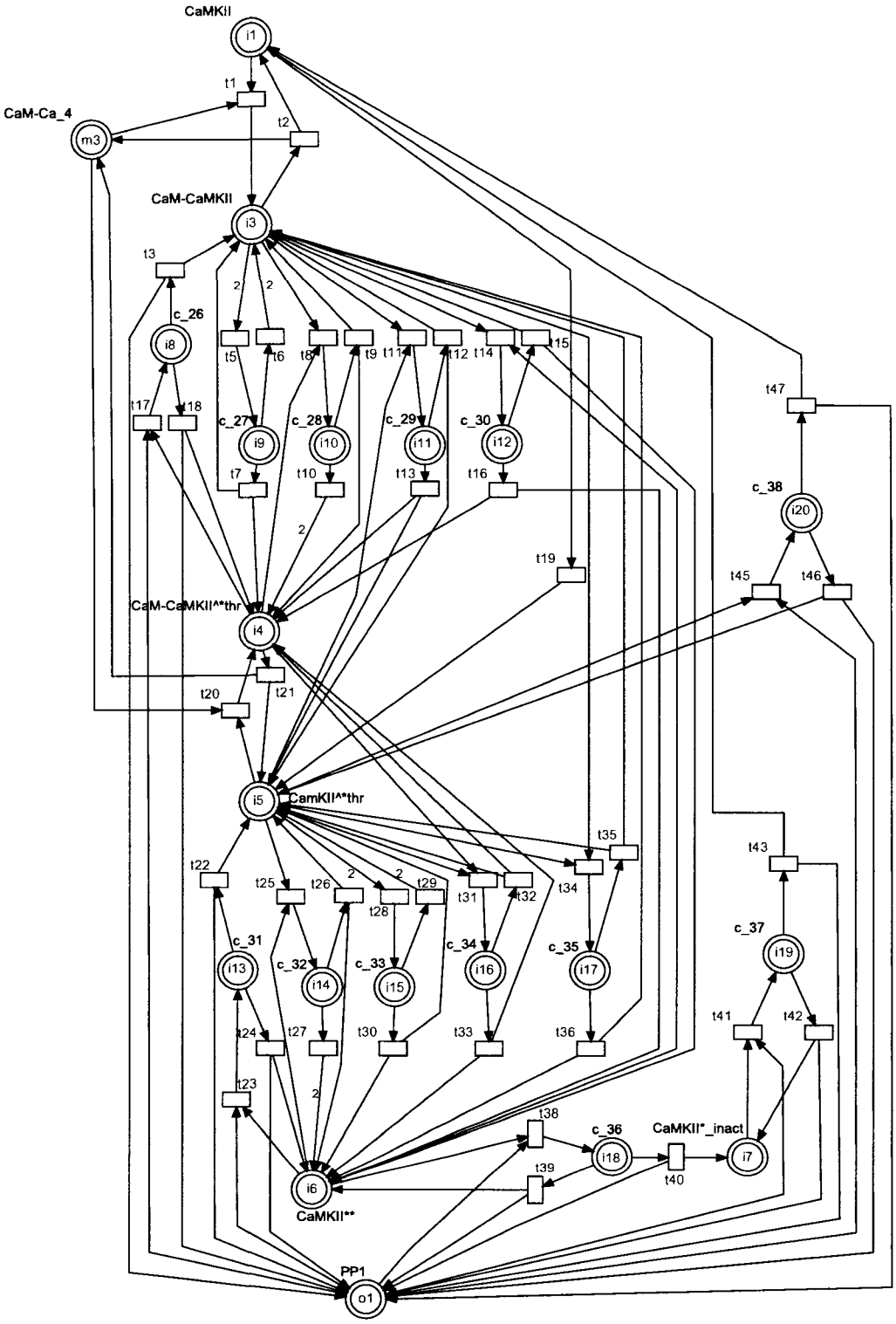


Figure 5.3: Continuous Petri net model of the calmodulin kinase pathway.

Tableau 5.1: Specifications of the places of the Petri net model of the calmodulin kinase pathway

Name	Substance	Variable
P_1	CaMKII	i_1
P_2	CaM.CamKII	i_3
P_3	CaM.CaMKII^{*thr}	i_4
P_4	CaMKII^{*thr}	i_5
P_5	CaMKII**	i_6
P_6	CaMKII$^{*}_{inactive}$	i_7
P_7	complex$_{26}$	i_8
P_8	complex$_{27}$	i_9
P_9	complex$_{28}$	i_{10}
P_{10}	complex$_{29}$	i_{11}
P_{11}	complex$_{30}$	i_{12}
P_{12}	complex$_{31}$	i_{13}
P_{13}	complex$_{32}$	i_{14}
P_{14}	complex$_{33}$	i_{15}
P_{15}	complex$_{34}$	i_{16}
P_{16}	complex$_{35}$	i_{17}
P_{17}	complex$_{36}$	i_{18}
P_{18}	complex$_{37}$	i_{19}
P_{19}	complex$_{38}$	i_{20}
P_{20}	CaM.Ca$_4$	m_3
P_{21}	PP1	o_1

Tableau 5.2: Specifications of the transitions and arcs of the Petri net model of the camodulin pathway

Name	Speed	From		To	
		variable	factor	variable	factor
t_1	$50 \times i_1 \times m_3$	i_1	1	i_3	1
		m_3	1		
t_2	$5 \times i_3$	i_3	1	i_1	1
				m_3	1
t_3	$0.35 \times i_8$	i_8	1	i_3	1
				o_1	1
t_5	$1.543 \times {i_3}^2$	i_3	2	i_9	1
t_6	$2 \times i_9$	i_9	1	i_3	2
t_7	$0.5 \times i_9$	i_9	1	i_3	1
				i_4	1
t_8	$1.543 \times i_3 \times i_4$	i_3	1	i_{10}	1
		i_4	1		
t_9	$2 \times i_{10}$	i_{10}	1	i_3	1
				i_4	1
t_{10}	$0.5 \times i_{10}$	i_{10}	1	i_4	2
t_{11}	$i_3 \times i_5$	i_3	1	i_{11}	1
		i_5	1		
t_{12}	$2 \times i_{11}$	i_{11}	1	i_3	1
				i_5	1
t_{13}	$0.5 \times i_{11}$	i_{11}	1	i_4	1
				i_5	1
t_{14}	$i_3 \times i_6$	i_3	1	i_{12}	1
		i_6	1		
t_{15}	$2 \times i_{12}$	i_{12}	1	i_3	1
				i_6	1
t_{16}	$0.5 \times i_{12}$	i_{12}	1	i_4	1
				i_6	1
t_{17}	$0.343 \times i_4 \times o_1$	i_4	1	i_8	1
		o_1	1		
t_{18}	$1.4 \times i_8$	i_8	1	i_4	1
				o_1	1
t_{19}	$0.003 \times i_1$	i_1	1	i_5	1
t_{20}	$1000.2 \times i_5 \times m_3$	i_5	1	i_4	1
		m_3	1		

Continued on next page

Tableau 5.2 – continued from previous page

Name	Speed	From		To	
		variable	factor	variable	factor
t_{21}	$0.1 \times i_4$	i_4	1	i_5	1
				m_3	1
t_{22}	$0.35 \times i_{13}$	i_{13}	1	i_5	1
				o_1	1
t_{23}	$0.343 \times i_6 \times o_1$	i_6	1	i_{13}	1
		o_1	1		
t_{24}	$1.4 \times i_{13}$	i_{13}	1	i_6	1
				o_1	1
t_{25}	$12 \times i_5 \times i_6$	i_5	1	i_{14}	1
		i_6	1		
t_{26}	$24 \times i_{14}$	i_{14}	1	i_5	1
				i_6	1
t_{27}	$6 \times i_{14}$	i_{14}	1	i_6	2
t_{28}	$12 \times {i_5}^2$	i_5	2	i_{15}	1
t_{29}	$24 \times i_{15}$	i_{15}	1	i_5	2
t_{30}	$6 \times i_{15}$	i_{15}	1	i_5	1
				i_6	1
t_{31}	$18.519 \times i_4 \times i_5$	i_4	1	i_{16}	1
		i_5	1		
t_{32}	$24 \times i_{16}$	i_{16}	1	i_4	1
				i_5	1
t_{33}	$6 \times i_{16}$	i_{16}	1	i_4	1
				i_6	1
t_{34}	$18.519 \times i_3 \times i_5$	i_3	1	i_{17}	1
		i_5	1		
t_{35}	$24 \times i_{17}$	i_{17}	1	i_3	1
				i_6	1
t_{36}	$6 \times i_{17}$	i_{17}	1	i_3	1
				i_6	1
t_{38}	$0.343 \times i_6 \times o_1$	i_6	1	i_{18}	1
		o_1	1		
t_{39}	$1.4 \times i_{18}$	i_{18}	1	i_6	1
				o_1	1
t_{40}	$0.35 \times i_{18}$	i_{18}	1	i_7	1
				o_1	1

Continued on next page

Tableau 5.2 – continued from previous page

Name	Speed	From		To	
		variable	factor	variable	factor
t_{41}	$0.343 \times i_7 \times o_1$	i_7	1	i_{19}	1
		o_1	1		
t_{42}	$1.4 \times i_{19}$	i_{19}	1	i_7	1
				o_1	1
t_{43}	$0.35 \times i_{19}$	i_{19}	1	i_1	1
				o_1	1
t_{45}	$0.343 \times i_5 \times o_1$	i_5	1	i_{20}	1
		o_1	1		
t_{46}	$1.4 \times i_{20}$	i_{20}	1	i_5	1
				o_1	1
t_{47}	$0.35 \times i_{20}$	i_{20}	1	i_1	1
				o_1	1

5.3.3 Marking invariants and mass conservation in biochemical models

The relationship between the marking invariants of a Petri net, also known as p-invariants, and the basic concepts of biochemical modeling has already been established. In models of metabolic pathways, marking invariants express conservation relations for metabolites (Voss et al., 2003; Zevedei-Oancea et Schuster, 2003). In signal transduction models, marking invariants represent a different kind of conservation relation (Sackmann et al., 2006). A signaling pathway is considered active when enzymes change state to transmit a signal, and the total concentration of all forms of an enzyme is modeled as a constant quantity. This quantity is a marking invariant in a Petri net model.

The p-invariants of a Petri net model are all the vectors $\mathbf{x}$ that solve equation 5.1. The non-zero values of these vectors indicate the places of the Petri net that are part

of a conservation component. The marking invariant associated with a p-invariant is the sum of the marks of the places of the conservation component, where each mark is multiplied by its associated value in the vector. To find the invariants of simple Petri net models, David et Alla (2004) described an algorithm. There are more efficient algorithms using linear programming (Takano et al., 2001), and Petri net analysis software is also freely available (Starke, 2003). The marking invariants of the Petri net model of the CaMKII pathway are shown in Table 5.3. These invariants correspond to the three enzymes present in the model: CaMKII, CaM and PP1. The marking invariants identify the places representing the different substances containing at least one molecular form of the enzymes. When a coefficient other than one precedes the marking of a place, it means that there is an equivalent number of enzyme molecules assembled in a multi-enzyme complex. To obtain a constant sum, the concentration of these complexes has to be weighted, so as to form a constant positive linear combination.

Tableau 5.3: The marking invariants of the Petri net model of the calmodulin kinase pathway

Substance	P-invariant Marking invariant	
CaMKII	$\mathbf{x}_1$	$= [1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 0\ 0]$
	N_{CC_1}	$= i_1 + i_3 + i_4 + i_5 + i_6 + i_7 + i_8 + 2i_9 + 2i_{10} + 2i_{11} + 2i_{12} + i_{13} + 2i_{14} + 2i_{15} + 2i_{16} + 2i_{17} + i_{18} + i_{19} + i_{20}$
CaM	$\mathbf{x}_2$	$= [0\ 1\ 1\ 0\ 0\ 0\ 1\ 2\ 2\ 1\ 1\ 0\ 0\ 0\ 1\ 1\ 0\ 0\ 0\ 1\ 0]$
	N_{CC_2}	$= i_3 + i_4 + i_8 + 2i_9 + 2i_{10} + i_{11} + i_{12} + i_{16} + i_{17} + m_3$
PP1	$\mathbf{x}_2$	$= [0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 0\ 1]$
	N_{CC_3}	$= i_8 + i_{13} + i_{18} + i_{19} + i_{20} + o_1$

Often, biochemical models are open systems with the entry and exit of mass in the systems causing a perturbation or stimulation. Because of this, some places modeling external substances are not part of any conservation component. This is not

the case for the Petri net model of the CaMKII pathway, in which every place is part of at least one conservation component. However, it is the case for some places of the complete complex signaling network model of the hippocampal CA1 neuron, where these places are not part of any conservation component, like the places modeling the calcium ions and the cAMP. The temporary removal of the places and transitions responsible for the input/production and output/consumption of these external substances enables the identification of a new type of component related to p-invariants. After their removal, new solutions might be found to equation 5.1, also corresponding to mass conservation relationships. We call these new solutions *modified p-invariants*. The places of the sets found after this modification of the model represent molecular entities partly formed of the external substances. The positive linear combinations of the modified p-invariants no longer produce constant values. However, these combinations do provide useful information. For example, in the complete model of the complex signaling network of the hippocampal CA1 neuron, the positive linear combinations of the two modified p-invariants are equations giving the total concentration of calcium ions and cAMP in the system, in one form or another, at each time step. The modified p-invariants can come into play in the process of painting a model.

5.3.4 Local relative and marking invariant-based painting rules

To display the simulation data of a model with the Petri net painting feature, a painting rule must be defined. A painting rule is an equation transforming the mark of a place into a value from 0 to 1. This value is then associated with a tint that is applied to the place in the Petri net graphical representation. In this section, we present two different painting rules. The first is associated with the data visualization approach used by existing software tools, while the second is a new approach. In the

next section, the two approaches are compared.

The local relative approach converts raw simulation data into relative values using the range between the minimum and maximum values of each variable. This produces a percentage, a local relative value, which can be transformed into a color or a scaling factor for data visualization. With the first painting rule, given by equation 5.5, a local relative value $paint_{P_{i,t}}$ is calculated for each place P_i at time t.

$$paint_{P_{i,t}} = (mark_{P_{i,t}} - min_{P_i})/(max_{P_i} - min_{P_i}) \tag{5.5}$$

where $mark_{P_{i,t}}$ represents the mark of the place P_i at time t, and max_{P_i} and min_{P_i} represent the minimum and maximum marks for the place P_i during the entire simulation.

With the second painting rule, we present a novelty: the use of marking variants in a painting rule. The rule is the following:

$$paint_{P_{i,t}} = mark_{P_{i,t}}/N_{CC_x} \tag{5.6}$$

where the place P_i is part of the conservation component CC_x having N_{CC_x} for marking invariant. This rule computes the ratio between the concentration of a single molecular form at simulation time t and the concentration of this substance in all forms.

Because a place can be part of several conservation components, a place can logically have x paint values, where x is the number of conservation components of which this place is a part. To solve this potential problem, we suggest that users assign priorities to conservation components. By doing so, they decide on the paint from which conservation components will be preferred in the display. Thus, it is possible

to ensure that every place which is part of multiple conservation components has a single tint in the final display.

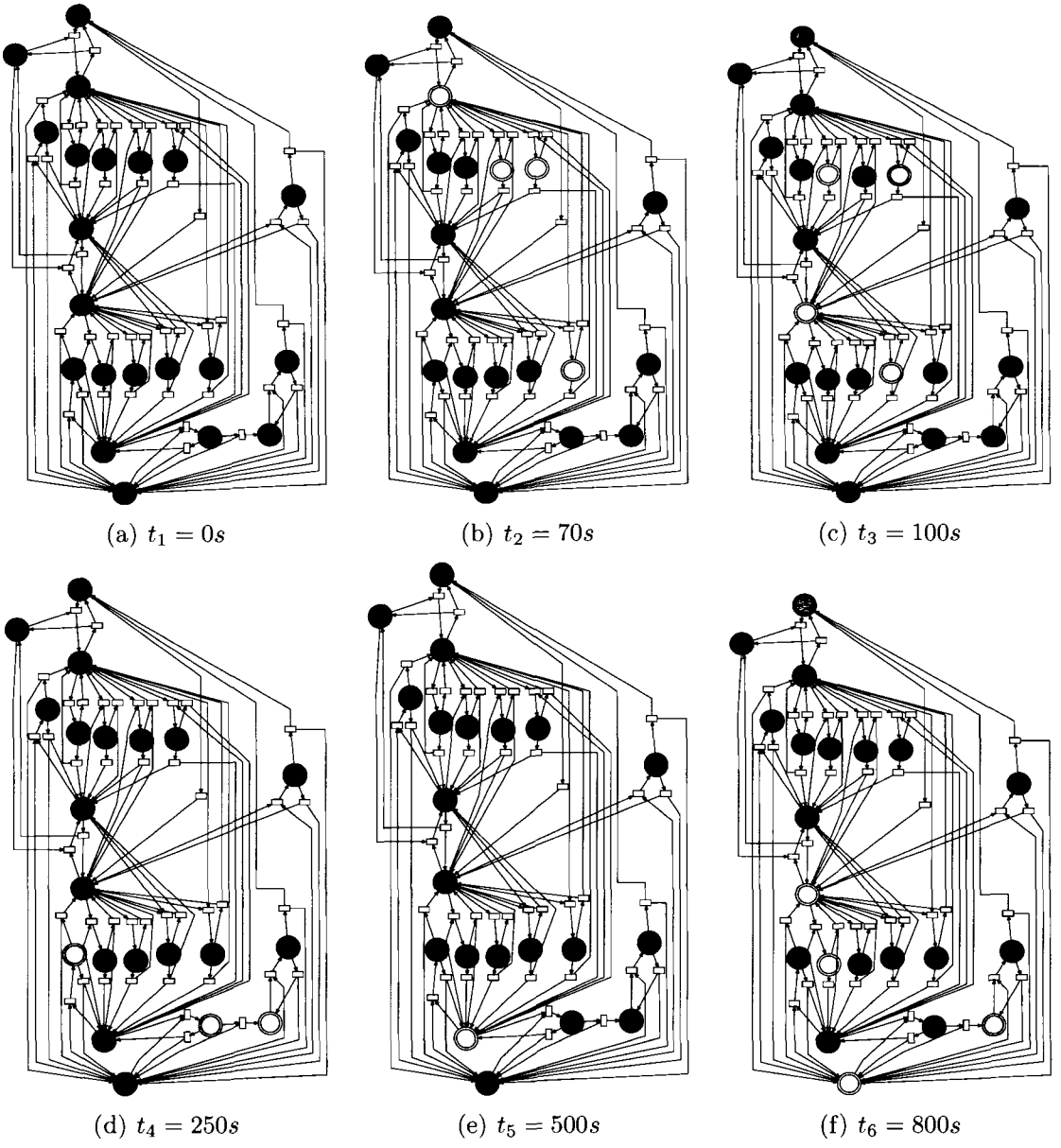
5.4 Results

In this section, we show an initial display of the simulation data of the CaMKII pathway which uses the local relative painting rule, and a second display which uses the marking invariant-based painting rule. The complete Petri net model of the complex signaling network of the hippocampal CA1 neuron has been simulated, but only the data pertaining to the CaMKII pathway are shown. A calcium inflow, reproducing a neuronal tetanic stimulation, is induced after 60 seconds of simulation time. The simulation software used is Genomic Object Net (a hybrid functional Petri net simulator commercially known as Cell Illustrator)(Nagasaki et al., 2003). Both displays are animated representations, where each frame shows the places of the model with their respective paint at time t. The paints vary from blue to white to red on a spectrum of 100 discrete colors. Only when the simulations are completed can the paints be computed and the animated representations of the simulation data be displayed.

Figure 5.4 shows a series of snapshots of the animated painted Petri net model of the CaMKII pathway using the local relative painting rule of equation 5.5 at various moments during the simulation. In this display, the bluer a place is, the closer its mark is to the place's minimal value, and the redder a place is, the closer its mark is to the place's maximal value. A consequence of this painting rule is the absence of a relationship between the red paint of one place and the red paint of another place; the same red paint might represent very different concentration values. Their only similarity is in representing the maximal value of the mark of the place. In other

words, it is strictly local relative information. Neither the network nor any functional entity is taken into account. Despite this shortcoming, it is possible to describe the concentration variations of the CaMKII pathway model. At t_1, the inactive state of CaMKII (the place at the top of the model and identified as CaMKII) is red, and thus at its maximal concentration value. The concentration of PP1 is also at its highest. The concentrations of some CaMKII activated forms in the autophosphorylation cycle are also high. At t_2, 10 seconds after the calcium stimulation, many places of the CaMKII activated forms in the autophosphorylation cycle are redder. At t_3, we can observe a slight decrease in the level of the inactivate CaMKII and PP1 reaches its minimal level, due, as we have already explained, to regulation outside the CaMKII pathway. At t_4, the only CaMKII form at its maximal concentration is the completely phosphorylated form (CaMKII**). At t_5, the CaMKII** concentration decreases, while the concentration of one of the two inactive CaMKII forms (CaMKII$^*_{inactive}$) increases. At t_6, the system slowly returns to its initial state, the levels of two inactive CaMKII forms are close to their maximal values and the PP1 concentration is increasing.

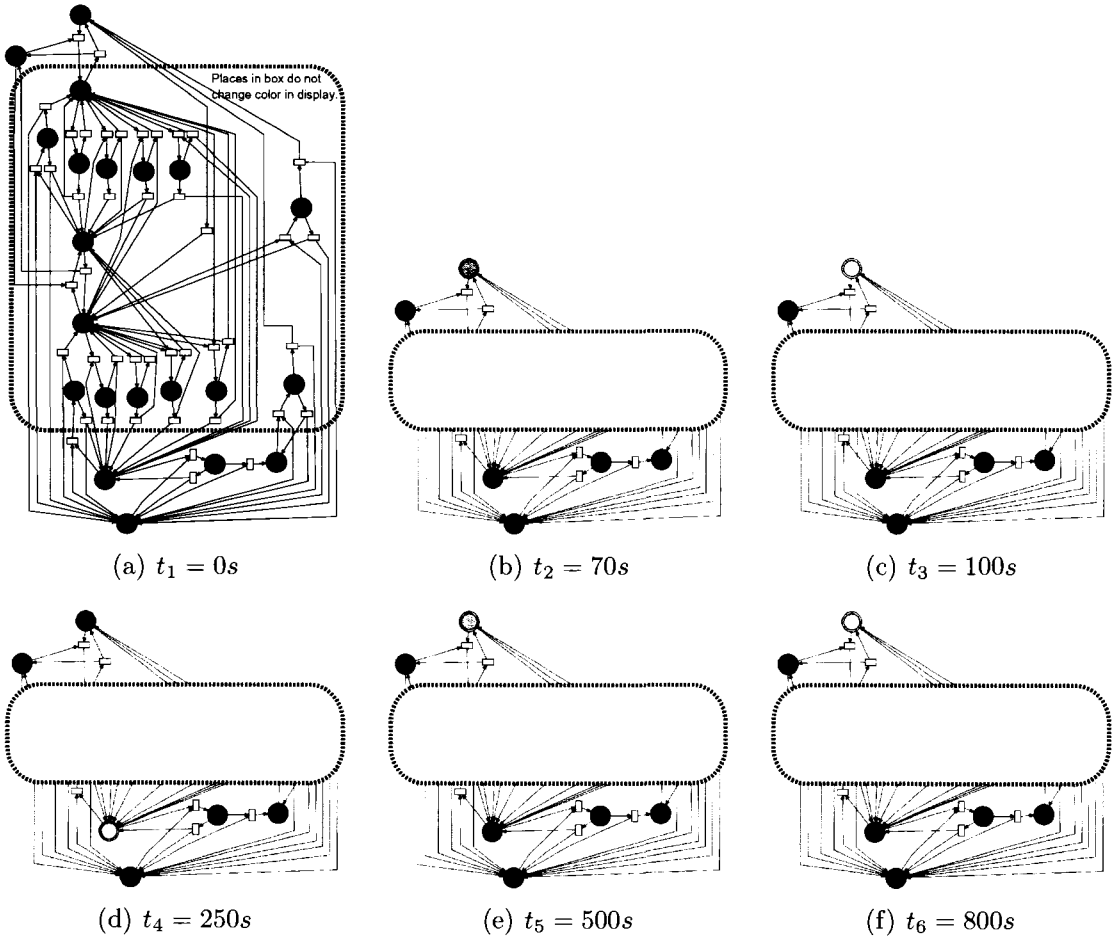
Figure 5.5 shows a series of snapshots of the animated painted Petri net model of the CaMKII pathway at various moments during the simulation using the marking invariant-based painting rule of equation 5.6. A first look at this display reveals that there is less red. The reason for this color distribution is that the paint values of the places represent absolute values in this display: the bluer a place is, the smaller its mark is compared to the marking invariant of its conservation component. In the case of the CaMKII conservation component, the sum of the concentrations of all the different forms is, according to the initial parameters of the model, 70 μM (this is the N_{CC_1} value of equation 5.6 for the places that are part of the CaMKII conservation component). With this painting rule, every concentration variation is scaled to the total enzyme concentration. This causes small concentration variations



(a) $t_1 = 0s$ (b) $t_2 = 70s$ (c) $t_3 = 100s$

(d) $t_4 = 250s$ (e) $t_5 = 500s$ (f) $t_6 = 800s$

Figure 5.4: A series of snapshots showing the animated painted Petri net model of the calmodulin kinase pathway at different moments in time using a local relative painting rule.



(a) $t_1 = 0s$ (b) $t_2 = 70s$ (c) $t_3 = 100s$

(d) $t_4 = 250s$ (e) $t_5 = 500s$ (f) $t_6 = 800s$

Figure 5.5: A series of snapshots showing the animated painted Petri net model of the calmodulin kinase pathway at different moments in time using a marking invariant-based painting rule.

to disappear from the display, and the focus is thus set on significant variations. In the CaMKII pathway model, there are places that are part of more than one conservative component. For these places, the painting priority is then given to the CaMKII conservation component.

The description of the simulation data of the CaMKII pathway shown in 5.5 is straightforward. At t_1, most of the CaMKII concentration is in an inactive form (the CaMKII place). After stimulation, at t_2 and t_3, there is a slow decrease in the concentration of the inactive form. At t_4, we see that almost half of all the CaMKII molecules are in the active form, CaMKII**. At t_5 and t_6, we can observe a slow return of the pathway to its initial inactive state.

5.5 Discussion

The two animated representations of the simulation data of the CaMKII pathway model presented in the previous section differ in their level of detail. The display of the simulation data with the local relative painting rule shows the detail of every concentration variation in the model, while the display with the marking invariant-based painting rule presents a clearer picture of the switching behavior of this pathway.

The local relative approach is less suitable for showing the dynamic behavior of signal transduction models than the marking invariant-based approach for two reasons. First, it is difficult to compare the paints of the places in the first display because of their relative nature. The minimum-maximum range can vary greatly from one place to another. For example, the initial concentration of CaMKII is 52.03 μM and the initial concentration of CaMKII^{*thr} is 0.014 μM, but, because these values constitute the local maxima, both places are painted red. It is difficult to describe the behavior of this pathway with the display shown in Fig. 5.4 without any knowledge of the

absolute concentrations of the molecular substances. As we have already pointed out, at t_1, the places representing the inactive CaMKII form and the active CaMKII^{*thr} form are both painted red because their values are close to the maximal concentration they reach during simulation. Does this mean that the CaMKII enzyme in its initial state is equally active and inactive? In Fig. 5.4, it is impossible to know if CaMKII is initially active or inactive, because the absolute values are absent from the first display.

A second problem with the local relative painting rule is the numerous color changes in the display of a complex signaling network. For some substances, the presence of many different intermediate molecular forms with a small minimum-maximum range introduces unwanted color flashings. In the snapshots of the animated display shown in Fig. 5.4, irrelevant color changes for several places hinder the visualization of the network behavior. The impact of this problem grows exponentially for complex models: the animated representation of the simulation data of the complete network composed of six similar interconnected modules shows 100 nodes changing color! The objective of easing the visualization of the dynamic behavior is not achieved. A global painting rule (the same global minimum value and global maximum value for every place) would not be a good solution either, because of the different orders of concentration in the model.

These two limitations of the local relative approach are not problematic for metabolic pathways. Local relative information is all that is needed to visualize the increase and decrease in the concentration of metabolites. Thus, it is sufficient to see the dynamic behavior of this family of biochemical systems. For signaling pathways, some kind of network-related information is required to attain this objective, and it can be attained with the marking invariants of Petri net models. The animated representation of the Petri net model of the CaMKII pathway painted with the marking

invariant rule shows color modifications which are relevant to the evolution of the distribution between the different forms of the CaMKII enzyme. The distribution for this enzyme, as for others present in signal transduction systems, is linked to its functional activation or inactivation. For example, the distribution of CaMKII in phosphorylated and dephosphorylated forms indicates whether this enzyme is "on" or "off". This painting rule can also eliminate concentration variations extraneous to the dynamic behavior of enzymes, and results in the appropriate level of detail for understanding systemic behavior: the simple display of the dynamic behavior of one enzyme is an asset for the visualization of the simulation data of a complex signaling pathway model with several enzymes.

We could rapidly conclude that a painting rule based on conservation components and marking invariants computes paints which are always more significant in terms of the dynamic behavior of a model than the paints computed with the local relative painting rule. Nonetheless, the marking invariant-based painting rule also has drawbacks.

While it is perfect for enzymes with a switching behavior, this rule is less effective for depicting an enzyme which performs its function with only a small part of its total concentration. For example, the paints of the $CaM.Ca_4$ and PP1 places in the CaMKII model are computed with their marking invariant value from the complete model, which is 20 μM for CaM and 1.8 μM for PP1. The range of the concentration variation of $CaM.Ca_4$ is 0.00001-0.0135 μM, and the range of the concentration variation of PP1 is 0.0461-0.285 μM. Because of the values of the marking invariants, the $CaM.Ca_4$ and PP1 marks in Fig. 5.5 are never represented with tints other than blue. Instead of revealing the dynamic behavior of these enzymes, the marking invariant-based painting rule conceals it.

The weakness of the first painting rule is the strength of the second, and vice-versa.

This is why we recommend visualizing the simulation data of the biochemical models with both types of animated representation. The marking invariant-based painting rule generates displays which are suitable for an initial exploration of the global dynamic behavior. The local relative painting rule generates displays which can give supplementary detailed information once the general behavior is understood. Thus, the use of more than one viewpoint is an efficient technique for exploring different aspects of the data.

5.6 Conclusion

In this article, we used the Petri net formalism and some of its properties for the development of a new visualization method tailored to the simulation data of complex signaling pathways. We exemplified the advantages of this approach by comparing two animated representations of the simulation data of the CaMKII pathway: the first animated representation was created with a local relative painting rule, and the second with a marking invariant-based painting rule. We have also explained how to transform an ODE model into a continuous Petri net model.

To model a biological system and to visualize the simulation data with the same graphical formalism is an advantage for modelers. No time is lost in becoming familiar with a new graphical representation once a system has been mathematically modeled. A single modeling framework is used from beginning to end. It was also demonstrated that a Petri net-based visualization approach for the simulation data of a signal transduction pathway using marking invariants emphasizes its dynamic behavior. The switching mechanisms of enzymes are particularly highlighted, because of the elimination of small perturbations from the animated representations. One of the disadvantages of this visualization approach is that, for some enzymes,

the elimination of small perturbations conceals their dynamic behavior from viewers. For that reason, we believe that the local relative and the marking invariant-based visualization approaches should be used together, in order to provide more than one viewpoint for data exploration.

The visualization of the simulation data of signaling pathway models raises questions about the visualization of signal fluxes. For example, could the Petri net formalism be used to visualize the transmission of signals in a biochemical model? In the near future, we will investigate this matter, and we will work on the development of a new Petri net-based technique to characterize the dynamics of regulation motifs and signal fluxes in signal transduction models.

Acknowledgment

This work was supported in part by grant A-0141 from the Natural Sciences and Engineering Research Council of Canada and an NSERC Postgraduate Scholarship.

CHAPITRE 6

PETRI NET-BASED METHOD FOR THE ANALYSIS OF THE DYNAMICS OF SIGNAL PROPAGATION IN SIGNALING PATHWAYS

Abstract

Motivation: Cellular signaling networks[1] are dynamic systems which propagate and process information, and, ultimately, cause phenotypical responses. Understanding the circuitry of the information flow in cells is one of the keys to understanding complex cellular processes. The development of computational quantitative models is a promising avenue for attaining this goal. Not only does the analysis of the simulation data based on the concentration variations of biological compounds yields information about systemic state changes, but it is also very helpful for obtaining information about the dynamics of signal propagation.

Results: This paper introduces a new method for analyzing the dynamics of signal propagation in signaling pathways using Petri net theory. The method is demonstrated with the Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) regulation network. The results constitute temporal information about signal propagation in the network, a simplified graphical representation of the network and of the signal propagation dynamics and a characterization of some signaling routes as regulation motifs.

[1]Cet article a été soumis à la revue *Bioinformatics*. Voir la référence (Hardy et Robillard, 2007a)

6.1 Introduction

With high-throughput technologies, cell signaling research has produced a great deal of data on the signaling networks of cells. To understand how complex biological processes perform and how biological information is processed, some researchers are investigating how multiple components interact in a systemic manner. Part of this understanding will come from models developed by computational biologists, and a number of modeling and simulation tools have been developed in the last two decades specifically for biological applications. Some mathematical approaches are aimed at developing quantitative models that are numerically simu-lated (Eungdamrong et Iyengar, 2004). Other approaches are based on graph theory and ignore quantitative information to focus on the topology of the interactions between biological compounds in cellular networks (Mason et Verwoerd, 2007).

Petri net theory is used in another group of theoretical approaches to studying biological systems. This mathematical and graphical formalism was created by Petri (1962) to study systems with causal, concurrent processes. Since then, Petri nets have been further developed and applied to different types of systems, such as communication and industrial processes. Reddy et al. (1993) first applied this theory to biological systems. Meanwhile, a number of the theoretical tools and methods provided by Petri net theory have been used to study metabolic networks, signal transduction pathways and gene regulation networks (see, for example, Goss et Peccoud, 1998; Voss et al., 2003; Li et al., 2006; Matsuno et al., 2006a; Sackmann et al., 2006; Steggles et al., 2007). Because of the versatility of Petri net theory, Petri net approaches can be used to complete either qualitative or quantitative studies.

This paper presents a new Petri net-based method for the analysis of the dynamics of signal propagation in quantitative models of signal transduction networks. This

method combines invariant analysis (a structural analysis of the dynamic properties of Petri net models) and topological analysis to analyze simulation data. This method gives temporal information about signal propagation, produces a simplified graphical representation of the network and of the signal propagation dynamics, and it characterizes some signaling routes as regulation motifs. This method can be part of a unified Petri net framework used by computational biologists for modeling, simulation and simulation data analysis.

We apply this method to the Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) regulation network, which is composed of several interconnected signaling pathways. The next section briefly presents the model of the CaMKII regulation network, gives a short introduction to Petri net basics, presents the signal dynamic analysis method and provides an example of it with a model of the calmodulin and calcineurin pathways. Section 6.3 presents the results of the analysis of the signal dynamics of the CaMKII regulation network.

6.2 Methods

6.2.1 The model of the Ca^{2+}/calmodulin-dependent protein kinase II regulation pathway

The model of the CaMKII regulation pathway that is analyzed in this paper is one of the three developed for a theoretical study of long-term potentiation in the hippocampal CA1 neuron (Bhalla et Iyengar, 1999). In the cited study, several kinetic models of signaling pathways were connected to form complex networks in order to investigate their signaling processing capabilities. The experimentally observed link between persistently activated CaMKII in the postsynaptic neuron and increased

synaptic responses was one of the motivations for conducting this quantitative study. The cyclic adenosine monophosphate (cAMP), protein kinase A (PKA) and protein phosphatase 1 (PP1) signaling pathways were connected to the calmodulin (CaM), calcineurin (CaN, also known as protein phosphatase 2B, PP2B) and CaMKII signaling pathways to create a single network, because earlier studies had shown evidence that the cAMP pathway gated CaMKII signaling through the regulation of protein phosphatases (Iyengar, 1996; Blitzer et al., 1998). The kinetic parameters of the model were derived from experimental observations, and are now available in the Database of Quantitative Cellular Signaling (Sivakumaran et al., 2003). The simulation data were analyzed with concentration plots to work out the dynamic behavior of the network. The modeled CaMKII regulation pathway is shown in Figure 6.4(a).

We provide an example of the Petri net-based signal analysis method in this section using a subset of the CaMKII regulation pathway model. This subset model is composed of the CaM and CaN modules. The results section presents the conclusions of the signaling dynamic analysis of the complete model of the CaMKII network.

6.2.2 Petri net modeling

A Petri net is a directed, bipartite graph. It contains two kinds of nodes, places and transitions, connected by directed arcs. In biochemical Petri net models, places represent biological compounds, like enzymes or metabolites, and transitions represent chemical reactions. Arcs can only link one node of each kind. They indicate the causal relations between biological compounds and chemical reactions. These relations are weighted. In biochemical Petri net models, arcs are labeled with weights corresponding to the stoichiometric parameters of the reaction equations. Places can contain dynamic objects, called tokens, which represent a certain quantity of a chemical compound. The mark of a place corresponds to its number of tokens, thus

to the number of molecules of the biological compounds.

Tokens are produced and consumed through the firing of transitions. A transition is able to fire if the marks of all its input places satisfy the token amount required by the weights of the arcs connecting these input places with the considered transition. This is the precondition of a transition. If a transition fires, it means that the chemical reaction has occurred. The weights of the outgoing arcs indicate the token amount that is added to the output places of the considered transition. This is the postcondition of a transition.

The distribution of tokens in the Petri net places is called the marking of the net and indicates the state of the system. For a biochemical model, the marking is the distribution of compounds. The initial marking is the marking of the start state.

The theoretical description of Petri nets given above is the description of the original theory (Petri, 1962). These nets are also called place/transition nets. In these, places contain a discrete quantity of tokens (marks are integers) and the firings of transitions are discrete events. Different extensions have been added to the original theory to augment the modeling possibilities of this formalism. In this paper, we also use elements with continuous properties. Continuous places contain a continuous quantity; and the marks of continuous places are real numbers. The firing of continuous transitions is a continuous flow, and continuous transitions have a specified speed. A kinetics model can be modeled with continuous Petri net elements: the mark of continuous places represents the concentration of chemical compounds, and the marking-dependant speed of the continuous transitions can correspond to the reaction-rate equations of chemical reactions. For more detailed information on using Petri nets for biochemical modeling, see Hardy et Robillard (2004).

The model of the CaM and CaN modules of the CaMKII regulation pathway, shown

in Figure 6.1, is an example of a biochemical Petri net model incorporating kinetics parameters. In this model, every place and transition is continuous. Tables 6.1 and 6.2 present the specifications of the Petri net model of the CaM and CaN modules. The continuous places are drawn as double circles, and the continuous transitions as white rectangles. The input places of a transition indicate that these biological compounds are the substrates of the chemical reaction, and the output places its products. This continuous model is mathematically equivalent to a model of ordinary differential equations, but with a formal graphical representation and numerous theoretical tools for model analysis.

Tableau 6.1: Specifications of the places of the Petri net model of the calmodulin kinase pathway

Name	Substance	Variable
CaM	Calmodulin	m_0
CaM-Ca_2	Calmodulin-Ca_2	m_1
CaM-Ca_3	Calmodulin-Ca_3	m_2
CaM-Ca_4	Calmodulin-Ca_4	m_3
Ca	Calcium	m_4
CaM-Ng	Calmodulin-Neurogranin complex	m_5
Ng	Neurogranin	m_6
Ng*	phosphorylated Neurogranin	m_7
PKC	Protein kinase C	m_8
$c_2 2$	CaM-Ng-PKC complex	m_9
$c_2 1$	Ng-PKC complex	m_{10}
Ca_4-PP2B	Calcineurin-Ca_4	n_0
Ca_2-PP2B	Calcineurin-Ca_2	n_1
PP2B	Calcineurin	n_2
Ca_4-PP2B-CaM-Ca_4	Calcineurin-Ca_4-Calmodulin-Ca_4 complex	n_4
Ca_4-PP2B-CaM-Ca_3	Calcineurin-Ca_4-Calmodulin-Ca_3 complex	n_5
Ca_4-PP2B-CaM-Ca_2	Calcineurin-Ca_4-Calmodulin-Ca_2 complex	n_6
$c_2 5$	Ca_4-PP2B-CaM-Ca_4-Ng* complex	n_7
$c_2 4$	Ca_4-PP2B-CaM-Ca_3-Ng* complex	n_8
$c_2 3$	Ca_4-PP2B-CaM-Ca_2-Ng* complex	n_9

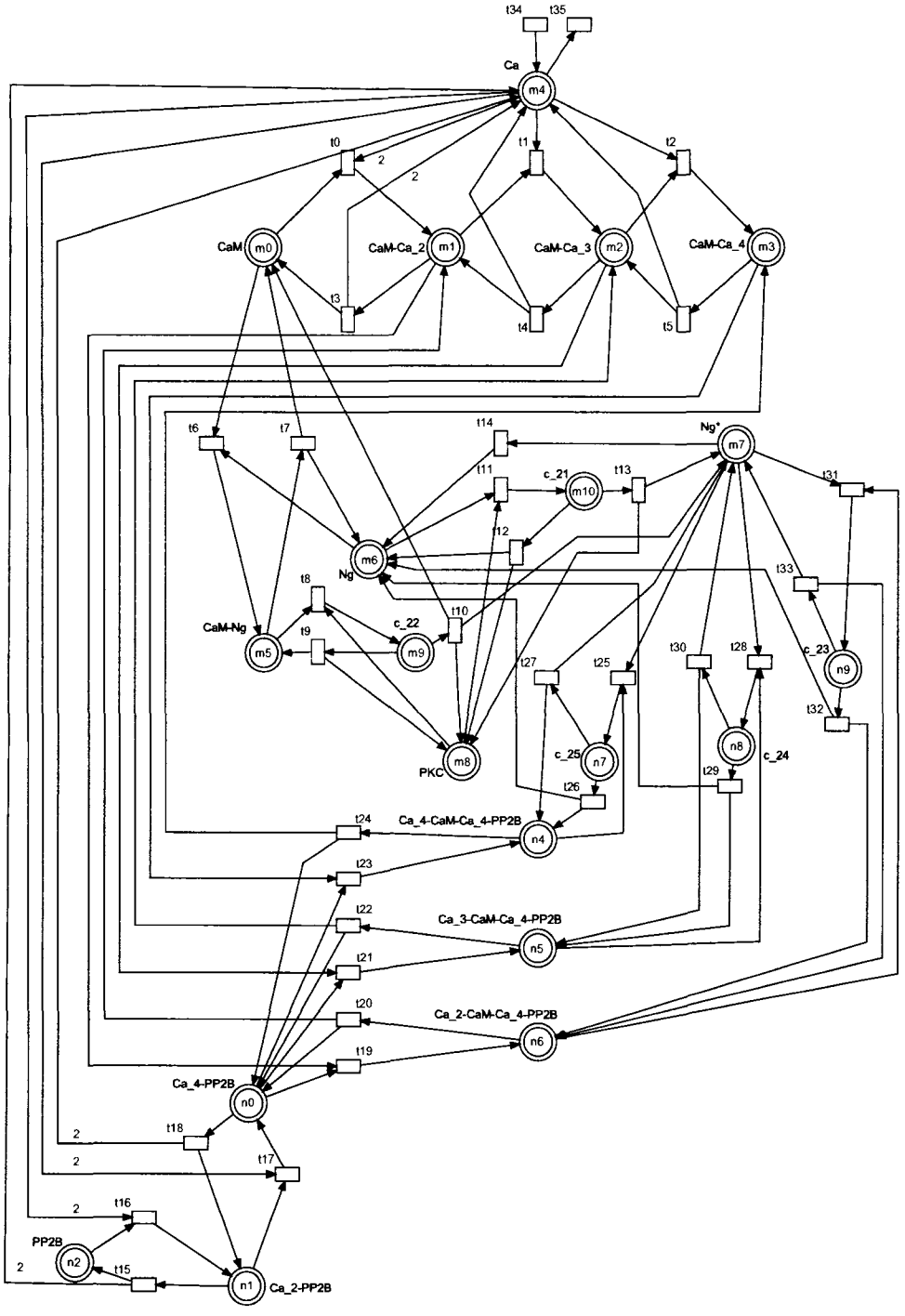


Figure 6.1: Petri net model of the calmodulin and calcineurin modules of the CaMKII regulation pathway

Tableau 6.2: Specifications of the transitions and arcs of the Petri net model of the camodulin pathway

Name	Speed	From		To	
		variable	factor	variable	factor
t_0	$72 \times m_0 \times {m_4}^2$	m_0	1	m_1	1
		m_4	2		
t_1	$3.6 \times m_1 \times m_4$	m_1	1	m_2	1
		m_4	1		
t_2	$0.465 \times m_2 \times m_4$	m_2	1	m_3	1
		m_4	1		
t_3	$72 \times m_1$	m_1	1	m_0	1
				m_4	2
t_4	$10 \times m_2$	m_2	1	m_1	1
				m_4	1
t_5	$10 \times m_3$	m_3	1	m_2	1
				m_4	1
t_6	$0.3 \times m_0 \times m_6$	m_0	1	m_5	1
		m_6	1		
t_7	m_5	m_5	1	m_0	1
				m_6	1
t_8	$0.0612 \times m_5 \times m_8$	m_5	1	m_9	1
		m_8	1		
t_9	$1.4 \times m_9$	m_9	1	m_5	1
				m_8	1
t_{10}	$0.35 \times m_9$	m_9	1	m_0	1
				m_7	1
				m_8	1
t_{11}	$0.101 \times m_6 \times m_8$	m_6	1	m_{10}	1
		m_8	1		
t_{12}	$2.32 \times m_{10}$	m_{10}	1	m_6	1
				m_8	1
t_{13}	$0.58 \times m_{10}$	m_{10}	1	m_7	1
				m_8	1
t_{14}	$0.005 \times m_7$	m_7	1	m_6	1
t_{15}	n_1	n_1	1	n_2	1
				m_4	2
t_{16}	$10008 \times n_2 \times {m_4}^2$	n_2	1	n_1	1
		m_4	2		

Continued on next page

Tableau 6.2 – continued from previous page

Name	Speed	From variable	From factor	To variable	To factor
t_{17}	$3.6 \times n_1 \times {m_4}^2$	n_1	1	n_0	1
		m_4	2		
t_{18}	n_0	n_0	1	n_1	1
				m_4	2
t_{19}	$0.24 \times n_0 \times m_1$	n_0	1	n_6	1
		m_1	1		
t_{20}	n_6	n_6	1	n_0	1
				m_1	1
t_{21}	$2.238 \times n_0 \times m_2$	n_0	1	n_5	1
		m_2	1		
t_{22}	n_5	n_5	1	n_0	1
				m_2	1
t_{23}	$600 \times n_0 \times m_3$	n_0	1	n_4	1
		m_3	1		
t_{24}	n_4	n_4	1	n_0	1
				m_3	1
t_{25}	$0.3346 \times n_4 \times m_7$	n_4	1	n_7	1
		m_7	1		
t_{26}	$0.67 \times n_7$	n_7	1	n_4	1
				m_6	1
t_{27}	$2.68 \times n_7$	n_7	1	n_4	1
				m_7	1
t_{28}	$0.3346 \times n_5 \times m_7$	n_5	1	n_8	1
		m_7	1		
t_{29}	$0.67 \times n_8$	n_8	1	n_5	1
				m_6	1
t_{30}	$2.68 \times n_8$	n_8	1	n_5	1
				m_7	1
t_{31}	$0.3346 \times n_6 \times m_7$	n_6	1	n_9	1
		m_7	1		
t_{32}	$0.67 \times n_9$	n_9	1	n_6	1
				m_6	1
t_{33}	$2.68 \times n_9$	n_9	1	n_6	1
				m_7	1
t_{34}–t_{35}	These transitions are simplifications for the calcium pump module.				

6.2.3 Invariant analysis

Invariants are dynamic properties of Petri nets, and are obtained from a structural computation. In biochemical models, they usually have a mass conservation meaning or represent a cyclic pathway. This subsection presents the theoretical background on Petri net invariants and the result of the invariant analysis of the CaM-CaN model.

The structure of a Petri net model, i.e. the arrangement of places, transitions and arcs, can be expressed in linear algebra by a two-dimensional matrix. This is the incidence matrix $\mathbf{W}$. One dimension of the incidence matrix is the number of places of the model and the other dimension is the number of transitions. In place-transition nets, each element w_{ij} of the matrix indicates the token change at place i after the firing of transition j. In continuous nets, each element w_{ij} of the matrix indicates the flow from transition j to place i. The element w_{ij} is negative if the flow direction is from a place to a transition. The flow is the product of the speed of a transition and the weight of the arc between the place and the transition. From the incidence matrix, it is possible to determine structural properties of Petri nets like invariants. Among all the attainable states of a model, called the reachable markings of a Petri net model, some quantities do not change, even when transitions are fired. This first type of invariant property is the marking invariant (p-invariant). Every p-invariant of a Petri net model is a positive vector $\mathbf{x}$ that is a solution of the following equation:

$$\mathbf{x} \cdot \mathbf{W} = \mathbf{0} \tag{6.1}$$

A p-invariant characterizes a conservation component of the model. A conservation

component is a set of places over which the weighted sum of the tokens is constant for every reachable marking. Relationships between p-invariants and biochemical modeling concepts have already been discussed. P-invariants express conservation relations of metabolites in metabolic pathway models (Voss et al., 2003; Zevedei-Oancea et Schuster, 2003). In signaling pathway models, p-invariants can represent a different kind of conservation relation (Sackmann et al., 2006). Biochemical processes in signaling pathways are performed by enzymes that change state to transmit a signal. The total concentration of all forms of an enzyme is modeled as a constant quantity. This quantity is a marking invariant of a Petri net model. Thus, the p-invariants and their associated conservation components identify all the places representing a specific form of an enzyme. The four conservation components of the CaM-CaN model are listed in Table 6.3.

Tableau 6.3: Conservation components and marking invariants of the Petri net model of the calmodulin and calcineurin modules of the CaMKII regulation pathway

Enzyme	Conservation component Marking invariant	
Calmodulin	$CC_1 =$	{CaM, CaM-Ca_2, CaM-Ca_3, CaM-Ca_4, CaM-Ng, c_{22}, Ca_4-PP2B-CaM-Ca_2, Ca_4-PP2B-CaM-Ca_3, Ca_4-PP2B-CaM-Ca_4, c_{23}, c_{24}, c_{25}}
	$N_{CC_1} =$	$m_0 + m_1 + m_2 + m_3 + m_5 + m_9 + n_4 + n_5 + n_6 + n_7 + n_8 + n_9$
Neurogranin	$CC_2 =$	{Ng, Ng*, CaM-Ng, c_{21}, c_{22}, c_{23}, c_{24}, c_{25}}
	$N_{CC_2} =$	$m_5 + m_6 + m_7 + m_9 + m_{10} + n_7 + n_8 + n_9$
PKC	$CC_3 =$	{PKC, c_{21}, c_{22}}
	$N_{CC_3} =$	$m_8 + m_9 + m_{10}$
Calcineurin	$CC_4 =$	{PP2B, Ca_2-PP2B, Ca_4-PP2B, Ca_4-PP2B-CaM-Ca_2 Ca_4-PP2B-CaM-Ca_3, Ca_4-PP2B-CaM-Ca_4, c_{23}, c_{24}, c_{25}}
	$N_{CC_4} =$	$n_0 + n_1 + n_2 + n_4 + n_5 + n_6 + n_7 + n_8 + n_9$

Among all the possible firing sequences of a model, some can repeat themselves,

creating a cycle of successive states. This second type of invariant is the firing invariant (t-invariant). The t-invariant of a Petri net model is a positive vector $\mathbf{y}$ that is a solution of the following equation:

$$\mathbf{W} \cdot \mathbf{y} = \mathbf{0} \tag{6.2}$$

A t-invariant characterizes a repetitive component of a model. A repetitive component is a set of transitions causing a return to a previous state of a model. In other words, all the firings of the transitions of a repetitive component together have a null effect on the marking of the model. In Petri net models of signaling networks, the t-invariant property has been used for the identification methods of functional units (Sackmann et al., 2006) and of transduction activation components (Li et al., 2006).

For continuous Petri nets, the t-invariant definition has to be adapted. The firing of continuous transitions is a continuous flow and not a discrete event. Because a continuous flow cannot be repeated, the concept of a repetitive component makes no sense. In its place, we suggest using the concept of a steady component. From an unstable state, a continuous Petri net model evolves to a steady state. In a stable steady state and without any external perturbations, the marking of the model no longer changes. In this stable state, a subset of the solutions of equation 6.2 lists the steady components of a model, which are the sets of transitions for which the total flow is null. To be part of this subset of solutions, these special t-invariants must be vectors composed only of 0 and 1 (a flow can either exist or not, it cannot be multiplied). Hence, the non-zero elements of these t-invariants correspond to the transitions composing a steady component. In the case of a continuous Petri net with marking-dependent speeds, equation 6.2 must be solved with the incidence matrix filled with the steady state flow values in order to find the steady components. Linear

combinations of the t-invariants are also solutions of equation 6.2; however, only the vectors that do not include other solutions are searched. These are the minimal steady components. The minimal steady components of the CaM-CaN model are listed in Table 6.4.

Tableau 6.4: Steady components of the Petri net model of the calmodulin and calcineurin modules of the CaMKII regulation pathway

$SC_1 = \{t_0, t_3\}$	$SC_6 = \{t_{17}, t_{18}\}$
$SC_2 = \{t_1, t_4\}$	$SC_7 = \{t_{19}, t_{20}\}$
$SC_3 = \{t_2, t_5\}$	$SC_8 = \{t_{21}, t_{22}\}$
$SC_4 = \{t_{15}, t_{16}\}$	$SC_9 = \{t_{23}, t_{24}\}$
$SC_5 = \{t_6, t_7, t_8, t_9, t_{10}, t_{11}, t_{12}, t_{13}, t_{14}, t_{25}, t_{26}, t_{27}, t_{28}, t_{29}, t_{30}, t_{31}, t_{32}, t_{33}\}$	

6.2.4 Identification of the information flux segments and nodes of a signaling network

To detect signal propagation through a signaling network, places and transitions must be identified to serve as measuring points. These are called information flux nodes and segments. Information flux segments are the portions of a pathway composed of places and transitions that might transmit or route a signal in the network. Information flux segments are positioned between two groups of one or more places described as information flux nodes. The flow circulating on an information flux segment is bidirectional. At steady state, the flow in each direction, on every possible pathway linking the nodes, is equal. This is the equal flow condition that will be used in Algorithm 6.1 to identify information flux segments and nodes. This condition is true only for the places and transitions of information flux segments. As we will see in the next subsection, a comparison of the flow in one direction with the flow in the opposite direction on a segment reveals the direction of a signal. The steady

components of a Petri net model provide the initial subnets used in the search for the information flux segments of the signaling network.

The search for the information flux nodes and segments can be accomplished with the following algorithm:

Algorithm 6.1. Search for the information flux segments and nodes of a steady component.

Step 1. Let SC be a given steady component and SPN the subnet of a given Petri net with only the transitions of SC and all the places directly linked to these transitions. Generate B, the set of all the pairs of places from SPN. Initialize the set of information flux nodes $IFN \leftarrow \varnothing$. IFN is a set containing pairs of sets of places.

Step 2. If $B \neq \varnothing$, then pull the pair $\langle \{b_1\}, \{b_2\} \rangle$ from B and perform steps i) to v), otherwise go to 3).

Step 2i. Let D_1 be the set of all the paths in SPN from b_1 to b_2 and D_2 be the set of all the paths in SPN from b_2 to b_1. A path is a succession of places and transitions linked by arcs. The paths with a transition having an outgoing arc towards the initial place of the path are rejected. Initialize cumulative flow values $F_{1_{in}} = F_{1_{out}} = F_{2_{in}} = F_{2_{out}} = 0$. If D_1 or D_2 is empty, stop.

Step 2ii. Add to $F_{1_{in}}$ the value of the speeds of every transition directly linked to b_1 in the paths from D_2 (i.e. every transition t_y that is part of the last place-transition-place triplet $p_x \rightarrow t_y \rightarrow b_1$ of a path). For triplets where p_x is not b_2 (i.e. for indirect paths), withdraw from $F_{1_{in}}$ the values of the speed of any transition t_z in SPN in a place-transition-place triplet $b_1 \rightarrow t_z \rightarrow p_x$ where place p_x is the same as one of the triplets of the first category.

Step 2iii. Add to $F_{2_{out}}$ the value of the speeds of every transition directly linked to b_1 in the paths from D_1 (i.e. every transition t_m that is part of the last place-transition-place triplet $b_1 \rightarrow t_m \rightarrow p_y$ of a path). For triplets where p_y is not b_2 (i.e. for indirect paths), withdraw from $F_{2_{out}}$ the values of the speed of any transition t_n in SPN in a place-transition-place triplet $p_y \rightarrow t_n \rightarrow b_1$ where place p_y is the same as one of the triplets of the first category.

Step 2iv. Add to $F_{2_{in}}$ the value of the speeds of every transition directly linked to b_2 in the paths from D_1 (i.e. every transition t_y that is part of the last place-transition-place triplet $p_x \rightarrow t_y \rightarrow b_2$ of a path). For triplets where p_x is not b_1 (i.e. for indirect paths), withdraw from $F_{2_{in}}$ the values of the speed of any transition t_z in SPN in a place-transition-place triplet $b_2 \rightarrow t_z \rightarrow p_x$ where place p_x is the same as one of the triplets of the first category.

Step 2v. Add to $F_{2_{out}}$ the value of the speeds of every transition directly linked to b_2 in the paths from D_2 (i.e. every transition t_m that is part of the last place-transition-place triplet $b_2 \rightarrow t_m \rightarrow p_y$ of a path). For triplets where p_y is not b_1 (i.e. for indirect paths), withdraw from $F_{2_{out}}$ the values of the speed of any transition t_n in SPN in a place-transition-place triplet $p_y \rightarrow t_n \rightarrow b_2$ where place p_y is the same as one of the triplets of the first category.

Step 2vi. If $F_{1_{in}} = F_{2_{out}}$, $F_{1_{out}} = F_{2_{in}}$ and these values are not null, then add $\langle \{b_1\}, \{b_2\} \rangle$ to IFN. Go to 2). *Step 3.* If two pairs $\langle B_x, B_y \rangle$ in IFN share an identical element, determine whether or not the two pairs also share the same transitions. If they do, form the union of the two different sets of places and replace the two pairs in IFN by the new unified pair.

□

In step 1 of the above algorithm, the main operation is the construction of the set B of all possible pairs of places. These places are part of the subnet SPN corresponding to the steady component SC. Also, the set IFN that will contain the information flux nodes identified by the algorithm is initialized. In step 2, each pair of sets of places $\langle\{b_1\},\{b_2\}\rangle$ of set B is tested with the equal flow condition to determine whether or not they are part of an information flux segment. In step 2i, the sets D_1 and D_2 are constructed to contain every possible path linking places b_1 and b_2 in SPN. In steps 2ii to 2v, four cumulative flow values are computed; these values are the flows going in and out of places b_1 and b_2. In step 2vi, the cumulative flow values are compared to verify the equal flow condition. If it is verified, the pair is added to IFN. In step 3, the pairs of sets of places in INF sharing an identical set are merged into a single pair only if they are part of the same segment. Table 6.5 shows the results of the search for the informa-tion flux nodes and segments of the steady components of the CaM-CaN model.

6.2.5 Analysis of the information flux of a signaling network

Once the information flux segments and nodes have been identified, these measuring points can be used, together with the topology of the network, to characterize the dynamics of a signal. At steady state, the difference between the flows in each direction on an information flux segment is zero. However, after a perturbation, this difference indicates the direction of the propagation of a signal, which can then be followed throughout the structure of the model. With formal definitions about the pathway structure of regulation motifs, such as positive and negative feedback loops, different types of signal routes can also be identified.

Here, we complete the introduction of the Petri net-based signal analysis method by presenting an analysis of the information flux caused by the inflow of calcium

Tableau 6.5: Information flux segments of the Petri net model of the calmodulin and calcineurin modules of the CaMKII regulation pathway

Steady component	Segment	Nodes and speeds of the input and output transitions of the segment
SC_1	S_1	<{Ca, CaM},{CaM-Ca_2}> $F_{1_{in}} = speed_{t_0}$, $F_{1_{out}} = speed_{t_3}$ $F_{2_{in}} = speed_{t_3}$, $F_{2_{out}} = speed_{t_0}$
SC_2	S_2	<{Ca, CaM-Ca_2},{CaM-Ca_3}> $F_{1_{in}} = speed_{t_1}$, $F_{1_{out}} = speed_{t_4}$ $F_{2_{in}} = speed_{t_4}$, $F_{2_{out}} = speed_{t_1}$
SC_3	S_3	<{Ca, CaM-Ca_3},{CaM-Ca_4}> $F_{1_{in}} = speed_{t_2}$, $F_{1_{out}} = speed_{t_5}$ $F_{2_{in}} = speed_{t_5}$, $F_{2_{out}} = speed_{t_2}$
SC_4	S_4	<{Ca, PP2B},{Ca_2-PP2B}> $F_{1_{in}} = speed_{t_{16}}$, $F_{1_{out}} = speed_{t_{15}}$ $F_{2_{in}} = speed_{t_{15}}$, $F_{2_{out}} = speed_{t_{16}}$
SC_5	S_5	<{CaM},{CaM-Ng}> $F_{1_{in}} = speed_{t_6}$, $F_{1_{out}} = speed_{t_9} - speed_{t_8}$ $F_{2_{in}} = speed_{t_7} + speed_{t_{10}}$, $F_{2_{out}} = speed_{t_6}$
	S_6	<{Ng},{Ng*}> $F_{1_{in}} = speed_{t_{13}} + speed_{t_{10}}$, $F_{1_{out}} = speed_{t_{14}} + speed_{t_{25}} - speed_{t_{27}} + speed_{t_{28}} - speed_{t_{30}} + speed_{t_{31}} - speed_{t_{33}}$ $F_{2_{in}} = speed_{t_{14}} + speed_{t_{26}} + speed_{t_{29}} + speed_{t_{32}}$, $F_{2_{out}} = speed_{t_{11}} - speed_{t_{12}} + speed_{t_6} - speed_{t_7}$
SC_6	S_7	<{Ca, Ca_2-PP2B},{Ca_4-PP2B}> $F_{1_{in}} = speed_{t_{17}}$, $F_{1_{out}} = speed_{t_{18}}$ $F_{2_{in}} = speed_{t_{18}}$, $F_{2_{out}} = speed_{t_{17}}$
SC_7	S_8	<{CaM-Ca_2, Ca_4-PP2B},{Ca_4-PP2B-CaM-Ca_2}> $F_{1_{in}} = speed_{t_{19}}$, $F_{1_{out}} = speed_{t_{20}}$ $F_{2_{in}} = speed_{t_{20}}$, $F_{2_{out}} = speed_{t_{19}}$
SC_8	S_9	<{CaM-Ca_3, Ca_4-PP2B},{Ca_4-PP2B-CaM-Ca_3}> $F_{1_{in}} = speed_{t_{21}}$, $F_{1_{out}} = speed_{t_{22}}$ $F_{2_{in}} = speed_{t_{22}}$, $F_{2_{out}} = speed_{t_{21}}$
SC_9	S_{10}	<{CaM-Ca_4, Ca_4-PP2B},{Ca_4-PP2B-CaM-Ca_4}> $F_{1_{in}} = speed_{t_{23}}$, $F_{1_{out}} = speed_{t_{24}}$ $F_{2_{in}} = speed_{t_{24}}$, $F_{2_{out}} = speed_{t_{23}}$

in the CaM-CaN model. The CaM-CaN pathway was modeled and simulated with the software Genomic Object Net (GON, commercially known as Cell Illustrator) (Nagasaki et al., 2003). The simulation period lasts 10 seconds. The calcium inflow occurs at the first second and lasts 1 second. This inflow perturbs the CaM-CaN model and a propagation of a signal ensues.

The speed differential of an information flux segment is defined as the difference between its $F_{1_{in}}$ and $F_{2_{in}}$ values. Figure 6.2 presents the normalized speed differential plots of the information flux segments computed with the data generated by the simulation of the CaM-CaN model. The plots show an increase in calcium binding with calmodulin and calcineurin, an increase in the dephosphorylation of neurogranin and a decrease followed by an increase in the binding of calmodulin with neurogranin. The speed differential plots give a first indication of the consequences of the calcium inflow, but with Petri net properties and the topological data of the network, a deeper analysis of the dynamics of the signal propagation can be performed and a graphical representation of the signal propagation can be generated.

The initial data needed for the signal analysis of a Petri net model is the source of the signal and the duration of the stimulation. For the CaM-CaN model, the Ca place is the source and the stimulation lasts 1 second. At every simulation step of the stimulation period, the differential speeds of the information flux segments linked to the source of the signal are checked. For the CaM-CaN example, the speed differentials of segments S_1, S_2, S_3, S_4 and S_7 are verified, because the Ca place is part of these segments. When the relative difference between $F_{1_{in}}$ and $F_{2_{in}}$, or vice-versa, is above an arbitrary threshold (10% in this example), the segment is considered to be propagating a signal. When this occurs, the speed differentials of other segments connected to the nodes of the activated segment are examined to determine whether or not the propagation continues. In our example, following signaling activation of

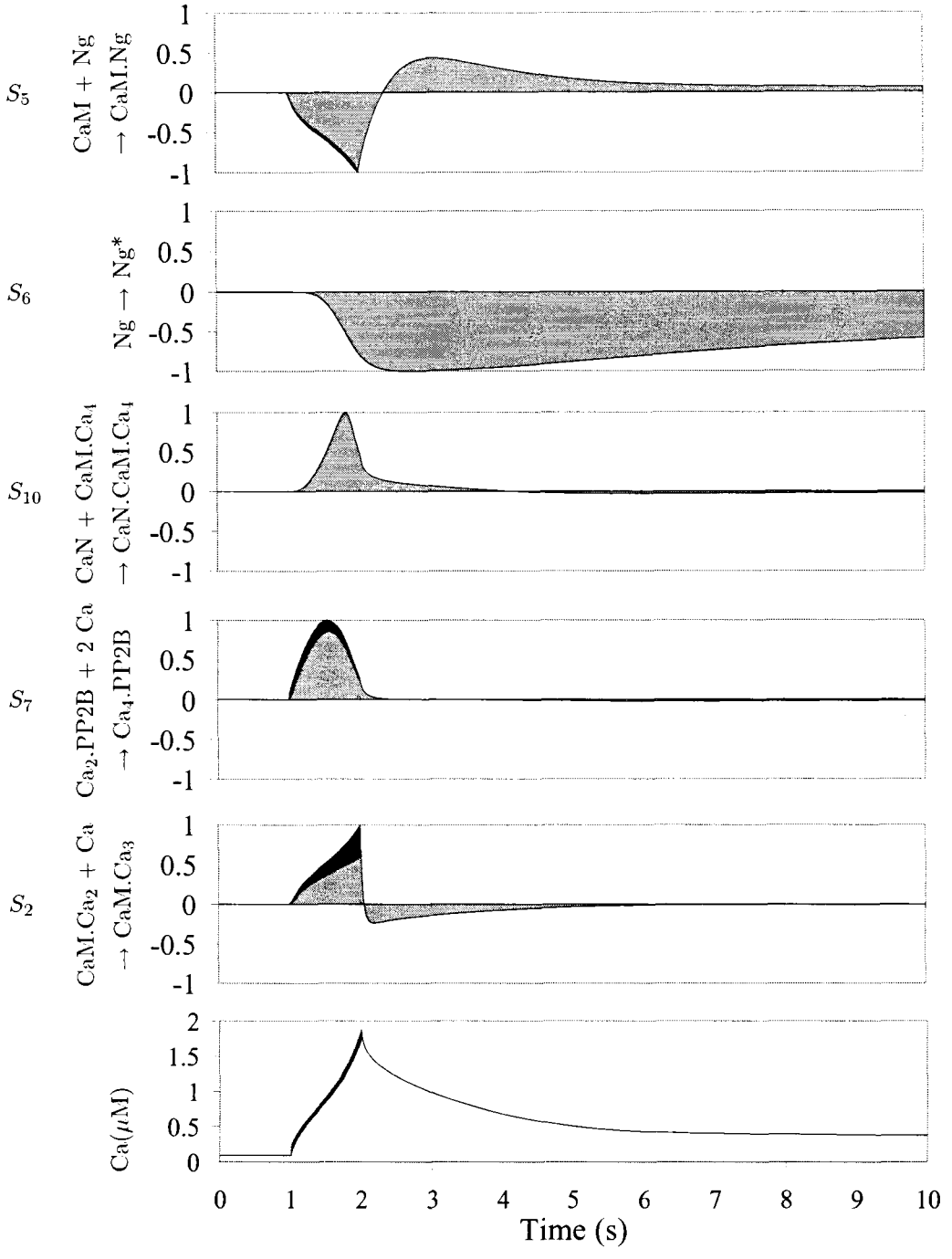
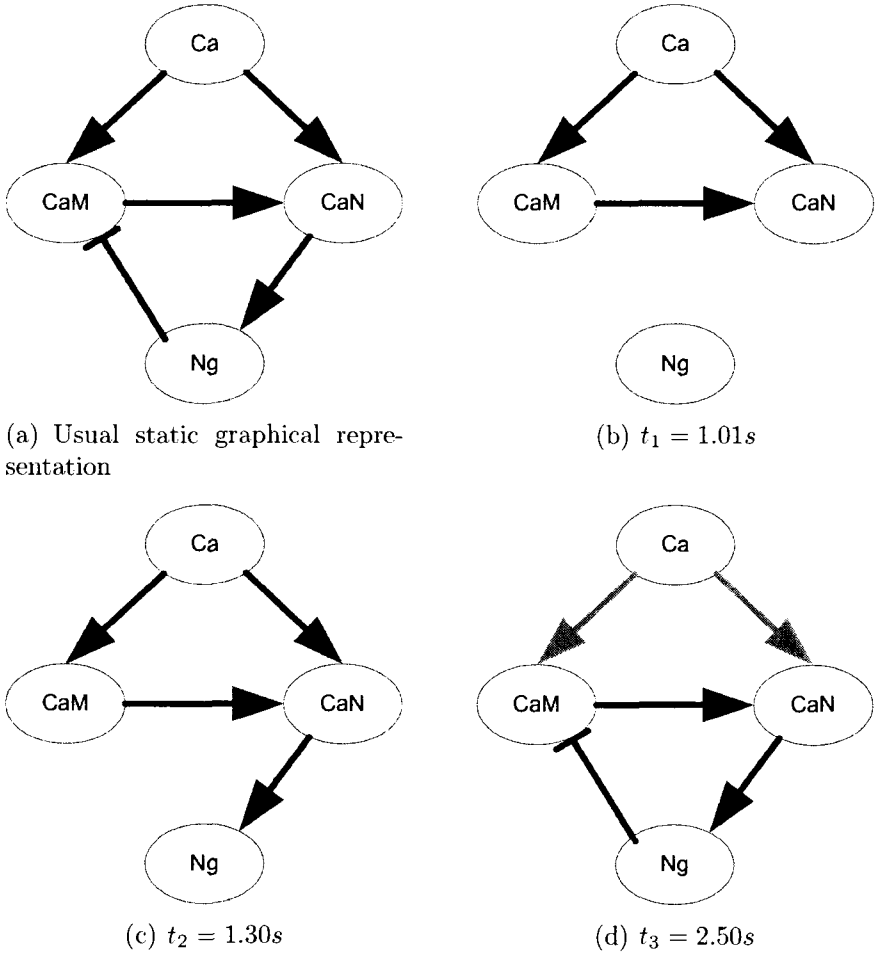


Figure 6.2: Concentration plot of calcium and normalized speed differential plots for some information flux segments of the Petri net model of the calmodulin and calcineurin modules of the CaMKII regulation pathway

segments S_1, S_2, S_3, S_4 and S_7 due to the calcium inflow, the speed differentials of segments S_8, S_9 and S_{10} are evaluated. This assessment of the speed differentials of the connected in-formation flux segments continues as the signal is propagated. In our example, following signaling activation of segments S_8, S_9 and S_{10}, the activation of segment S_6 and eventually of segment S_5, is observed.

More than just observing successive signaling activations, the Petri net-based signal analysis method can yield temporal information about these activations and produce a simplified graphical representation of the network and of the signal propagation dynamics, as well as characterizing some signaling routes as regulation motifs. These three features of the method are shown in Figure 6.3. In this figure, the precise moments of the signaling activation of the information flux segments are indicated. Also, instead of displaying the detailed model of the network, this figure shows a simpler structure. The results of the p-invariant analysis of the model provide the information to generate this view: the nodes of this abridged graph correspond to the conservation components of the Petri net model. Finally, the definitions of the graph structure of two regulation motifs, positive and negative feedback loops, are used to detect special signaling pathways. A positive feedback loop is identified when a signal is propagated to an information flux segment that was previously activated by the same signal. The signal then promotes the activation of the segment in the same direction in which it was initially activated. In other words, the signal loops back to sustain the signaling activation of an already activated segment. A negative feedback loop is identified in two situations. The first situation is when a signal is propagated to a segment that was previously activated by the same signal, and the signaling flux of the segment is then reversed. The second situation is when a signal is propagated to a segment in such a way that it competes with a connected segment already activated by this signal. In both situations, the signal loops back to hinder the signaling activation of an already activated segment. A negative feedback loop

was identified in the CaN-CaM model and is shown in Figure 6.3(d). The activation of the segment S_6, (CaM + Ng $\rightarrow$ CaM-Ng), competes with the activated segment S_1 (CaM + 2 Ca $\rightarrow$ CaM-Ca_2) because they both share the CaM place. This process is known as the inhibition of calmodulin by neurogranin. In our example, this inhibition occurs at 2.5 seconds, 0.5 second after the end of the calcium inflow.



(a) Usual static graphical representation

(b) $t_1 = 1.01s$

(c) $t_2 = 1.30s$

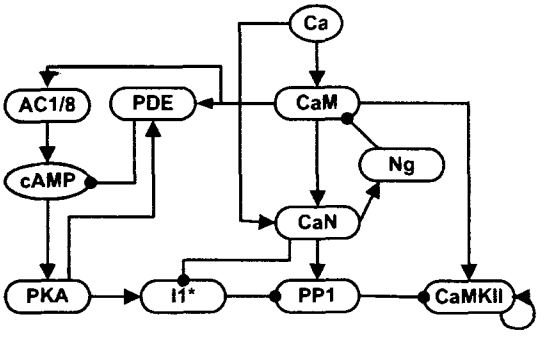
(d) $t_3 = 2.50s$

Figure 6.3: Graphical representations of the propagation of a signal in the calmodulin and calcineurin modules of the CaMKII regulation pathway

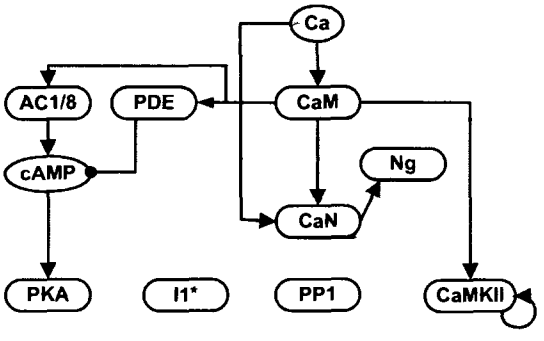
6.3 Results

This section presents the results of the signal dynamic analysis of the CaMKII regulation pathway with a Petri net-based method. The Petri net model of the CaMKII regulation pathway was simulated with GON for 800 seconds. A calcium inflow reproducing a tetanic stimulation is induced at the sixtieth second and lasts 1 second. A p-invariant analysis of the model resulted in the identification of 15 conservation components. Also, using the concentration values of the steady state of the model, 30 information flux segments have been identified. Figure 6.4 shows the dynamics of the signal propagation in the network. This figure is a simplified graphical representation of the Petri net model of the network produced with the p-invariant analysis data. It gives temporal information about the signaling activation of the information flux segments of the network. The initial and final states are not shown because they are steady states in which no signal is propagated (no arrows displayed). Figure 6.4 also shows the activity of three feedback loops. The first loop, the negative feedback loop leading to the inhibition of calmodulin by neurogranin, has already been discussed and is shown in an active state in Figure 6.4(c). Another negative feedback loop has been detected: the increase of the degradation of cAMP into AMP because of the activation of phosphodiesterase (PDE) by PKA. This regulation motif is shown in an active state in Figure 6.4(c) and 6.4(d). The third loop is a positive feedback loop. It corresponds to the autophosphorylation cycle of CaMKII. This motif is active immediately following calcium stimulation at 61 seconds and lasts for more than 500 seconds. This loop is represented as a circular arrow on the CaMKII conservation component in every subfigure of Figure 6.4. The details of this motif and the cyclic activation of two information flux segments are hidden, because both seg-ments are enclosed in the CaMKII conservation component.

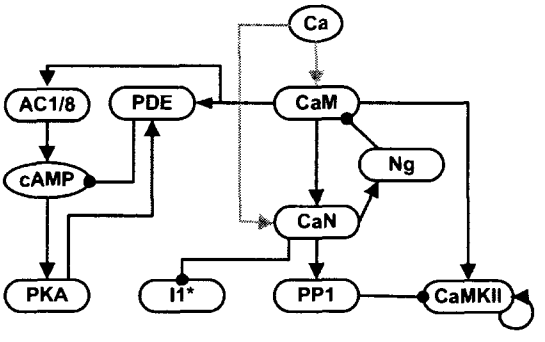
Figure 6.4 illustrates key moments of the signal propagation due to the calcium



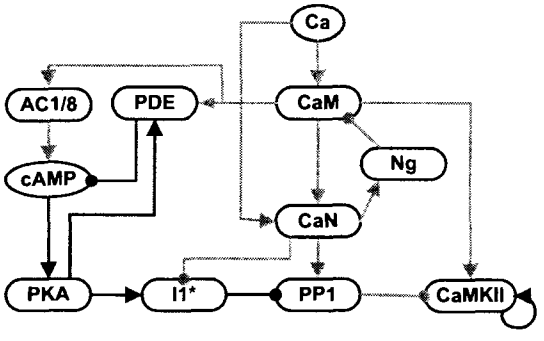
(a) Usual static graphical representation



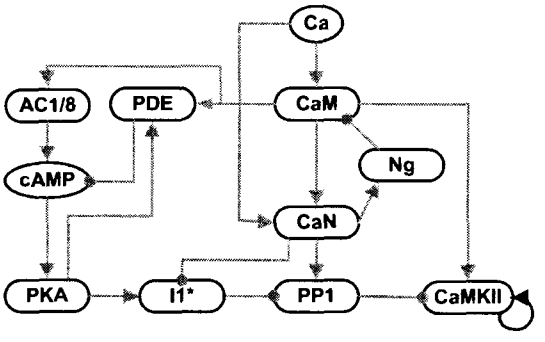
(b) $t_1 = 61s$



(c) $t_2 = 63s$



(d) $t_3 = 89s$



(e) $t_4 = 450s$

Figure 6.4: Graphical representations of the propagation of a signal in the CaMKII regulation network

inflow in the CaMKII regulation network. Figure 6.4(b) shows the propagation of the signal immediately after the end of the stimulus. At 61 seconds, the CaM activation initiates the production of cAMP through the activation of adenylyl cyclases 1 and 8 (AC1/8), which in turn immediately starts activating PKA. CaM activation also starts the autophosphorylation cycle of CaMKII and with CaN, dephosphorylates Ng. Two seconds later, as shown in Figure 6.4(c), the Ng and PKA negative feedback loops are active. Also, CaN has activated PP1 and the inhibited Inhibitor-1 protein (I1), which in turn has a deactivation effect on CaMKII. Figure 6.4(d) shows that the PP1 activation by CaN is overcome by the activation of PKA, because it causes the phosphorylation of I1 and therefore the inhibition of PP1. Figure 6.4(e) shows that the positive feedback loop of the CaMKII is still active, for a long period of time after the stimulus has been withdrawn.

6.4 Conclusion

This paper describes a method derived from Petri net theory to analyze the dynamics of signal propagation in a signaling network. The Petri net formalism is useful for depicting biochemical networks and performing simulations, but, more importantly, it provides techniques for the validation of the model and for the analysis of the simulation data. Because of its theoretical richness, Petri net theory offers a unified framework to computational biologists. The techniques that have been exploited in this paper are the calculation of p-invariants (conservation components) and t-invariants (steady components) of a system. The method has been applied to the Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) regulation network, an important pathway in the synaptic plasticity of neurons, and resulted in a new portrayal of the dynamics of the signal propagation in the network. We believe that this Petri net-based method will be of great value to computational biologists who

need to rapidly interpret the simulation data of their signaling network models and gain insights into the systemic dynamics of signal processing components.

Understanding the dynamics of signal propagation in signaling network models is usually achieved with the analysis of the concentration plots of the simulation data. However, studying the concentration variations of a model yields more information about the changes of state of the model components than the dynamics of the signal propagation, and, as a result, the signal propagation dynamics have to be deduced from the data. The Petri net-based method presented in this paper is a new analysis tool, which highlights, in a comprehensive way, the propagation of a signal in cellular network models and facilitates a system-level understanding.

The study of this method has been limited to models with a single signal source. Its application to the signaling dynamics of networks with simultaneous signals from multiple sources has yet to be performed.

New definitions of the graph structure of regulation motifs such as scaffolds and bifans can be added to this method. This will enrich its analytical possibilities and enhance its ability to help in deciphering the signal processing functions of cellular networks. Also, in the future, this method will be adapted to the particularities of hybrid Petri nets.

Acknowledgment

This work was supported in part by grant A-0141 from the Natural Sciences and Engineering Research Council of Canada and an NSERC Postgraduate Scholarship.

CHAPITRE 7

DISCUSSION GÉNÉRALE

Une discussion sur l'objet de cette thèse multidisciplinaire ouvre la porte à de nombreux sujets. Au confluent de la biologie, de l'informatique et du génie, j'ai choisi de me tourner vers deux pistes de réflexion fort différentes l'une de l'autre. Premièrement, je discuterai des aspects techniques de la simulation de modèles complexes avec les réseaux de Pétri. Celle-ci présente des difficultés qui pourraient faire obstacle à l'utilisation de cette méthode. Ensuite, j'aborderai deux thèmes de nature philosophique et historique qui sont liés à la biologie systémique et aux idées avancées dans cette thèse. S'il est possible de diviser les scientifiques en deux groupes, les premiers étant ceux qui comme Richard Feynman, prix Nobel de physique 1965, croient que la philosophie des sciences est aussi utile aux scientifiques que l'ornithologie l'est aux oiseaux, et les seconds, ceux pour la philosophie des sciences s'avèrent un efficace outil de réflexion sur les concepts mis en place par les scientifiques et sur leur construction d'une représentation rationnelle du monde, alors j'appartiens vraisemblablement au deuxième groupe !

7.1 Aspects méthodologiques de la simulation avec les réseaux de Pétri

La simulation des systèmes biologiques présente d'importantes difficultés. La nature même de ces systèmes pose certains problèmes. Et nous ne parlons ici même pas des complexités expérimentales des mesures en milieu *in vivo*, ni de la démarche ardue qui consiste à sélectionner les éléments à assembler dans un modèle pour reproduire le comportement cellulaire observé. Le développement de modèles biologiques réalistes

suppose l'intégration dans un même modèle de phénomènes se déroulant à différentes échelles de temps et de grandeur. Certaines composantes ont un comportement stochastique. Les modèles sont hétérogènes, de grande dimension et de plus en plus complexes à mesure que les données expérimentales s'accumulent.

7.1.1 Réseaux de Pétri versus équations différentielles

La simulation de tels modèles nécessite de considérables ressources de calcul. Une approche de modélisation basée sur les équations différentielles, qui est l'approche la plus communément utilisée en raison d'un imposant héritage théorique et d'une utilisation massive dans pratiquement tous les domaines d'application de modélisation, propose des techniques pour optimiser la simulation des modèles de grande taille. En recourant à des grappes multiprocesseurs et à des techniques de traitement parallèle comme la décomposition par domaine, la résolution numérique de larges modèles d'équations différentielles peut s'effectuer sans que le temps de simulation ne devienne dissuasif. Notons qu'avec les équations différentielles partielles, les variables spatiales peuvent s'ajouter au temps et qu'avec les équations différentielles stochastiques, l'inclusion de processus stochastiques est permise. Toutefois, les possibilités de ce type de modèle se limitent à la modélisation continue.

L'intérêt d'une approche hybride se situe en partie dans l'extension des possibilités de modélisation. En effet, l'ajout de quantités et d'événements discrets, pouvant aussi avoir un caractère stochastique, combine le pouvoir de modélisation de méthodes continues comme les équations différentielles à celui des approches discrètes. Une approche hybride semble mieux répondre aux besoins de modélisation réaliste des systèmes biologiques (Jordan et al., 2000). De plus, si nous considérons l'apport de la théorie des réseaux de Pétri et de son formalisme graphique, nous sommes enclins à penser que nous tenons là une approche ayant beaucoup de potentiel. Des

publications récentes montrent des signes probants d'une maturation des méthodes basées sur les réseaux de Pétri pour les applications biologiques (Sackmann et al., 2006; Li et al., 2006). Nous pensons aussi que la méthode que nous avons développée pour l'analyse de la dynamique de la propagation des signaux cellulaires, présentée au chapitre 6, contribue à ce mouvement. Malgré cela, l'inefficacité computationnelle de la simulation de modèles de réseaux de Pétri hybrides pourrait devenir un obstacle sérieux entraînant un désintérêt de cette approche dans le cas de vastes analyses quantitatives.

À notre connaissance, les modèles de réseaux de Pétri que nous avons simulés sont les plus grands modèles à avoir été présentés dans la littérature. Au cours de nos recherches, nous avons rencontré certains problèmes qu'il faut souligner. Ceux-ci ont été riches d'enseignements et nous amènent à proposer une méthodologie.

7.1.2 Réduction de la précision des simulations

De prime abord, le temps de calcul pour simuler un modèle de réseaux de Pétri hybrides est plus important qu'un modèle équivalent d'équations différentielles. Les techniques de résolution numérique sont différentes et les outils ne sont pas aussi performants. Le seul logiciel de simulation de réseaux de Pétri hybrides spécifique à la biologie, Genomic Object Net (GON), a été développé en Java, un langage de programmation dont on a démontré qu'il était moins performant que d'autres (Prechelt, 2000). Un premier compromis possible porte sur la précision de la simulation. Une simulation moins précise est moins vorace en ressources de calcul, mais les résultats peuvent aussi être corrompus en raison de l'imprécision. La figure 7.1 illustre les conséquences d'une résolution numérique avec un certain degré d'imprécision. La courbe de ce graphe de concentration s'avère tout de même une bonne approximation de la courbe malgré les oscillations provoquées par un grand pas de simulation.

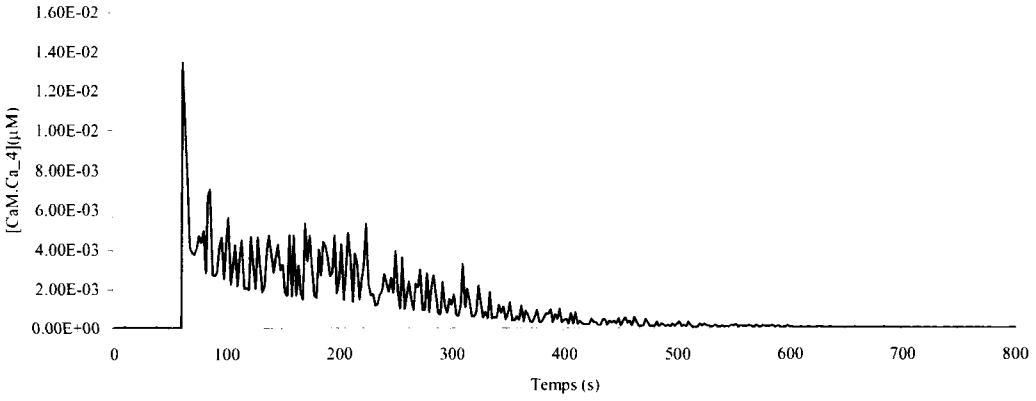


Figure 7.1: Concentration de $CaM.Ca_4$, illustrant la problématique de l'imprécision numérique lors d'une simulation

Heureusement, ces oscillations n'ont pas introduit de perturbations inattendues et indésirables dans le système et n'ont donc pas causé de vices de simulation. Dans un contexte de recherche où il n'y a pas de résultats attendus pour valider les résultats obtenus par simulation, il est impossible de déterminer avec certitude le seuil de précision nécessaire afin d'éviter une résolution numérique biaisée. C'est pourquoi nous déconseillons une diminution de la précision afin de simuler plus rapidement, plus particulièrement, lors des premières étapes de développement du modèle.

7.1.3 Simulation dans un environnement multiprocesseur

Nous avons aussi fait des tests de performance d'exécution de GON sur une grappe de multiprocesseurs puisque le traitement parallèle est une avenue moins dispendieuse que les superordinateurs pour obtenir une grande puissance de calcul. La figure 7.2 montre les résultats de ces tests, effectués sur la grappe Gropp du laboratoire DRAP (design et réalisation d'application parallèles) de l'École Polytechnique de Montréal. Il s'agit d'une grappe de 64 processeurs Opteron distribués sur 32 noeuds qui sont interconnectés à l'aide d'un réseau de fibre optique Myrinet et par un réseau Gigabit Ethernet. Cette grappe nous a servi de plateforme pour la simulation du modèle

Figure 7.2: Temps d'exécution d'une simulation en fonction du nombre de processeurs utilisés

des voies de la calmoduline et de la calcineurine du chapitre 6. La courbe de ce graphe montre un très faible gain de performance, ce qui révèle une inadéquation de GON à un environnement multiprocesseur. Ces résultats étaient prévisibles puisque l'architecture de ce logiciel n'a pas été conçue pour ce genre d'environnement. Pour l'instant, les équations différentielles semblent donc supérieures aux réseaux de Pétri hybrides pour leur temps de résolution. Toutefois, ceci est en partie le résultat du peu de recherche sur ce sujet.

La littérature scientifique contient bien des travaux sur l'utilisation de traitement parallèle (par exemple, voir Chiola et Ferscha (1993); Nicol et Mao (1995); Nketsa et Khalifa (2001); Knoke et al. (2005)) et aussi d'autres travaux portant sur l'optimisation de la simulation des réseaux de Pétri (par exemple, voir Gaeta (1996); Mortensen (2001)). Mais seuls les réseaux discrets temporisés ou colorés sont considérés dans ces études. L'optimisation et le traitement parallèle de réseaux de Pétri hybrides est donc est champs de recherche vierge.

7.1.4 Approche mixte de simulation

Pour la simulation de modèles complexes de réseaux de Pétri hybrides, où de larges pans sont de nature continue, nous avons conclu que pour des considérations d'efficacité, il vaut mieux adopter une approche mixte qui combine la simulation d'équations différentielles et la simulation de réseaux de Pétri. Lors du développement d'un modèle, nous suggérons de procéder par incrémentation, c'est-à-dire par la modélisation successive de parties du modèle. Ces sous-modèles sont individuellement validés pour ensuite être reliés. Aussi, toute simulation consommant beaucoup de ressources devrait d'abord être faite avec des modèles d'équations différentielles, où les éléments discrets sont approximés par des éléments continus. Cette technique permet d'obtenir plus rapidement des données de simulation qui permettent de confirmer la justesse de la modélisation. Lorsque des résultats satisfaisants sont obtenus et que les paramètres du modèle sont presque fixés, on peut alors passer aux réseaux de Pétri hybrides et inclure les aspects discret et stochastique. Cette approche de modélisation et de simulation profite de l'efficacité de la simulation d'équations différentielles ainsi que de la puissance de la modélisation des réseaux de Pétri hybrides fonctionnels.

7.2 Réflexions philosophiques

L'épistémologie est le domaine de la philosophie qui désigne la réflexion sur ce qu'est le savoir. L'origine grecque du mot signifie discours sur la connaissance. C'est notamment le cas de la philosophie des sciences qui étudie la connaissance scientifique d'un point de vue critique et qui a aussi parfois recours à l'histoire des sciences. L'essor actuel de la biologie est une excellente occasion pour jeter un regard critique sur cette discipline.

Dans un premier texte critique, je porterai mon attention sur le concept de fonction en biologie. Central dans cette thèse, ce concept est aussi à la base de la démarche biologique. Les biologistes cherchent à comprendre le fonctionnement du vivant et ils attribuent des "fins", des "fonctions" et même des "dessins" aux composantes des systèmes biologiques ainsi qu'aux systèmes eux-mêmes. Identifier la raison d'être de leurs objets d'étude amène les biologistes à utiliser un langage téléologique. Ce langage est aussi présent lorsque, adoptant un point de vue d'ingénierie en biologie, l'on parle du design du vivant. Pourtant, un tel langage est honni en chimie et en physique. Il est considéré non-scientifique par certains puisque insuffisamment réducteur et trop éloigné des lois naturelles. Selon les biologistes, ce langage est néanmoins nécessaire pour obtenir des explications satisfaisantes. Ces derniers sont aussi d'avis que la théorie de l'évolution par sélection naturelle de Darwin a évacué la "mauvaise" téléologie. Puisque j'ai abondamment utilisé le concept de fonction dans cette thèse, je me demanderai dans un premier texte si le recours à la téléologie est justifié en biologie.

Le second texte critique porte une discussion sur le caractère prétendument révolutionnaire de la biologie systémique. Génomique, protéomique, biologie systémique, épigénomique, et autres -omique ont semblé naître à un rythme accéléré au cours des deux dernières décennies. Dans mon mémoire de maîtrise, j'ai dressé un portrait de ce bouillonnement animant la recherche en biologie et de ce que cela signifiait pour la démarche de compréhension du vivant (Hardy, 2004). Je relayais en grande partie les points de vue de publications de vulgarisation et spécialisées, très souvent dithyrambiques à propos de la "nouvelle" biologie. J'ai utilisé le mot "révolution" à plusieurs reprises puisqu'il se retrouvait dans les articles que j'ai consultés. Mais la biologie systémique est-elle réellement une révolution scientifique ? Les nouvelles technologies et les nouvelles découvertes changent-elles la conceptualisation du vivant ? S'est-il produit un changement de paradigme en passant d'une approche réductionniste à

une approche systémique ?

7.2.1 La téléologie en biologie

Une courte mise en contexte s'impose afin d'exposer la présence d'un langage téléologique tant chez Darwin que dans un débat plus récent en philosophie de la biologie. La téléologie - du grec *télos*, fin ou but, et *logos*, discours - est l'étude des causes finales, ou finalisme. Aristote développa le principe de causalité pour en faire ressortir une finalité sous-jacente à la nature des choses. Après les Grecs, le finalisme fut assimilé à la représentation chrétienne du monde dans une croyance que tout dans la nature et ses processus avait une raison d'être, un but prédéterminé : une téléologie cosmique. Au XVIIIe siècle, Kant décrivit en termes strictement mécaniques ou newtoniens la nature inanimée, mais, en raison du peu de savoir biologique dont on disposait à l'époque, il ne put voir à l'oeuvre, dans tous les processus du monde vivant, que des forces téléologiques. Selon Mayr (2001), un des plus grands accomplissements philosophiques de Darwin fut d'expliquer par la sélection naturelle tous les phénomènes naturels pour lesquels Kant avait dû invoquer la téléologie. Ainsi, la théorie de l'évolution par sélection naturelle de Darwin - et ce malgré l'illusion de progrès et d'amélioration que le mot évolution induit - mit fin à la téléologie cosmique (Darwin, 1859). Pourtant, il est possible de voir la téléologie réintroduite par Darwin dans le fonctionnement même des êtres vivants. À propos des organes ayant diverses fonctions peu spécifiques à accomplir, Darwin (1859, p. 149) écrit : "In the same way that a knife which has to cut all sorts of things may be of almost any shape ; whilst a tool for some particular object had better be of some particular shape[1]." Darwin, implicitement, voit-il donc les organes accomplissant une *fonction*

[1]Traduction de cette citation dans l'édition française, p.202 : "On pourrait comparer ces organes à un couteau destiné à toutes sortes d'usages, et qui peut, en conséquence, avoir une forme quelconque, tandis qu'un couteau destiné à un usage déterminé doit prendre une forme particulière."

précise comme ayant un *but déterminé* ? La voie tracée par Darwin mènerait-elle à une interprétation téléologique du vivant, qui n'est plus cosmique, mais d'une autre nature ?

Le biologiste Mayr reconnaît qu'un langage téléologique est souvent utilisé en biologie pour formuler des affirmations à propos des fonctions des organes, des processus physiologiques ainsi qu'à propos du comportement et des actions d'espèces ou d'individus. Ce langage est caractérisé par l'utilisation des mots *fonction*, *raison d'être* et *but* ou par des affirmations que quelque chose existe ou est fait *afin de*. Les physiciens s'opposent à l'utilisation d'un tel langage, arguant qu'il est subjectif, chargé d'un sens métaphysique et qu'il introduit de l'anthropomorphisme en biologie. Les biologistes rétorquent que l'emploi d'un langage téléologique est légitime et qu'éliminer ce langage provoque la perte d'explications significatives. Mayr (1988) affirme que ce différend proviendrait d'une confusion sur la définition de téléologie. Pour y remédier, il identifie et distingue les différentes catégories de processus téléologiques. Le philosophe Nagel (1977) conteste cette catégorisation des processus téléologiques et il critique tout particulièrement le concept de programme de Mayr.

Dans ce court texte, j'explorerai les positions de Mayr sur les processus téléologiques en biologie à la lumière de la théorie de Darwin sur l'évolution par la sélection naturelle. Je tenterai donc de répondre à la question : L'utilisation d'une heuristique téléologique, telle que formulée par Mayr, est-elle fondée en biologie dans le cadre de la théorie de l'évolution ?

7.2.1.1 La téléologie : un concept que définit Ernst Mayr

Mayr est d'avis que le mot téléologie est historiquement galvaudé ; plusieurs phénomènes hétérogènes ont été maladroitement rassemblés sous la bannière "téléologique".

C'est pourquoi il propose quatre catégories pour faciliter l'analyse de la téléologie en biologie. Selon lui, les trois premières ont une solide base empirique alors qu'il n'y a aucune preuve de l'existence de la quatrième catégorie, la téléologie cosmique. Cette dernière catégorie a été brièvement abordée dans l'introduction et je n'y reviendrai plus. Je ne décrirai pas non plus la catégorie des dispositifs adaptatifs, qui n'est pas pertinente pour cette discussion.

Mayr regroupe dans une première catégorie tous les phénomènes téléologiques qui sont des processus dont l'état final est causé et déterminé par des lois naturelles (par exemple, la gravité ou la deuxième loi de la thermodynamique). Il qualifie ces processus de *téléomatiques*. Les processus téléomatiques sont orientés vers une fin, mais uniquement de façon passive et automatique.

Dans une seconde catégorie, Mayr regroupe tous les phénomènes téléologiques dont l'orientation vers un but déterminé est contrôlée par les opérations d'un programme. Il qualifie ces processus de *téléonomiques*. Selon Mayr, ce type de processus est possiblement la caractéristique la plus répandue dans le monde des organismes vivants. Il donne en exemple les activités reliées à la migration, à l'alimentation, à la séduction, à l'ontogenèse ainsi qu'à toutes les phases de la reproduction. Il y a aussi tous les processus dépendant du "programme" au coeur des cellules, le code génétique. Contrairement aux processus téléomatiques, les processus téléonomiques impliquent un processus dynamique plutôt que des conditions statiques. Ces processus ont deux caractéristiques : ils sont guidés par un programme, et ils dépendent d'un but qui est prévu par le programme et qui régule leur comportement. Mayr précise que tous les programmes biologiques sont le résultat de la sélection naturelle et qu'ils sont constamment ajustés selon la valeur sélective de leur résultat.

7.2.1.2 Des processus téléonomiques, une particularité du "vivant" ?

La deuxième catégorie de processus orientés vers un but que propose Mayr, les processus téléonomiques, repose sur l'idée de *programme*. Mayr (1992, p. 127) définit ce concept comme suit : " Un programme consiste en de l'information codée ou préarrangée qui contrôle un processus menant à un but et aussi en des instructions pour l'utilisation de cette information. " Il distingue deux types de programme : les programme fermés, comme l'ADN ; et les programmes ouverts, ou somatiques, auxquels de l'information additionnelle est incorporée par apprentissage durant la vie d'un organisme.

Nagel (1977) critique la conceptualisation de la téléologie par programme proposée par Mayr. Il émet quatre objections et il conclut que le concept de programme ne permet pas de décrire adéquatement les processus biologiques. À l'aide des définitions de Mayr, comment distinguer, en effet, les processus biologiques qui dépendent de lois naturelles (par exemple, les réactions chimiques déterminées par les lois de la thermodynamique) et qui sont donc téléomatiques, des autres processus biologiques guidés dynamiquement par un programme et qui sont donc téléonomiques ? Le concept de programme de Mayr ne révèle-t-il pas l'impossibilité passée et actuelle d'atteindre le niveau de réductionnisme physico-chimique nécessaire pour expliquer les comportements cellulaires "guidés" par l'ADN ? En d'autres mots, les comportements cellulaires seraient compris comme étant guidés par le programme de l'ADN parce qu'on ne possède pas de connaissances plus précises sur les phénomènes physico-chimiques qui les ramèneraient à des processus téléomatiques.

Malgré leur divergence, les propos du biologiste et du philosophe sont en accord avec la théorie de Darwin. Les deux hommes rejettent la téléologie cosmique et les explications non empiriques. Mayr a réagi aux critiques de Nagel. À son avis, la nouveauté du

concept de programme, issu de la récente théorie de l'information créée par Shannon (1949), générerait cette incompréhension. Mayr écarte rapidement toute intentionnalité en affirmant que les programmes biologiques sont le fruit de l'évolution par sélection naturelle. Pourtant, les programmes, tels qu'on les connaît aujourd'hui, sont produits par des individus ayant l'intention de créer des processus exécutés par un ordinateur, et ce pour atteindre un but. Ce processus est clairement téléologique. En est-il de même du "programme" du vivant ? Dans le cas des programmes somatiques, qui nécessitent un conditionnement, il est possible d'identifier une intention ou des états internes analogues à la volonté d'atteindre un but. Dans le cas des programmes fermés, particulièrement pour l'ADN, on ne peut pas identifier d'intention à moins d'être un apôtre de la théorie du dessein intelligent. Ce constat, à mon avis, affaiblit la proposition de Mayr puisqu'il révèle l'inadéquation de son analogie informationnelle (i.e. les programmes de l'ADN ne peuvent pas être comparés aux programmes informatiques).

À ce sujet, il est intéressant de s'arrêter à l'expression "code génétique". Est-ce un "code" parce que les triplets de nucléotides de l'ADN sont décodés par les ribosomes pour assembler des acides aminés en protéines ? Ou bien est-ce un "code" parce qu'il s'agit du programme de la cellule qui conduit son développement ? L'analogie informationnelle fallacieuse de Mayr est peut-être même présente dans ce concept des biologistes.

La téléologie renferme plusieurs pièges. Les philosophes et les physiciens ont raison d'être prudents à son sujet. Toutefois, rejeter le langage téléologique des biologistes comme ils le font est mal avisé. La réalité expérimentale des biologistes et des physiciens est radicalement différente. Les physiciens ont généralement affaire à des systèmes fortement homogènes, dont la modélisation mathématique et la déduction de lois sont possibles. Les biologistes font face à des systèmes d'une extraordinaire com-

plexité dont certains mécanismes sont encore difficiles, voire impossibles, à mesurer à l'échelle moléculaire. Des explications causales strictement physico-chimiques leur sont souvent inaccessibles avec les technologies actuelles. Sur ce point, je me range aux arguments de Mayr. La téléologie a une puissante valeur heuristique. Alors que les questions débutant par "Quoi ?" et "Comment ?" sont suffisantes en physique, les questions débutant par "Pourquoi ?" sont génératrices de nouvelles hypothèses et de savoir en biologie. Darwin n'a-t-il pas lui-même utilisé une heuristique téléologique pour comprendre la spécificité des variations entre individus et espèces, ce qui l'a mené à sa théorie ?

À la question qui a inspiré ce texte, je répondrai donc que l'utilisation d'une heuristique téléologique est fondée en biologie et qu'elle est compatible avec la théorie de l'évolution. Une approche téléologique peut parfois être l'approche la plus féconde de savoir, faute de moyens techniques. Les méthodes informatiques présentées dans cette thèse s'inscrivent dans ce type d'heuristiques. Chercher des fonctions systémiques aux macromolécules, visualiser le comportement dynamique des enzymes et caractériser la dynamique de la propagation d'un signal dans une cellule et ses motifs de régulation sont des méthodes qui, à divers degrés, tentent de répondre à des questions débutant par "Pourquoi ?". Malgré l'utilité que je reconnais au langage téléologique en biologie, je ne peux pas accepter la définition de processus téléonomiques de Mayr en raison des difficultés soulevées quant au concept de programme biologique. La vague déferlante de la génomique a mené les biologistes jusqu'au contenu du code génétique, mais sans livrer les clés de la compréhension du phénotype. L'heuristique téléonomique de Mayr a été utile en biologie moléculaire. Avec prudence, on peut utiliser les mots fonction, signal et information cellulaire. mais le paravent mental du "programme", si les biologistes le conservent intact, pourra devenir un obstacle à la compréhension des phénomènes physico-chimiques du "vivant".

7.2.2 La biologie systémique : une réelle révolution scientifique ?

Pour offrir une réponse à cette question, il faut définir le terme "révolution scientifique." Et, bien entendu, la réponse dépendra fortement de la définition ! Révolution conceptuelle, révolution technologique, grande découverte, tous ces éléments peuvent faire partie de la définition d'une révolution scientifique. Différentes définitions ont servi de prémisses à quelques discussions sur l'assertion que la biologie systémique est une révolution scientifique. Bothwell (2006), en se basant sur la définition développée par Kuhn (1962), réfute la thèse révolutionnaire. Wellstead (2006) a manifesté son désaccord avec ce point de vue. Il voit s'opérer en biologie un changement d'approche fondamental, annonciateur de révolution et il fonde son argumentaire révolutionnaire, entre autres, sur l'idée d'innovation disruptive de Christensen (1997). Le deuxième propos du texte de Wellstead est pourtant d'affirmer que la systématisation de la biologie est en partie l'héritage des idées d'Erwin Schrödinger, prix Nobel de physique 1933. Un héritage vieux de soixante ans peut-il constituer aujourd'hui une révolution ? Je présenterai brièvement l'héritage biologique de Schrödinger pour ensuite exposer la critique qu'en a fait Kupiec (2003), une critique audacieuse qui s'étend à toute la biologie moléculaire moderne. Je terminerai avec les commentaires de Lazebnik (2004) qui montre les limites conceptuelles de la biologie pour étudier les systèmes complexes. Mais avant, Bothwell et Wellstead soulèvent tous deux des argumentaires qui méritent considération. Ils sont contradictoires à première vue, mais nous verrons comment ils décrivent une réalité complémentaire.

7.2.2.1 Pas de changement de paradigme, dit Bothwell

Bothwell (2006) amorce son article en citant des exclamations triomphalistes sur la perspective systémique en biologie : qualifiée de "révolution" (Aderem, 2005) et

de "changement de paradigme des sciences de la vie modernes" (Aggarwal et Lee, 2003), la biologie des systèmes "promet de révolutionner notre compréhension des systèmes biologiques complexes de régulation" (Kitano, 2002b). Pourtant, les révolutions scientifiques décrites par Kuhn sont rares. Entre les révolutions, les scientifiques font de la science dite normale puisqu'ils adhèrent à un ensemble cohérent de lois, de procédures et d'hypothèses, ce que Kuhn nomme un paradigme. Le paradigme est peu remis en question et l'essentiel du travail scientifique, en période de science normale, est de faire des ajouts minimes au paradigme et de l'utiliser pour résoudre des énigmes de la nature. Le paradigme est ébranlé lorsque plusieurs anomalies le contredisant s'accumulent. Ces anomalies signifient, non pas que le paradigme est incomplet, mais qu'il est en partie incorrect. La réalisation de ce fait, jumelée à la proposition d'un nouveau paradigme adopté par une majorité de scientifiques, indique une rupture avec la tradition que l'on peut qualifier de révolution. Les deux paradigmes, l'ancien et le nouveau, sont alors incommensurables, c'est-à-dire qu'ils sont fondamentalement incompatibles.

Bothwell reconnaît la popularité de la biologie systémique : de nouveaux instituts de recherche sont inaugurés et des programmes universitaires sont créés, il y a une croissance exponentielle des articles inspirés par cette nouvelle discipline et les nouveaux périodiques consacrés à ce sujet sont nombreux. Selon Bothwell, il y a un engouement réel pour l'approche systémique, mais ceci n'est pas dû à un changement de paradigme, donc pas une révolution scientifique. Les scientifiques optent pour une nouvelle approche, pas pour un nouvel ensemble de théories, de procédés et d'hypothèses. L'argument principal de Bothwell est qu'on ne peut pas opposer le paradigme réductionniste (comprendre le vivant en décomposant les systèmes biologiques en une liste de composantes) au paradigme de la biologie systémique (comprendre le vivant en étudiant les fonctions et les processus qui émergent de l'interaction des composantes biologiques). Il s'agit en fait de la continuation d'un seul et même paradigme.

Bothwell fait appel à Aristote, dont le tiers des oeuvres portent sur la biologie et qui fut un des premiers à suggérer que pour comprendre un système complexe, il faut le décomposer en ses constituants, définir l'arrangement de ceux-ci, caractériser les changements qui se produisent dans le système et déterminer sa fonction. Pour Bothwell, la biologie systémique est donc la suite logique du réductionnisme biologique selon l'approche épistémologique aristotélicienne.

7.2.2.2 La recherche en biologie est en profonde mutation et la révolution est inévitable, selon Wellstead

Wellstead (2006), contrairement à l'analyse principalement philosophique que fait Bothwell, adopte une perspective plus historique. Selon lui, la recherche en biologie subit actuellement une systématisation, identique à celle qui a transformé la conception des produits de fabrication humaine complexes. Les manipulations en laboratoire sont automatisées et on a recours plus que jamais à la modélisation mathématique et à la simulation. Le développement de la théorie des systèmes et de certaines de ses constituantes, soient la théorie de la rétroaction, l'analyse dynamique et la modélisation mathématique, fournit maintenant des approches structurées pour la résolution de problèmes scientifiques à l'aide d'une approche systémique. Pour Wellstead, il y a aussi eu un synchronisme d'intérêt entre les biologistes et les théoriciens des systèmes et des technologies sont arrivées à maturité pour supporter les études systémiques.

Wellstead fait un parallèle entre la révolution industrielle et la révolution biologique. Par exemple, des technologies cruciales pour l'investigation biologique, comme la cytométrie, ont connu une accélération de développement exponentielle ces dernières années. Les courbes du progrès technologique de la biologie moléculaire des dernières années (ces courbes tracent la distribution temporelle des innovations majeures) montrent des similitudes frappantes avec celles du développement technolo-

gique de la révolution industrielle. Selon Wellstead, nous sommes à la fin de la période pré-révolutionnaire et il juge que la révolution scientifique est inévitable.

Les conclusions de Bothwell et de Wellstead sont diamétralement opposées parce qu'ils analysent des aspects différents de la biologie et de l'approche systémique. Il n'y a pas de révolution scientifique selon Bothwell puisqu'il n'y a pas de révolution conceptuelle. Il y a révolution scientifique selon Wellstead puisqu'il se produit une révolution technologique majeure. Mais Wellstead affirme aussi que la biologie systémique, avec ses présupposés actuels, est en partie l'héritage de Schrödinger parce qu'il a conceptualisé le fonctionnement des organismes vivants comme s'approchant d'un fonctionnement purement mécanique, donc déterministe. Selon Wellstead, il faut concevoir les organismes vivants ainsi pour être en mesure d'appliquer la théorie des systèmes. Cette affirmation circonscrit la position de Wellstead et la rapproche de celle de Bothwell. La révolution décrite par Wellstead n'a rien d'un changement paradigmatique. Voyons pourquoi.

7.2.2.3 Un dogme pour la biologie moléculaire, inspiré par Schrödinger

Le dogme central de la biologie moléculaire (c'est réellement le mot dogme qui est communément employé, ce qui n'est pas sans rappeler Kuhn, pour qui l'enseignement de la science normale est un endoctrinement) est que l'information génétique contenue dans l'ADN, le génotype, est retranscrite en ARN pour ensuite être décodée et transformée en protéines, le phénotype. Ce flux d'information confère un rôle particulier à l'ADN, celui d'une macromolécule contenant sous forme d'information codée le plan architectural des organismes.

La paternité de cette idée revient à Schrödinger (1944), pour qui les systèmes biologiques diffèrent des systèmes physiques puisqu'on ne peut pas y appliquer la loi des

grands nombres, ce principe créant l'ordre macroscopique à partir d'un désordre microscopique. La seule autre avenue possible, spécifique aux systèmes biologiques, est celle d'un principe d'ordre microscopique menant à un ordre macroscopique. Schrödinger propose que l'information génétique soit ce principe d'ordre microscopique. Inspiré par Delbrück, il propose même qu'un cristal apériodique soit le support de cette information, et ce, un peu moins de dix ans avant la découverte de l'ADN par Watson et Crick! Le raisonnement de Schrödinger, père de la biologie moléculaire, a ainsi été validé par une découverte scientifique majeure. Son influence est encore bien présente puisque le dogme central de la biologie moléculaire constitue une des théories fondamentales enseignées aux jeunes biologistes.

7.2.2.4 L'erreur de Schrödinger : les anomalies soulignées par Kupiec

L'ordre microscopique, tel que conçu par Schrödinger, a mené les biologistes moléculaires à la stéréospécificité des protéines. La stéréospécificité, définie par la forme et la charge de chaque protéine, limite à un nombre très réduit les possibilités d'associations entre protéines. Subséquemment, les protéines peuvent générer des chaînes de réactions causales qui permettent la transmission d'information. On remarquera que les modèles présentés et les méthodes avancées dans cette thèse sont fondés sur cette hypothèse.

Kupiec (2003) est un détracteur du concept de stéréospécificité. Dans une analyse théorique, il compare ce concept ainsi que celui de l'information génétique à la métaphysique aristotélicienne (encore Aristote!). En effet, puisque les biologistes supposent que l'organisme s'explique par ses gènes, le réductionnisme génétique constitue une actualisation de la cause finale d'Aristote. Kupiec s'oppose au rôle particulier d'ontogénèse que les biologistes attribuent aux gènes. Kupiec affirme aussi que la stéréospécificité des protéines n'a pas réellement été démontrée expérimentalement.

Il cite en exemple les études *in vitro* des facteurs de transcription (les protéines qui régulent spécifiquement l'expression des gènes en se liant à l'ADN) qui n'ont jamais pu montrer que l'activation et la répression spécifiques des gènes reposent sur la stéréospécificité.

Si on adopte l'ontologie kuhnienne, les problèmes identifiés par Kupiec sont des anomalies qui menacent le paradigme actuel de la biologie. Kuhn (1962) explique clairement la mécanique du monde scientifique : face aux anomalies, les scientifiques peuvent choisir de les ignorer, de modifier légèrement leur paradigme pour les résoudre, ou finalement, d'adopter un nouveau paradigme, ce qui déclenche une révolution scientifique.

Kupiec a proposé un nouveau paradigme. Il s'agit d'une biologie moléculaire darwinienne où le modèle de hasard-sélection s'applique à la cellule. Selon cette nouvelle définition de la biologie, différentes structures peuvent apparaître, avec des durées de vie et des probabilités différentes, et le comportement des protéines comme des cellules est exprimé en termes probabilistes. Le paradigme de la biologie moléculaire darwinienne n'a pas entraîné de révolution scientifique. Il faut des anomalies sérieuses pour que les scientifiques, qui ont toujours démontré un certain conservatisme conceptuel, modifient leur paradigme. Les anomalies soulevées par Kupiec et sa solution n'ont pas suscité une adhésion massive à ses idées. Néanmoins, voilà, contrairement à la biologie systémique, un exemple de germes conceptuels d'une révolution scientifique.

7.2.2.5 Alors, pourquoi la biologie systémique ?

Je crois bien avoir démontré que l'arrivée de la biologie systémique n'est pas un changement de paradigme pour la biologie moléculaire. Même si des mathématiciens, des

informaticiens, des ingénieurs, des chimistes et des physiciens se joignent maintenant aux équipes de recherche oeuvrant en biologie, les concepts fondamentaux de la biologie demeurent inchangés. Au-delà de la technologie, on peut alors se demander où se situe l'innovation scientifique de la biologie systémique puisque cette nouvelle discipline ne vient pas bouleverser les concepts établis. L'innovation provient du fait que la biologie systémique crée un pont entre deux univers conceptuels existants pour permettre la résolution de nouveaux problèmes.

Lazebnik (2004), dans un article au ton humoristique mais tout de même très convainquant, imagine un biologiste réparer une radio avec les méthodes qu'il utilise au laboratoire. Il veut ainsi démontrer que les notations, les concepts et les outils que les biologistes utilisent aujourd'hui sont trop primitifs pour étudier les systèmes biologiques complexes. Il est d'avis que l'apport du domaine du génie pourrait être bénéfique pour la biologie. Ces propos recoupent ceux de Csete et Doyle (2002) que nous avons présentés en détail au chapitre 2 et confirment le grand intérêt de leur démarche.

Nonobstant notre conclusion que la biologie systémique n'est pas une révolution scientifique, elle est un formidable moteur pour de nouvelles découvertes. Avec ces découvertes viendront probablement la constatation d'importantes anomalies, qu'il sera alors impossible d'ignorer, et ce sera la révolution !

CONCLUSION ET RECOMMANDATIONS

Les contributions originales de cette thèse à l'avancement des connaissances, via ses trois articles, sont nombreuses. Parmi celles-ci, il y a deux ajouts à la théorie des réseaux de Pétri : les réseaux de Pétri peinturés, une technique efficace pour visualiser le marquage d'un modèle complexe ; et les composantes d'équilibres, qui sont un sous-ensemble des T-invariants pour les réseaux de Pétri continus et hybrides.

Une autre contribution, présentée au chapitre 5 est la résolution des problèmes des méthodes actuelles de visualisation des données de simulation des modèles biochimiques. Ces méthodes sont mieux adaptées aux réseaux métaboliques qu'aux voies de transduction. La méthode de visualisation que nous avons développée est basée sur des techniques issues de la théorie des réseaux de Pétri. Elle propose une visualisation qui tient compte de l'organisation des voies de transduction et de leurs enzymes. Cette information est obtenue par une analyse des P-invariants. Ainsi, cette méthode peut montrer plus clairement la dynamique de commutation des enzymes, qui est le mécanisme fondamental de transmission de signaux cellulaires.

La dernière contribution, qui tout de même la plus significative aux yeux de l'auteur, est la méthode d'analyse de la dynamique de la propagation de signal d'une voie de transduction présentée au chapitre 6. Cette méthode, entièrement automatisable, transforme les données de variation de concentrations moléculaires en données de transmission de signaux cellulaire. La première étape de cette transformation est de créer une représentation graphique plus simple de la structure du système en trouvant ses P-invariants. La deuxième étape est de produire un portrait dynamique de la propagation du signal dans le réseau en ayant au préalable identifié les T-invariants et les segments de flux d'information du système. Dans une troisième étape, les chemins de propagation sont analysés pour détecter la présence de motifs

de régulation comme les boucles de rétroaction. En résumé, cette méthode offre un accès rapide aux mécanismes systémiques de traitement d'information de la cellule.

L'originalité des travaux de recherche rapportés dans cette thèse est aussi d'un autre ordre. Les applications des réseaux de Pétri à la biologie moléculaire se sont toujours cantonnées dans deux types d'utilisation : les analyses qualitatives et les analyses quantitatives. Les analyses qualitatives sont principalement utilisées à l'étape de modélisation pour effectuer une validation du modèle. En d'autres mots, les analyses qualitatives sont utilisées pour s'assurer que le modèle a les propriétés attendues. Dans les analyses quantitatives, les réseaux de Pétri servent de support à la simulation. Les données de simulation sont, quant à elles, analysées par d'autres méthodes. Dans cette thèse, j'ai proposé d'utiliser les réseaux de Pétri pour procéder à l'analyse des données. Au meilleur de mes connaissances, ceci est une première. Ces travaux sont d'autant plus intéressants que les techniques d'analyses de ce formalisme permettent d'obtenir des informations sur les caractéristiques systémiques d'un modèle. Ces techniques sont remarquablement utiles pour développer des méthodes d'analyse destinées à la biologie systémique. Pour cette raison, elles ont facilité l'atteinte de l'objectif de recherche des travaux de recherche présentés dans cette thèse.

Les résultats de mes travaux de recherche renforcent l'idée que les réseaux de Pétri est un formalisme de choix pour la modélisation biologique. Nous pouvons même imaginer une méthodologie complète où modélisation, simulation et analyse des données sont effectuées avec les réseaux de Pétri. Cette méthodologie pourrait être implémentée dans un logiciel facilement utilisable par les biologistes. L'application de cette méthodologie aux modèles de grande taille, qui seront de plus en plus nombreux en biologie, est néanmoins encore laborieuse en raison des difficultés rencontrées lors de la simulation de tels modèles. Ces problèmes ont été l'objet de la discussion du chapitre 7 et un moyen d'y pallier a été présenté. Cette solution n'est pas idéale

puisqu'il faut recourir à un autre formalisme pour accélérer les simulations lors des phases critiques de la modélisation. De meilleures solutions seront possibles lorsque davantage de travaux de recherche sur l'optimisation de la simulation des réseaux de Pétri hybrides et sur l'utilisation du traitement parallèle auront été menés.

Malgré tous les avantages mentionnées dans cette thèse et qui encouragent l'utilisation des réseaux de Pétri pour la modélisation biologique, ce formalisme n'a pas été encore adopté par un grand nombre d'équipes de chercheurs. À mon avis, cette situation est due à l'absence d'une démonstration convaincante de son utilité avec un modèle original. Les applications de ce formalisme à la modélisation biologique ont toujours été réalisées par des scientifiques dont la formation principale est en informatique. Cette critique s'applique aussi à cette thèse. Je suis un ingénieur qui a proposé de nouvelles méthodes computationnelles en vue d'applications en biologie moléculaire. Pour démontrer l'intérêt de ces méthodes, j'ai utilisé un modèle développé par des biologistes. Ce modèle, quoique formidable par sa complexité, n'exploite pas encore tout le pouvoir de modélisation des réseaux de Pétri. S'exprime ici l'immense défi de la biologie systémique. Bâtir un pont entre les sciences appliquées et la biologie est une opération délicate puisque deux langages, deux pratiques et deux communautés se rencontrent. Je compte moi-même relever ce défi, en allant compléter un stage postdoctoral sous la direction d'un biologiste.

BIBLIOGRAPHIE

Aderem, A. (2005). Systems biology : Its practice and challenges. *Cell*, **121**(4), 511–513.

Aderem, A. et Smith, K. D. (2004). A systems approach to dissecting immunity and inflammation. *Seminars in Immunology*, **16**(1), 55–67.

Aggarwal, K. et Lee, H. K. (2003). Functional genomics and proteomics as a foundation for systems biology. *Briefings in Functional Genomics and Proteomics*, **2**(3), 175–184.

Ajay, S. M. et Bhalla, U. S. (2006). Synaptic plasticity in vitro and in silico : Insights into an intracellular signaling maze. *Physiology*, **21**(4), 289–296.

Allen, C. et Bekoff, M. (1995). Biological function, adaptation, and natural design. *Philosophy of Science*, **62**(4), 609–622.

Alm, E. et Arkin, A. P. (2003). Biological networks. *Current Opinion in Structural Biology*, **13**(2), 193–202.

Alon, U. (2003). Biological networks : the tinkerer as an engineer. *Science*, **301**(5641), 1866–1867.

Alves, R., Antunes, F., et Salvador, A. (2006). Tools for kinetic modeling of biochemical networks. *Nature Biotechnology*, **24**(6), 667–672.

Angeli, D., Ferrell, J. E., et Sontag, E. D. (2004). Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(7), 1822–1827.

Angeli, D. et Sontag, E. D. (2003). Monotone control systems. *IEEE Transactions on Automatic Control*, **48**(10), 1684 – 1698.

Ariew, A. (2003). Ernst mayr's 'ultimate/proximate' distinction reconsidered and reconstructed. *Biology and Philosophy*, **18**(4), 553–565.

Batchelor, E., Silhavy, T. J., et Goulian, M. (2004). Continuous control in bacterial regulatory circuits. *Journal of Bacteriology*, **186**(22), 7618–7625.

Becker, M. Y. et Rojas, I. (2001). A graph layout algorithm for drawing metabolic pathways. *Bioinformatics*, **17**(5), 461–467.

Bhalla, U. S. (2002). Use of kinetikit and GENESIS for modeling signaling pathways. *Methods in Enzymology*, **345**, 3–23.

Bhalla, U. S. (2004). Signaling in small subcellular volumes. II. stochastic and diffusion effects on synaptic network properties. *Biophysical Journal*, **87**(2), 745–753.

Bhalla, U. S. et Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. *Science*, **283**, 381–387.

Bhalla, U. S. et Iyengar, R. (2001). Robustness of the bistable behavior of a biological signaling feedback loop. *Chaos*, **11**(1), 221–226.

Bhalla, U. S., Ram, P. T., et Iyengar, R. (2002). Map kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science*, **297**(5583), 1018–1023.

Bliss, T. V. P. et Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J Physiol*, **232**(2), 331–356.

Blitzer, R. D. (2005). Long-term potentiation : Mechanisms of induction and maintenance. *Science's STKE*, **2005**(309), tr26–.

Blitzer, R. D., Connor, J. H., Brown, G. P., Wong, T., Shenolikar, S., Iyengar, R., et Landau, E. M. (1998). Gating of CaMKII by cAMP-Regulated Protein Phosphatase Activity During LTP. *Science*, **280**(5371), 1940–1943.

Bothwell, J. H. F. (2006). The long past of systems biology. *New Phytologist*, **170**(1), 6–10.

Carlson, J. M. et Doyle, J. C. (1999). Highly optimized tolerance : a mechanism for power laws in designed systems. *Physical Review E*, **60**(2 Pt A), 1412–1427.

Carlson, J. M. et Doyle, J. C. (2002). Complexity and robustness. *Proceedings of the National Academy of Sciences of the United States of America*, **99 Suppl 1**, 2538–2545.

Chen, B.-S., Wang, Y.-C., Wu, W.-S., et Li, W.-H. (2005). A new measure of the robustness of biochemical networks. *Bioinformatics*, **21**(11), 2698–2705.

Chen, M. et Hofestädt, R. (2003). Quantitative Petri net model of gene regulated metabolic networks in the cell. *In Silico Biology*, **3**(3), 347–365.

Chiola, G. et Ferscha, A. (1993). Distributed simulation of petri nets. *IEEE Parallel & Distributed Technology : Systems & Applications*, **1**(3), 33–50.

Cho, K.-H., Choo, S.-M., Wellstead, P., et Wolkenhauer, O. (2005a). A unified framework for unraveling the functional interaction structure of a biomolecular network based on stimulus-response experimental data. *FEBS Letters*, **579**(20), 4520–4528.

Cho, K.-H., Johansson, K. H., et Wolkenhauer, O. (2005b). A hybrid systems framework for cellular processes. *Biosystems*, **80**(3), 273–282.

Christensen, C. M. (1997). *The innovator's dilemma : when new technologies cause great firms to fail.* Harvard Business School Press, Boston, Massachusetts.

Csete, M. E. et Doyle, J. C. (2002). Reverse engineering of biological complexity. *Science*, **295**(5560), 1664–1669.

Csete, M. E. et Doyle, J. C. (2004). Bow ties, metabolism and disease. *Trends in Biotechnology*, **22**(9), 446–450.

Darwin, C. (1859). *On the Origin of Species.* John Murray, London.

David, R. et Alla, H. (2004). *Discrete, Continuous, and Hybrid Petri Nets.* Springer, Berlin. 524 p.

de Jong, H. (2002). Modeling and simulation of genetic regulatory systems : a literature review. *Journal of Computational Biology*, **9**(1), 67–103.

Deavours, D., Clark, G., Courtney, T., Daly, D., Derisavi, S., Doyle, J., Sanders, W., et Webster, P. (2002). The mobius framework and its implementation. *Software Engineering, IEEE Transactions on*, **28**(10), 956–969.

Doi, A., Fujita, S., Matsuno, H., Nagasaki, M., et Miyano, S. (2003a). Constructing biological pathway models with hybrid functional Petri nets. *In Silico Biology*, **4**(2), 271–291.

Doi, A., Nagasaki, M., Fujita, S., Matsuno, H., et Miyano, S. (2003b). Genomic Object Net : II. Modeling biopathways by hybrid functional Petri net with extension. *Applied Bioinformatics*, **2**(3), 185–188.

Doi, A., Nagasaki, M., Matsuno, H., et Miyano, S. (2006). Simulation-based validation of the p53 transcriptional activity with hybrid functional petri net. *In Silico Biology*, **6**, 0001.

Doyle, J. et Csete, M. (2005). Motifs, control, and stability. *PLoS Biology*, **3**(11), 1868–1872.

El-Samad, H., Kurata, H., Doyle, J. C., Gross, C. A., et Khammash, M. (2005). Surviving heat shock : control strategies for robustness and performance. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(8), 2736–2741.

Espina, V., Woodhouse, E. C., Wulfkuhle, J., Asmussen, H. D., Petricoin, E. F., et Liotta, L. A. (2004). Protein microarray detection strategies : focus on direct detection technologies. *Journal of Immunological Methods*, **290**(1-2), 121–133.

Eungdamrong, N. J. et Iyengar, R. (2004). Computational approaches for modeling regulatory cellular networks. *Trends in Cell Biology*, **14**(12), 661–669.

Formanowicz, D., Sackmann, A., Formanowicz, P., et Blazewicz, J. (2007). Petri net based model of the body iron homeostasis. *Journal of Biomedical Informatics*, **40**(5), 476–485.

Fujita, S., Matsuis, M., Matsuno, H., et Miyano, S. (2004). Modeling and simulation of fission yeast cell cycle on hybrid functional petri net. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, **E87-A**(11), 2919 – 28.

Gaeta, R. (1996). Efficient discrete-event simulation of colored petri nets. *Software Engineering, IEEE Transactions on*, **22**(9), 629–639.

German, R., Kelling, C., Zimmermann, A., et Hommel, G. (1995). Timenet : a toolkit for evaluating non-markovian stochastic petri nets. *Performance Evaluation*, **24**(1-2), 69–87.

Gilbert, D. et Heiner, M. (2006). From petri nets to differential equations–an integrative approach for biochemical network analysis. In *Proceedings Application and Theory of Petri Nets 2006, Turku, Finland, June*, volume 4024 of *Lecture Notes in Computer Science*, Berlin, pages 181–200. Springer.

Goss, P. J. E. et Peccoud, J. (1998). Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proceedings of the National Academy of Sciences of the United States of America*, **95**(12), 6750–6755.

Goss, P. J. E. et Peccoud. J. (1999). Analysis of the stabilizing effect of Rom on the genetic network controlling ColE1 plasmid replication. In *Proceedings of the Pacific Symposium on Biocomputing (PSB'99)*, volume 4, pages 65–76.

Hardy, S. (2004). Modélisation et simulation de systèmes biologiques à l'aide de réseaux de Pétri. Master's thesis, École Polytechnique de Montréal, Montréal. 99 pages.

Hardy, S. et Robillard, P. N. (2003). Modélisation et simulation de processus de biologie moléculaire basées sur les réseaux de pétri : une revue de littérature. Technical report, École Polytechnique de Montréal, Montréal. 19 p. EPM/RT-2003-10.

Hardy, S. et Robillard, P. N. (2004). Modeling and simulation of molecular biology systems using petri nets : modeling goals of various approaches. *Journal of Bioinformatics and Computational Biology*, **2**(4), 595–613.

Hardy, S. et Robillard, P. N. (2005). Phenomenological and molecular-level petri net

modeling and simulation of long-term potentiation. *Biosystems*, **82**(1), 26–38.

Hardy, S. et Robillard, P. N. (2006a). A functional representation of the simulation data of biochemical models based on molecular activity. *In Silico Biology*, page (article soumis).

Hardy, S. et Robillard, P. N. (2006b). Painted petri net and functional abstraction to visualize dynamic modeling. *European Simulation and Modelling Conference 2006. ESM 2006.*

Hardy, S. et Robillard, P. N. (2007a). Petri net-based method for the analysis of the dynamics of signal propagation in signaling pathways. *Bioinformatics*, page (article soumis).

Hardy, S. et Robillard, P. N. (2007b). Petri net-based visualization of signal transduction pathway simulations. *IET Systems Biology*, page (article soumis).

Hardy, S. et Robillard, P. N. (2007c). Visualization of the simulation data of biochemical network models : a painted petri net approach. In *Proceedings of the Summer Computer Simulation Conference 2007 (SCSC2007).*

Hartwell, L. H., Hopfield, J. J., Leibler, S., et Murray, A. W. (1999). From molecular to modular cell biology. *Nature*, **402**(6761 Suppl), 47–52.

Heiner, M. et Koch, I. (2004). Petri net based model validation in systems biology. In *Proceedings Application and Theory of Petri Nets 2004, Bologna, Italy, June.*, volume 2679 of *Lecture Notes in Computer Science*, Berlin, pages 216–237. Springer.

Heiner, M., Koch, I., et Will, J. (2004). Model validation of biological pathways using petri nets–demonstrated for apoptosis. *Biosystems*, **75**(1-3), 15–28.

Hofestadt, R. et Thelen, S. (1998). Quantitative modeling of biochemical networks. *In Silico Biology*, **1**(1), 39–53.

Hofestädt, R. (1994). A Petri net application of metabolic processes. *Journal of System Analysis, Modeling and Simulation*, **16**, 113–122.

Ideker, T., Galitski, T., et Hood, L. (2001a). A new approach to decoding life : Systems biology. *Annual Review on Genomics and Human Genetics*, **2**, 343–372.

Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., et Hood, L. (2001b). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, **292**(5518), 929–934.

Iyengar, R. (1996). Gating by Cyclic AMP : Expanded Role for an Old Signaling Pathway. *Science*, **271**(5248), 461–463.

Jacob, F. et Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, **3**, 318–356.

Jensen, K. (1992). *Coloured Petri nets : basic concepts, analysis methods, and practical use.* Monographs on Theoritical Computer Science. Springer-Verlag, New York. 184 p.

Jordan, J. D., Landau, E. M., et Iyengar, R. (2000). Signaling networks : the origins of cellular multitasking. *Cell*, **103**(2), 193–200.

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, **8**(1), 1 – 8.

Küffner, R., Zimmer, R., et Lengauer, T. (2000). Pathway analysis in metabolic

databases via differential metabolic display (DMD). *Bioinformatics*, **16**(9), 825–836.

Khammash, M. et El-Samad, H. (2004). Systems biology : from physiology to gene regulation. *IEEE Control Systems Magazine*, **24**(4), 62 – 76.

Kikuchi, S., Fujimoto, K., Kitagawa, M., Fuchikawa, T., Abe, M., Oka, K., Takei, K., et Tomita, M. (2003). Kinetic simulation of signal transduction system in hippocampal long-term potentiation with dynamic modeling of protein phosphatase 2A. *Neural Networks*, **16**(9), 1389–1398.

Kindler, E. et Pales, C. (2004). 3d-visualization of petri net models : Concept and realization. *Proceedings of Applications and Theory of Petri Nets*, **3099**, 464–473.

Kitano, H. (2002a). Computational systems biology. *Nature*, **420**(6912), 206–210.

Kitano, H. (2002b). Systems biology : a brief overview. *Science*, **295**(5560), 1662–1664.

Kitano, H. (2004). Biological robustness. *Nature Reviews Genetics*, **5**(11), 826–837.

Kitano, H., Funahashi, A., Matsuoka, Y., et Oda, K. (2005). Using process diagrams for the graphical representation of biological networks. *Nature Biotechnology*, **23**(8), 961 – 966.

Knoke, M., Kühling, F., Zimmermann, A., et Hommel, G. (2005). Performance of a distributed simulation of timed colored petri nets with fine-grained partitioning. In *Proceedings of Design, Analysis, and Simulation of Distributed Systems Symposium, (DASD 2005), San Diego, USA*, pages 63–71.

Koch, I., Junker, B. H., et Heiner, M. (2005). Application of petri net theory for

modelling and validation of the sucrose breakdown pathway in the potato tuber. *Bioinformatics*, **21**(7), 1219–1226.

Koh, G., Teong, H. F. C., Clement, M.-V., Hsu, D., et Thiagarajan, P. (2006). A decompositional approach to parameter estimation in pathway modeling : a case study of the Akt and MAPK pathways and their crosstalk. *Bioinformatics*, **22**(14), e271–280.

Ku, J., Feng, X.-J., et Rabitz, H. (2004). Closed-loop learning control of bio-networks. *Journal of Computational Biology*, **11**(4), 642–659.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. The University of Chicago Press, Chicago, Illinois.

Kupiec, J.-J. et Sonigo, P. (2003). *Ni Dieu ni gène : pour une autre théorie de l'hérédité*. Points. Science. Éditions du Seuil.

Lazebnik, Y. (2004). Can a biologist fix a radio ? - or, what i learned while studying apoptosis. *Biochemistry (Moscow)*, **69**(12), 1403 – 1406.

Lee, D.-Y., Zimmer, R., Lee, S.-Y., Hanisch, D., et Park, S. (2004). Knowledge representation model for systems-level analysis of signal transduction networks. *Genome Informatics*, **15**(2), 234–243.

Lee, D.-Y., Zimmer, R., Lee, S. Y., et Park, S. (2006). Colored petri net modeling and simulation of signal transduction pathways. *Metabolic Engineering*, **8**(2), 112–122.

Li, C., Ge, Q.-Q., Nakata, M., Matsuno, H., et Miyano, S. (2007). Modelling and simulation of signal transductions in an apoptosis pathway by using timed petri nets. *Journal of Biosciences*, **32**(1), 113–127.

Li, C., Suzuki, S., Ge, Q.-Q., Nakata, M., Matsuno, H., et Miyano, S. (2006). Structural modeling and analysis of signaling pathways based on petri nets. *Journal of Bioinformatics and Computational Biology*, **4**(5), 1119–1140.

Ma'ayan, A. et Iyengar, R. (2006). From components to regulatory motifs in signalling networks. *Briefings in Functional Genomics and Proteomics*, **5**(1), 57–61.

Ma'ayan, A., Jenkins, S. L., Neves, S., Hasseldine, A., Grace, E., Dubin-Thaler, B., Eungdamrong, N. J., Weng, G., Ram, P. T., Rice, J. J., Kershenbaum, A., Stolovitzky, G. A., Blitzer, R. D., et Iyengar, R. (2005). Formation of regulatory patterns during signal propagation in a mammalian cellular network. *Science*, **309**(5737), 1078 – 1083.

MacEachren, A. M. (2004). *How Maps Work : Representation, Visualization, and Design*. The Guilford Press. 513 p.

Mason, O. et Verwoerd, M. (2007). Graph theory and networks in biology. *Systems Biology, IET*, **1**(2), 89–119.

Matsui, M., Fujita, S., Suzuki, S.-i., Matsuno, H., et Miyano, S. (2004). Simulated cell division processes of the xenopus cell cycle pathway by genomic object net. *Journal of Integrative Bioinformatics*, **1**(1), 95–104.

Matsuno, H., Doi, A., Nagasaki, M., et Miyano, S. (2000). Hybrid petri net representation of gene regulatory network. *Pacific Symposium on Biocomputing*, pages 341–352.

Matsuno, H., Fujita, S., Doi, A., Nagasaki, M., et Miyano, S. (2003a). Towards biopathways modeling and simulation. In van der Aalst, W. M. P. et Best, E., éditeurs, *Proceedings of the 24th Conference on Applications and Theory of Petri Nets (ICATPN 2003)*, volume 2679 of *Lecture Notes in Computer Science*, pages

3–22. Springer-Verlag.

Matsuno, H., Inouye, S.-I. T., Okitsu, Y., Fujii, Y., et Miyano, S. (2006a). A new regulatory interaction suggested by simulations for circadian genetic control mechanism in mammals. *Journal of Bioinformatic and Computational Biology*, **4**(1), 139–153.

Matsuno, H., Li, C., et Miyano, S. (2006b). Petri Net Based Descriptions for Systematic Understanding of Biological Pathways. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, **E89-A**(11), 3166–3174.

Matsuno, H., Murakami, R., Yamane, R., Yamasaki, N., Fujita, S., Haruka, Y., et Miyano, S. (2003b). Boundary formation by notch signaling in drosophila multicellular systems : Experimental observations and gene network modeling by Genomic Object Net. In *Proceedings of the Pacific Symposium on Biocomputing*, volume 8, pages 152–163.

Matsuno, H., Tanaka, Y., Aoshima, H., Doi, A., Matsui, M., et Miyano, S. (2003c). Biopathways representation and simulation on hybrid functional Petri net. *In Silico Biology*, **3**(3), 389–404.

Mayr, E. (1988). *Toward a New Philosophy of Biology, Observations of an Evolutionist*, chapter The multiple meanings of teleological, pages 38–66. Havard University Press, Cambridge, MA.

Mayr, E. (1992). The idea of teleology. *Journal of the History of Ideas*, **53**(1), 117–135.

Mayr, E. (2001). The philosophical foundations of darwinism. *Proceedings of the American Philosophical Society*, **145**(4), 488–495.

Mendes, P. (1997). Biochemistry by numbers : simulation of biochemical pathways with gepasi 3. *Trends in Biochemical Sciences*, **22**(9), 361–363.

Mesarovic, M. D., Sreenath, S. N., et Keene, J. (2004). Search for organising principles : understanding in systems biology. *IEE Systems Biology*, **1**(1), 19 – 27.

Müller, W. et Schumann, H. (2003). Visualization methods for time-dependent data - an overview. *Proceedings of the 2003 Winter Simulation Conference*, **Vol.1**, 737 – 745.

Morohashi, M., Winn, A. E., Borisuk, M. T., Bolouri, H., Doyle, J., et Kitano, H. (2002). Robustness as a measure of plausibility in models of biochemical networks. *Journal of Theoretical Biology*, **216**(1), 19–30.

Mortensen, K. H. (2001). Efficient data-structures and algorithms for a coloured petri nets simulator. In Jensen, K., éditeur, *3rd Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools (CPN'01)*, pages 57–74. DAIMI PB-554, Aarhus University.

Nagasaki, M., Doi, A., Matsuno, H., et Miyano, S. (2003). Genomic Object Net : I. A platform for modeling and simulating biopathways. *Applied Bioinformatics*, **2**(3), 181–184.

Nagasaki, M., Doi, A., Matsuno, H., et Miyano, S. (2004). A versatile petri net based architecture for modeling and simulation of complex biological processes. *Genome Informatics*, **15**(1), 180–197.

Nagasaki, M., Doi, A., Matsuno, H., et Miyano, S. (2005). *Bioinformatics Technologies*, chapter Computational Modeling of Biological Processes with Petri Net-Based Architecture, pages 179–242. Springer Press.

Nagel, E. (1977). Goal-directed processes in biology. *The Journal of Philosophy*, **74**(5), 261–279.

Nicol, D. M. et Mao, W. Z. (1995). Automated parallelization of timed petri-net simulations. *Journal of Parallel and Distributed Computing*, **29**(1), 60–74.

Nketsa, A. et Khalifa, N. B. (2001). Timed petri nets and prediction to improve the chandy-misra conservative-distributed simulation. *Applied Mathematics and Computation*, **120**(1-3), 235–254.

Oliveira, J. S., Jones-Oliveira, J. B., Dixon, D. A., Bailey, C. G., et Gull, D. W. (2004). Hyperdigraph-theoretic analysis of the egfr signaling network : Initial steps leading to gtp :ras complex formation. *Journal of Computational Biology*, **11**(5), 812–842.

Peleg, M., Rubin, D., et Altman, R. B. (2005). Using petri net tools to study properties and dynamics of biological systems. *Journal of the American Medical Informatics Association*, **12**(2), 181–199.

Petri, C. A. (1962). *Kommunikation mit Automaten*. PhD thesis, Institut fur Instrumentelle Mathematik, Bonn.

Pinney, J. W., Westhead, D. R., et McConkey, G. A. (2003). Petri net representations in systems biology. *Biochemical Society Transactions*, **31**(Pt 6), 1513–1515.

Popova-Zeugmann, L., Heiner, M., et Koch, I. (2005). Time petri nets for modelling and analysis of biochemical networks. *Fundamenta Informaticae*, **67**(3), 149–162.

Prechelt, L. (2000). An empirical comparison of seven programming languages. *Computer*, **33**(10), 23–29.

Priami, C. et Quaglia, P. (2004). Modelling the dynamics of biosystems. *Briefings in Bioinformatics*, **5**(3), 259–269.

Qeli, E., Freisleben, B., Degenring, D., Wahl, A., et Wiechert, W. (2003). Metvis : a tool for designing and animating metabolic networks. *The European Simulation and Modelling Conference 2003*, pages 333 – 338.

Qeli, E., Wiechert, W., et Freisieben, B. (2004a). 3d visualization and animation of metabolic networks. *18th European Simulation Multiconference : Networked Simulations and Simulated Networks - ESM'2004*, pages 258 – 262.

Qeli, E., Wiechert, W., et Freisieben, B. (2004b). Visualization of sensitivity matrices generated during simulations of metabolic network models. *Proceedings of the IASTED International Conference on Applied Simulation and Modelling*, pages 583 – 588.

Qeli, E., Wiechert, W., et Freisieben, B. (2005a). Visual exploration of time-varying matrices. *Proceedings. Ninth International Conference on Information Visualisation*, pages 889 – 895.

Qeli, E., Wiechert, W., et Freisleben, B. (2004c). Visualizing time-varying matrices using multidimensional scaling and reorderable matrices. *Proceedings of the International Conference on Information Visualization*, **8**, 561 – 567.

Qeli, E., Wiechert, W., et Freisleben, B. (2005b). The time-dependent reorderable matrix method for visualizing evolving tabular data. *Proceedings of SPIE - The International Society for Optical Engineering*, **5669**, 199 – 207.

Reddy, V. N. (1994). Modeling biological pathways : A discrete event systems approach. Master's thesis, University of Maryland. 82 p.

Reddy, V. N., Liebman, M. N., et Mavrovouniotis, M. L. (1996). Qualitative analysis of biochemical reaction systems. *Computers in biology and medicine*, **26**(1), 9–24.

Reddy, V. N., Mavrovouniotis, M. L., et Liebman, M. N. (1993). Petri net representation in metabolic pathways. In Hunter, L., Searls, D., et Shavlik, J., éditeurs, *Proceedings of the First International Conference on Intelligent Systems for Molecular Biology (ISMB)*, pages 328–336. AIII Press.

Reisig, W. (1985). *Petri nets : an introduction*. Monographs on Theoritical Computer Science. Springer-Verlag, New York. 161 p.

Roberts, J. C. (1998). On encouraging multiple views for visualization. *Proceedings of the IEEE Conference on Information Visualization*, pages 8 – 14.

Robillard, P. N. et Hardy, S. (2001). Les protéines : des molécules de génie, rapport sur la recherche en bioinformatique, en protéomique et en génie protéique. Technical report, École Polytechnique de Montréal, Montréal. 105 p. EPM/RT-2001-04.

Rost, U. et Kummer, U. (2004). Visualisation of biochemical network simulations with simwiz. *IEE Systems Biology*, **1**(1), 184 – 189.

Sackmann, A., Formanowicz, D., Formanowicz, P., Koch, I., et Blazewicz, J. (2007). An analysis of the petri net based model of the human body iron homeostasis process. *Computational Biology and Chemistry*, **31**(1), 1–10.

Sackmann, A., Heiner, M., et Koch, I. (2006). Application of petri net based analysis techniques to signal transduction pathways. *BMC Bioinformatics*, **7**(1), 482.

Saez-Rodriguez, J., Kremling, A., Conzelmann, H., Bettenbrock, K., et Gilles, E. D. (2004). Modular analysis of signal transduction networks. *IEEE Control Systems Magazine*, **24**(4), 35 – 52.

Saez-Rodriguez, J., Kremling, A., et Gilles, E. D. (2005). Dissecting the puzzle of life : Modularization of signal transduction networks. *Computers and Chemical Engineering*, **29**(3), 619 – 629.

Sauer, U., Heinemann, M., et Zamboni, N. (2007). Genetics : Getting closer to the whole picture. *Science*, **316**(5824), 550–551.

Schmid, C. et Hinterberger, H. (1994). Comparative multivariate visualization across conceptually different graphic displays. *Proceedings of the Seventh International Working Conference on Scientific and Statistical Database Management*, pages 42 – 51.

Schmidt, H., Cho, K.-H., et Jacobsen, E. W. (2005). Identification of small scale biochemical networks based on general type system perturbations. *FEBS Journal*, **272**(9), 2141–2151.

Schmidt, H. et Jacobsen, E. W. (2004a). Identifying feedback mechanisms behind complex cell behavior. *IEEE Control Systems Magazine*, **24**(4), 91 – 102.

Schmidt, H. et Jacobsen, E. W. (2004b). Linear systems approach to analysis of complex dynamic behaviours in biochemical networks. *IEE Systems Biology*, **1**(1), 149 – 58.

Schrödinger, E. (1944). *What is life ?* Cambrigde University Press.

Shannon, C. E. (1949). The mathematical theory of communication. *The Bell System Technical Journal*, **27**, 379–423, 623–656.

Sivakumaran, S., Hariharaputran, S., Mishra, J., et Bhalla, U. S. (2003). The database of quantitative cellular signaling : management and analysis of chemical kinetic models of signaling networks. *Bioinformatics*, **19**(3), 408–415.

Sontag, E. D. (2004). Some new directions in control theory inspired by systems biology. *IEE Systems Biology*, **1**(1), 9 – 18.

Starke, P. H. (2003). INA : Integrated net analyzer [logiciel]. Berlin. http ://www2.informatik.hu-berlin.de/ starke/ina.html.

Steggles, L. J., Banks, R., Shaw, O., et Wipat, A. (2007). Qualitatively modelling and analysing genetic regulatory networks : a petri net approach. *Bioinformatics*, **23**(3), 336–343.

Stelling, J. (2004). Mathematical models in microbial systems biology. *Current Opinion in Microbiology*, **7**(5), 513–518.

Stelling, J. et Gilles, E. D. (2004). Mathematical modeling of complex regulatory networks. *IEEE Transactions on Nanobioscience*, **3**(3), 172 – 179.

Stelling, J., Gilles, E. D., et Doyle, F. J. (2004a). Robustness properties of circadian clock architectures. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(36), 13210–13215.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., et Gilles, E. D. (2002). Metabolic network structure determines key aspects of functionality and regulation. *Nature*, **420**(6912), 190–193.

Stelling, J., Sauer, U., Szallasi, Z., Doyle, F. J., et Doyle, J. (2004b). Robustness of cellular functions. *Cell*, **118**(6), 675–685.

Strogatz, S. H. (2001). Exploring complex networks. *Nature*, **410**(6825), 268–276.

Takai-Igarashi, T. (2005). Ontology based standardization of petri net modeling for signaling pathways. *In Silico Biology*, **5**(5-6), 529–536.

Takai-Igarashi, T. et Mizoguchi, R. (2004). Cell signaling networks ontology. *In Silico Biology*, **4**(1), 81–87.

Takano, K., Taoka, S., Yamauchi, M., et Watanabe, T. (2001). Experimental evaluation of two algorithms for computing petri net invariants. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, **E84-A**(11), 2871–2880.

Timmer, J., Muller, T. G., Swameye, I., Sandra, O., et Klingmuller, U. (2004). Modeling the nonlinear dynamics of cellular signal transduction. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, **14**(6), 2069 – 79.

Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T. S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J. C., et Hutchison III, C. A. (1999). E-CELL : software environment for whole-cell simulation. *Bioinformatics*, **15**(1), 72–84.

Troncale, S., Tahi, F., Campard, D., Vannier, J.-P., et Guespin, J. (2006). Modeling and simulation with hybrid functional petri nets of the role of interleukin-6 in human early haematopoiesis. *Pacific Symposium on Biocomputing*, **11**, 427–438.

Van Regenmortel, M. H. (2004). Reductionism and complexity in molecular biology. scientists now have the tools to unravel biological and overcome the limitations of reductionism. *EMBO Reports*, **5**(11), 1016–1020.

Verbeek, E. et van der Aalst, W. M. (2000). Woflan 2.0 a petri-net-based workflow diagnosis tool. In Nielsen, M. et Simpson, D., éditeurs, *Proceedings Application and Theory of Petri Nets 2000, Aarhus, Denmark, June.*, volume 1825 of *Lecture Notes in Computer Science*, Berlin, pages 475–484. Springer.

Voss, K., Heiner, M., et Koch, I. (2003). Steady state analysis of metabolic pathways

using Petri nets. *In Silico Biology*, **3**(3), 367–387.

Wegner, K. (2005). Simwiz3d - visualising biochemical simulation results. In *Medical Information Visualisation - Biomedical Visualisation, 2005. (MediVis 2005). Proceedings. Third International Conference on*, pages 77–82.

Wellstead, P. (2006). Schrödinger's legacy and the systemization of biology. Conférence de biologie systémique, École Polytechnique de Montréal.

Weng, G., Bhalla, U. S., et Iyengar, R. (1999). Complexity in biological signaling systems. *Science*, **284**(5411), 92–96.

Wiles, J. et Watson, J. (2005). Patterns in complex systems modeling. *Lecture Notes in Computer Science*, **3578**, 532 – 539.

Wilhelm, T., Behre, J., et Schuster, S. (2004). Analysis of structural robustness of metabolic networks. *IEE Systems Biology*, **1**(1), 114 – 20.

Wolkenhauer, O. (2001). Systems biology : the reincarnation of systems theory applied in biology ? *Briefings in Bioinformatics*, **2**(3), 258–270.

Wolkenhauer, O., Sreenath, S. N., Wellstead, P., Ullah, M., et Cho, K.-H. (2005a). A systems- and signal-oriented approach to intracellular dynamics. *Biochemical Society Transactions*, **33**(Pt 3), 507–515.

Wolkenhauer, O., Ullah, M., Kolch, W., et Cho, K.-H. (2004). Modeling and simulation of intracellular dynamics : Choosing an appropriate framework. *IEEE Transactions on Nanobioscience*, **3**(3), 200 – 207.

Wolkenhauer, O., Ullah, M., Wellstead, P., et Cho, K.-H. (2005b). The dynamic systems approach to control and regulation of intracellular networks. *FEBS Letters*,

579(8), 1846–1853.

Ye, X., Chu, J., Zhuang, Y., et Zhang, S. (2005). Multi-scale methodology : a key to deciphering systems biology. *Frontiers in Bioscience*, **10**, 961–965.

Zevedei-Oancea, I. et Schuster, S. (2003). Topological analysis of metabolic networks based on Petri net theory. *In Silico Biology*, **3**(3), 323–345.

ANNEXE I

PAINTED PETRI NET AND FUNCTIONAL ABSTRACTION TO VISUALIZE DYNAMIC MODELING

Abstract

The Painted Petri Net[1] and Functional Abstraction techniques are combined in a new visualization approach aimed at representing the dynamic behavior of simulated complex Petri net models. This approach consists in creating dynamic representations of a model at different levels of abstraction. At the first level, the Petri net model structure is displayed, and the model dynamics is represented by variations in the tints coloring the Petri net model places. The degree of abstraction can be progressively increased based on the organizational layers of the model which are defined according to the functionalities of the model elements. This process is called functional abstraction, and it is used here to create views where specific functionalities are isolated. The approach is illustrated with a model from systems biology: visualization of the dynamic behavior of a biochemical model obtained from the hybrid Petri net model of the CaMKII regulation pathway.

I.1 Introduction

The essence of systems biology is to develop a system-level understanding of biological systems. This is achieved by studying the structure and dynamics of these

[1]Cet article a été présenté à la *European Simulation and Modelling Conference 2006.* Voir la référence (Hardy et Robillard, 2006b)

systems, as well as their control mechanisms and design methods (Kitano, 2002b). With advances in the high-throughput technologies in the last decade, biology has evolved from a mostly experimental hypothesis-driven science to one with a more data-driven approach, and now relies also on a model-driven approach based on computer simulation. A tight coupling of the three approaches is at the heart of systems biology (Ye et al., 2005), and specific tools are needed for each approach. Computational methods have been developed or adapted to the modeling and simulation of biological systems (de Jong, 2002; Eungdamrong et Iyengar, 2004), and methodologies from systems and control theory have been successfully applied to the analysis of biological models (El-Samad et al., 2005; Schmidt et Jacobsen, 2004a; Stelling, 2004). As a result, biologists now have a system-level understanding of certain biological systems.

Among the methodologies available based on modeling and simulation, a few exploit information visualization and visual data mining to acquire system knowledge. (Keim, 2002) clearly stated the benefits of visual data exploration, which integrates human perceptual abilities in the interpretation of the data and is highly useful in the exploratory steps of data analysis. (Qeli et al., 2005a) developed such an approach to visualize time-varying matrices by combining multidimensional scaling and the reorderable matrix method. The objective of their static representation method is to analyze the data matrices of a simulated metabolic network in order to identify the parts of a model that could be simplified. (Rost et Kummer, 2004) developed a visualization software tool, SimWiz, which produces a dynamic representation of the simulation data of biochemical networks by presenting the data in a comprehensive way and by preserving the topology of the modeled system.

The Petri Net technique is a modeling approach used for biological modeling (Hardy et Robillard, 2004). Different types of Petri nets have been created, based on the orig-

inal theory, to increase the modeling expressiveness of this formal notation: timed, stochastic, colored and hybrid, to name a few. The Hybrid Petri Net approach combines variables of a discrete and continuous nature. A variant of this modeling approach, the Hybrid Functional Petri Net, was designed specifically for biological modeling and has been successfully used to study a number of different biological systems (Matsuno et al., 2003b, 2006a). In biological modeling, time-varying data are usually generated by an ordinary differential equation (ODE) model that is simulated. ODEs do not have a formal graphical representation. Since Hybrid Petri Net modeling has been demonstrated to be equivalent to a system of ODEs (Matsuno et al., 2000), it offers an interesting substitute for ODE modeling when a graphical representation is desired. The software that has been developed for the modeling and simulation of the Hybrid Functional Petri Net is the Genomic Object Net (GON) (Nagasaki et al., 2003).

However, the visualization of Petri net simulations and of the model dynamics by traditional means is arduous for large and complex models. Since the human mind has an incredible capacity to detect structures and patterns in images (Müller et Schumann, 2003), we have devised a visual dynamic representation of Petri net simulation data which integrates the structure and the dynamics of a system model and presents the model at different levels of abstraction. We call this the Painted Petri Net. The objective of our method is to provide insights leading to a system-level understanding of a dynamic model. This paper describes how this representational approach displays the model structure with the Petri net and makes use of relative values instead of absolute values to represent dynamic behaviors with paints. We also explain how our method can generate representations at different abstraction levels from the same simulation data with a modular and functional approach. Finally, we discuss the possible applications of our dynamic representation. Readers already familiar with the colored Petri net must not confuse that particular extension of Petri

net theory with the dynamic representation of systemic behavior presented in this paper. In the colored Petri net, coloration is an abstract concept which makes it possible to distinguish different data types inside a model. In the painted Petri net, tints or colors from a real palette are used to visually indicate the temporal variation of the modeled entities represented by places.

I.2 Visual dynamic representation of the Petri net system behavior

According to the Müller et Schumann (2003) taxonomy of methods of visualization of time-dependent data, the method we suggest is a dynamic representation of multivariate data with a continuous linear time axis. Being time-dependent means that the visual representation changes dynamically over time and is a function of time. Two aspects of the model representation have to be considered: how to represent the model *structure* or *topology* and how to represent the model *dynamics*. These two aspects correspond to the two data sources needed to implement this method: a file containing the organizational specifications of the model (Petri net elements and interactions) and a file containing the simulation data.

I.2.1 Representation of the model structure

The Petri Net is a modeling approach with a formal graphical representation (Reisig, 1985). The Petri Net model specifications are usually defined with the help of a graphical editor, with which the user draws the model structure by arranging Petri net elements in a drawing space. In this approach, three main elements are represented: places, transitions and arrows. In molecular biology modeling, places are associated with substances and are represented as circles; transitions are associated with chemical reactions and are represented as rectangles; and arrows link places

to transitions and represent the relation between substances and chemical reactions. The time-varying data of interest in a molecular biology Petri net model are the numerical values associated with places. Those values mainly represent substance concentrations or numbers of molecules. Figure I.1 shows a hybrid functional Petri net representation of the ODE model of the Ca2+ calmodulin-dependent protein kinase II (CaMKII) regulation pathway of Bhalla et Iyengar (1999). The structural elements animated in the dynamic representation are the places (painted circles). The Hybrid Petri net Net model was edited and simulated with GON. The CaMKII regulation pathway model is part of a larger network in the hippocampal CA1 neuron, which has already been thoroughly studied for its role in synaptic plasticity. However, its details and composition, as well as its simulation results, are outside the scope of this paper. The reader should consult the referenced paper for more information.

Other methods for visualizing Petri net dynamic behavior already exist. Three-dimensional visualization of the Petri net has been developed by Kindler et Pales (2004), but, in their work, a Petri net model represents real world objects: a train and a railway controller. Software is also available with GON, called Cell Animator, to create animations based on simulation data. For example, Cell Animator can be used to represent a cell with several enzymes and metabolites interacting with one another, their number and position varying according to the numerical values obtained from simulation. These animated representations are useful for rapidly communicating information to others who are unfamiliar with the model, particularly in a teaching context. In the type of problem that interests us, however, such representations are unsatisfactory or even impossible. A 3D-visualization of a model of cellular signaling pathways, for example, where the Petri net entities are associated with animated objects, is not feasible we are dealing with concentrations of diffused substances. An animation of moving molecules becomes quite confusing

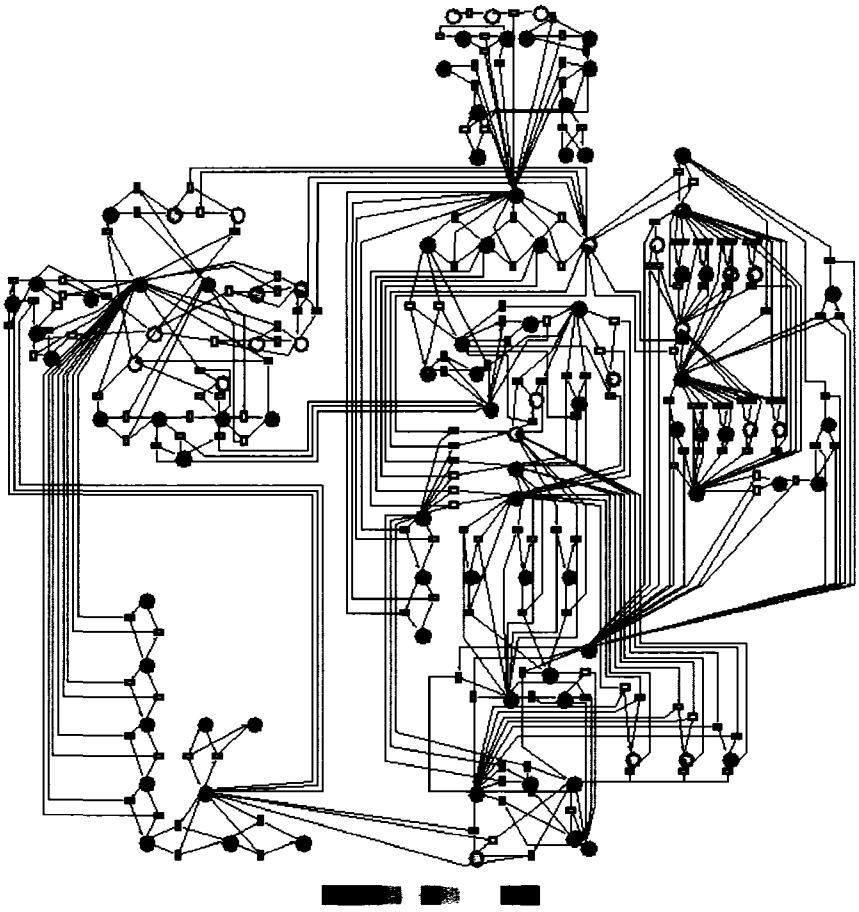


Figure I.1: A frame of the dynamic representation of a painted Petri net molecular biology model, simulation step 68 s

when tens, or even hundreds, of different substances are involved. More importantly, these animations do not incorporate the topology of the signaling pathways, and, as a result, crucial information about a model needed to reach a system-level understanding is lost. Keeping the Petri net structure for the dynamic representation is efficient for the modeler, as this individual is already accustomed to the model's abstracted elements and no transformation or rework is needed.

I.2.2 Representation of the model dynamics

We previously studied the modeling power of the Hybrid Functional Petri Net (Hardy et Robillard, 2005), and are now concerned with providing a single representation of the dynamic behavior of a complex system that incorporates all the simulation data. Such a comprehensive visualization tool is needed because traditional ways of representing dynamic data are sometimes too limited. Simulators produce plots and graphs for any variable of a mathematical model. For example, in molecular biology models, these are concentration graphs. When results are presented, only significant graphs illustrating important features of a system are displayed. However, if all the simulation data have to be explored, the modeler has to deal with one graph for each substance. This type of representation with graphs can rapidly become cumbersome in complex models involving many substances. For example, the model in Figure I.1 contains approximately one hundred substances. There are twice as many substances in a more complete model, adding other signaling pathways to the CaMKII regulation pathway to form a network (Bhalla et Iyengar, 1999), and there are five times as many substances in the latest enlarged version of the network model (Ma'ayan et al., 2005).

Color is commonly used in information visualization methods to improve their visual aspect. For our dynamic visual representation, we painted the simulation data using an 11-color palette (visible light spectrum: from violet to red). For the mapping, we favored a relative assignment of numerical values to colors rather than an absolute one, because the range of values for some concentrations is an order of magnitude smaller than the range of values of other concentrations. A relative assignment of tints gives an appreciation of the change for every substance. We believe that this advantage is greater than its disadvantage, which is that a tint does not represent the same quantity throughout the animated model. Numerical values are normalized

with equation I.1.

$$T_{i,t} = \left\lceil 10 \left(\frac{x_{i,t} - x_{i,min}}{x_{i,max} - x_{i,min}} \right) \right\rceil \tag{I.1}$$

In equation I.1, $T_{i,t}$ is the tint of the substance i at time t, $x_{i,t}$ is the concentration of the substance i at time t, and $x_{i,min}$ and $x_{i,max}$ are the minimum and maximum values of the concentration of substance i. The multiplication by ten and the application of the floor function return an integer value from 0 to 10, which corresponds to one of the 11 discrete tints of the palette. By applying this formula to the simulation data of a model, we obtain a tint for each substance at every simulation step (0 is mapped to violet and 10 is mapped to red). It is then possible to juxtapose the tints, which reflect the dynamic behavior of the model, to the model structure. This is done by painting the places of the Petri Net model, as shown in Figure I.1. The painted structure of the model at a given simulation step is one frame of the dynamic system representation. The Petri net model painted with the normalized concentrations for every time step results in successive frames that are assembled into an animation. This creates a movie-like representation of the Petri net dynamics.

I.2.3 Creation of abstraction levels to isolate specific functional activity and highlight system behavior

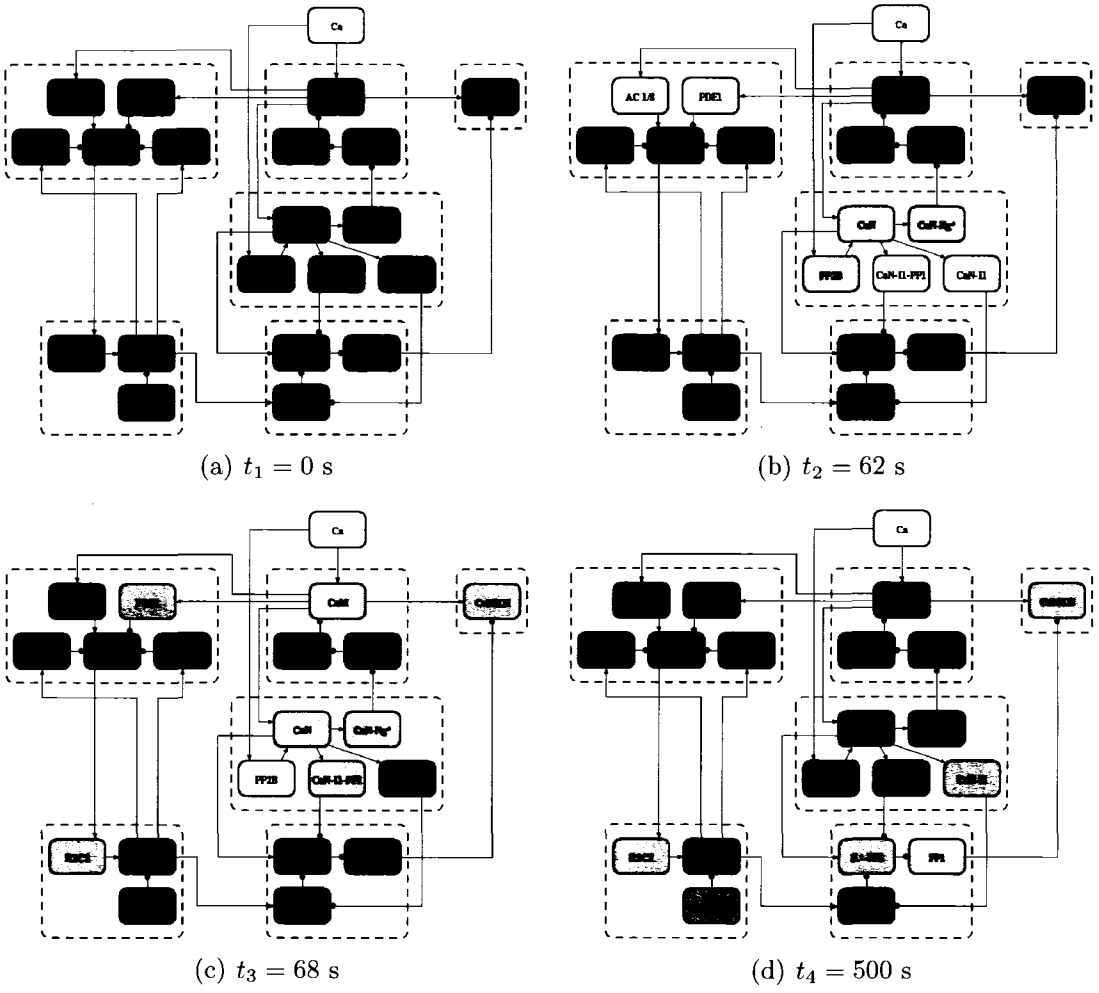
For large and complex models, there is a need to provide a hierarchy of abstraction levels in order to alleviate potential complexity pitfalls and to add systemic significance to the visualization. This hierarchy is created by recognizing the functional organization of the model. Figure I.2 depicts four frames of a representation of the CaMKII regulation pathway model at the second level of abstraction (Figure I.1 de-

picting the first level). To create this representation, several substances are grouped into new units according to their biochemical function. As we can see in Figure I.2, the Petri net model places are transformed into rounded rectangles linked by arrows. This view of the model has fewer elements and it is a coarse-grained abstraction of the network topology. The grouping into units was achieved with a detailed knowledge of the network's small components, which the modeler usually acquires during the modeling process. This grouping approach is similar to the ideas of modular cell biology of Hartwell et al. (1999), who argued for the recognition of functional modules as a critical level of biological organization in the cell. The grouping rules are rather simple. In the case model study in this paper, there are few different types of functional role. For example:

- a substance, or a group of substances, that assumes the role of a messenger molecule or an active enzyme;
- a substance, or a group of substances, that inhibits the activity of one or more other substances and thus participates in a negative feedback loop;
- a group of substances that relays an input activation to the production of an enzyme or a messenger molecule.

These three types of functional role are appropriate for the CaMKII regulation pathway model. Each application domain is likely to have some specific abstraction rules to define the model organization and appropriate abstraction levels.

The color palette of the model dynamics in the visual representation is similar throughout the abstraction levels, but the tint mapping is different. Since several substances can be combined into one unit, the normalization equation I.1 cannot be applied for abstraction levels other than the basic one. New equations are needed to



(a) $t_1 = 0$ s (b) $t_2 = 62$ s

(c) $t_3 = 68$ s (d) $t_4 = 500$ s

Figure I.2: Four frames of the dynamic representation of the functional units derived from a Petri net biological model at various moments of the simulation

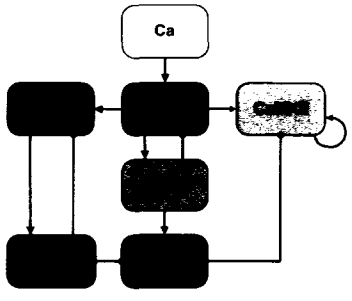


Figure I.3: A frame of the dynamic representation of the higher-level modules derived from a Petri net biological model, simulation step 68 s

map appropriate tints to the dynamic activity of the units. At the basic abstraction level, there is one concentration variable per equation. At higher levels of abstraction, the equation of a unit is composed of the concentration variables of the substances enclosed by the units. These new equations compute integer values from 0 to 10 that are associated with tints. With these equations, the tints no longer represent only a relative quantity on a min-max scale, but they are also designed to represent the functional activity level of the corresponding units.

In the model presented in this article, one organizational component is the module. Modules are pathway building blocks, and they are represented as dotted rectangles in Figure I.2. Combining several modules recreates a network. According to the grouping rules used at the second level of abstraction, substances acting together as an inhibitor inside a module are grouped into a unit, and substances acting as messenger molecules or enzymes are grouped into a unit as well, etc. Thus, the internal control structure of the units is preserved at the second abstraction level.

Figure I.3 shows a representation of the model at the third level of abstraction. At this level, there is only one unit for each module. Every substance in a module is abstracted by a single unit responsible for the module's main function. At this level of abstraction, details about the module's internal control mechanisms vanish and the representation granularity is coarser than at the previous abstraction level. The emphasis is on the interactions between modules. At this level of abstraction, the equations computing the tints represent the module activity level. For the CaMKII regulation pathway model, three levels of abstraction are enough to decompose the model into organizational layers. Any higher abstraction level would be useless for a model of this complexity, and additional abstraction levels could be useful for more complex models.

I.3 Discussion

The Painted Petri Net and Functional Abstraction techniques are the constituent parts of a new approach to visualizing the dynamic behavior of modeled systems. An example of this approach was given with the visualization of the molecular biological model of the CaMKII regulation pathway. In systems biology, modeling and simulation are used to study system behavior, and so this visualization approach can be useful in this area of research. However, we are extending the use of this visualization method to any application domain where complex Petri net models are used to study model dynamics by simulation. The Painted Petri net Net can be useful for modelers during the design of the model, as it can help rapidly provide them with a picture of the model model's dynamic behavior and help them find the source of problems or errors. Usually, Petri net simulators represent the dynamics of the model, with animations of discrete tokens being generated and consumed by the firing of discrete transitions or by displaying the real values of the markings of continuous places as continuous transitions are fired. Obtaining a global appreciation of the dynamics of a large Petri net model with these simulators is burdensome and inefficient, and the Painted Petri Net is a solution to this problem.

There are two advantages to the new views created by functional abstraction. First, by reducing the number of elements in the dynamic representation, the graphical display is simplified. This advantage will become more valuable as models increase in complexity. Second, grouping different substances linked to the same biochemical functionality adds a systemic knowledge filter that can facilitate the visualization of system behavior. The viewer can navigate between the different organizational layers of the model, thus helping him to refine his system-level understanding of the model. An important concept in systems biology is that of the emergent property, which is defined as a system-level characteristic that results from the complex and

nonlinear interactions of system elements and that cannot be predicted from what is known about these elements. In other words, we might understand the dynamics of the system elements taken separately, but when these elements are connected to form a complex network, it is impossible to predict the dynamics of the overall network. Emergent properties are nonintuitive, and systemic analysis is the main approach to understanding them. Indeed, the study of emergent properties was the main motivation behind combining the Painted Petri Net and Functional Abstraction techniques. By isolating specific functional activities and displaying their interactions in a dynamic representation, our visualization method can help a modeler detect and characterize the emergent properties of a model. Whether a network exhibits multistability, specific oscillatory patterns or synchronization, these complex properties may now be rapidly detected visually, revealing particular systemic properties of interest which would then be worth exploring with other analytical methods.

For this visualization approach, we favored grouping rules based on functional activity rather than rules based on other existing Petri net approaches like the hierarchical Petri net or structural reduction. Our main concern was to use the knowledge of the modeler that is specific to the application domain. However, in future studies, it will be interesting to verify whether or not these theoretical approaches can be combined with the hybrid Petri net, and to analyze what groupings the developed algorithms would identify. It would also be interesting to see whether or not the functional units identified by these theoretical approaches are identical to the functional units suggested by the application knowledge or if discrepancies appear.

Acknowledgement

This work was supported in part by NSERC grant A-0141 and an NSERC Postgraduate Scholarship.

ANNEXE II

VISUALIZATION OF THE SIMULATION DATA OF BIOCHEMICAL NETWORK MODELS : A PAINTED PETRI NET APPROACH

Abstract

The large quantity of data generated by the simulation[1] of complex biochemical models is difficult to interpret with traditional plots of time series data. A visualization method has been developed and software tools have been implemented. They are efficient with models of metabolic networks in which molecules are produced, consumed and degraded. However, models of signaling pathways in which enzymes are either activated or deactivated to transmit a signal are not handled well with the existing method. In this paper, we present a new visualization method using the painted feature of Petri nets and their invariant properties. We apply the first steps of this method to the continuous Petri net model of the calmodulin pathway. We also present the methodology that we will use to compare the existing visualization method with our Petri net-based method and to verify if it is more suited for signaling pathways.

II.1 Introduction

The visualisation of the simulation data of molecular biology models with traditional means – 2-dimensional plots of time series data – raises some problems when models

[1]Cet article a été présenté à la *Summer Computer Simulation Conference 2007.* Voir la référence (Hardy et Robillard, 2007c)

reach a certain level of complexity. These trajectories of concentrations are better interpreted than raw numbers, but it is increasingly confusing for a human user to make sense of relatively large quantities of data and substantial variables for different phenomena at the same time. Also, the topology of a biochemical model is not shown with these plots, thus there is an important loss of information when the networks are large and the topology cannot be memorized by the user. These problems were addressed by developing new visualisation software tools. MetVis (Qeli et al., 2003), SimWiz (Rost et Kummer, 2004) and BioPathwize [2] offer the possibility to visualize simulation data of metabolic networks in a more comprehensive way. Once a simulation is completed, these tools proceed similarly. The first step is to reconstruct the graph structure of the model from the differential equations. In this graph, the different nodes represent the various substances of the molecular model. The second step is to modify the graphical aspect of the graph nodes to show the variation of concentration, or of the number of molecules, obtained from simulation. The result is an animated view of the graph of the model displaying the simulation data. Depending of the software tool, the graphical aspect that changes to display the simulation data is either the color or the size of the nodes. Another feature of some tools is to display the variation of the fluxes between the molecular substances by modifying the graphical aspect of the edges linking the nodes of the graph. The tools MetVis and SimWiz have also a 3-dimensional version (Qeli et al., 2004a; Wegner, 2005).

The advantages of such a view are numerous. These tools take advantage of the benefits of visual data exploration, one of which is to integrate the human perceptual abilities in data interpretation, a useful asset in the exploratory steps of data analysis (Keim, 2002). They also give the possibility to make full use of the human mind capacity to detect structures and patterns in images, such as synchronicities, global changes of state and oscillations (Müller et Schumann, 2003). The developers

[2] http://bioanalyticsgroup.com

of MetVis and SimWiz insist on the usefulness of the topology of the biochemical networks for their animated graphical representation. To map the simulation data onto the model structure gives a good impression of the dynamical behavior of the system. Thus, a wealth of information is accessible to the user. However, a graphical representation of the models has to be generated since the models that are visualized with these tools are without any graphical nature. Since there is no standard for biochemical network graphical representation, each tool has its own format.

Petri net is a modeling language designed to study systems with causal concurrent properties. This modeling language has been used in molecular biology modeling for the first time by Reddy et al. (1993). Petri nets properties such as liveness, boundedness and reachability have been used to qualitatively analyze biological properties (Voss et al., 2003; Zevedei-Oancea et Schuster, 2003; Sackmann et al., 2006). The Petri net high-level extensions supporting simulation have been used to realize quantitative studies (Hofestadt et Thelen, 1998; Matsuno et al., 2003a, 2006a). For review articles on the subject, consult Hardy et Robillard (2004) or Matsuno et al. (2006b). A Petri net-based approach for biological modeling offers two main advantages. First, many theoretical elements of Petri nets with a mathematical basis are useful as a preliminary analysis tool for biological pathways. Second, biologists can easily model a biological system with Petri nets, partly due to its graphical nature, and then study it with the simulation capabilities of Petri net tools. In this paper, we start the exploration of a new kind of application of Petri nets for biological systems: the use of Petri nets for the visualization of the simulation data of biochemical models. We present a visualization method using the painted feature and the marking invariants of Petri nets. In this method, the marking invariant property of Petri nets is used to identify meaningful biological entities and to generate, with the painted feature of Petri nets, animated representations of the simulation data. Our hypothesis is that our method will highlight the switching behavior of enzymes during the visualization

of the simulation data of the models of signal pathways. In section II.2, we present the painted feature of Petri net and how we use it to reproduce the existing visualization method and to develop our new method. In section II.3, we explain how we plan to verify our hypothesis in future work.

II.2 Visualization of simulation data with the painted Petri net

Our visualization method uses a recent Petri net feature, the painted Petri net, first introduced in (Hardy et Robillard, 2006b). This feature is explained in the first subsection. Using this feature and a local relative painting rule, we describe in the second subsection how we can reproduce the existing visualization method, in order to compare it with our new visualization method. This method is based on marking invariants and it is presented in the third subsection.

II.2.1 Painted Petri net feature and a Petri net model of the calmodulin pathway

Several extensions of the Petri net theory have been used to complete quantitative analysis of different systems. The choice of the right extension depends of the nature of the studied system. However, no matter what the extension is, the data from the simulation of the Petri net model is generally communicated as raw numbers, directly on the model in the simulation environment during the simulation. Other tools also generate 2-dimensional histograms or plots. Unfortunately, it can be as confusing to make sense of the simulation data of complex Petri net models as for other formalisms with traditional means of data representation. To create a more intuitive and meaningfull way of representing the simulation data, we suggest the use of the painted Petri net feature to benefit from a more graphical representation.

The painted feature of Petri nets makes use of colors, called paints, that are applied on the places of a Petri net model, to give a visual indication of the evolution of the places content for the duration of the simulation of the model. The painted Petri net is not a standalone extension. This painted layer can be put on a Petri net model of any kind of extension used for simulation. The painted Petri net does not change any mathematical properties of the extension on which it is put on. The word paint is prefered to the word color to avoid any confusion with the colored Petri net. In the colored Petri net, coloration is an abstract concept, which makes it possible to distinguish different data types inside a model. In the painted Petri net, paints are used to visually indicate the temporal variation of the modeled entities represented by places. The paint of a place is determined for each time step of the simulation with a painting rule that uses the place's mark to generate a value from 0 to 1. The obtained numerical value is then associated to a paint from a spectrum (0 being blue at one end and 1 being red at the other end for example). The result is an animated graphical representation of the painted Petri net model that combines the topology of the model with the simulation data. Types of animated representations of a model, exposing different dynamical aspects of the model, can be created with different painting rules.

In this paper, the Petri net model of the calmodulin pathway is used to illustrate the first steps of the marking invariant-based visualization method. This model is part of a more complex model of the regulatory pathway of the Ca^{2+} calmodulin-dependent protein kinase II (CaMKII) in the hippocampal CA1 neuron. The original system was modeled with ordinary differential equations (ODE) by Bhalla et Iyengar (1999) and it has been converted into a Petri net model in a previous study (Hardy et Robillard, 2005). Figure II.1 shows a frame of the animated representation of the complete model painted using the local relative painting rule presented in the next subsection. The Petri net extension used to model this system has been specifically

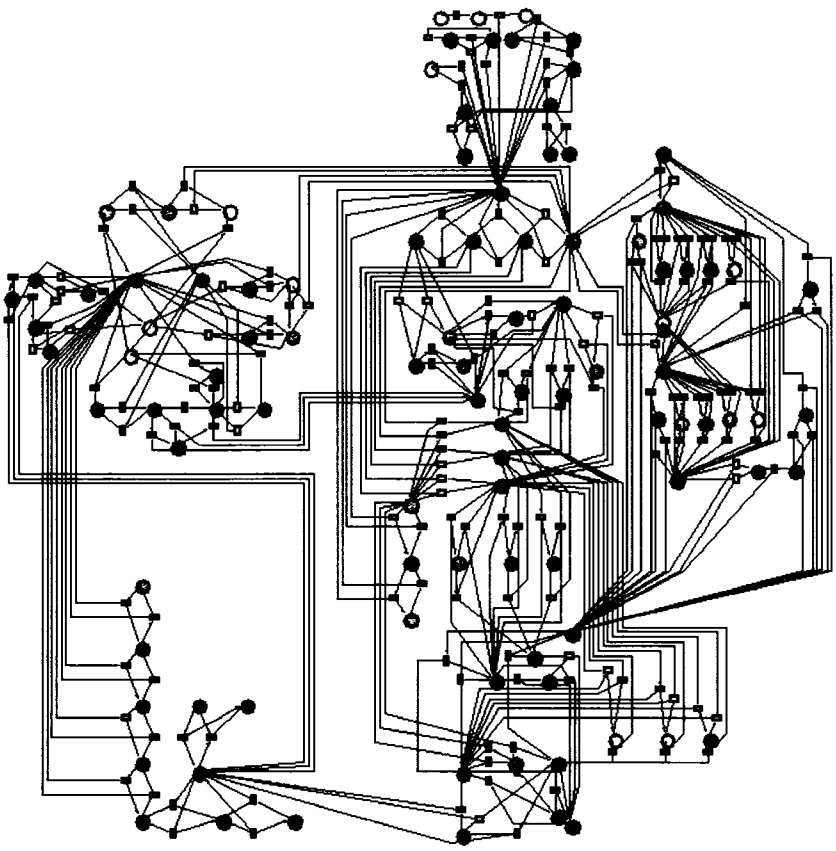


Figure II.1: A frame of the animated representation of the painted Petri net of the CaMKII regulation pathway model, simulation step 66 s

designed for biological modeling and simulation: the hybrid functional Petri net (HFPN) (Matsuno et al., 2003a). The HFPN extension enables the modeling of discrete and continuous elements and processes and the formation of three types of relationship between places and transitions with normal, inhibition and test arcs. A formal definition of the HFPN is given in Nagasaki et al. (2004). The software that has been developed for the modeling and simulation with HFPN is Genomic Object Net (GON), also commercially known as Cell Illustrator (Nagasaki et al., 2003).

The Petri net model of the calmodulin pathway is shown in Fig. II.2. All the elements of this Petri net model are continuous. Each place of the Petri net model (drawn as

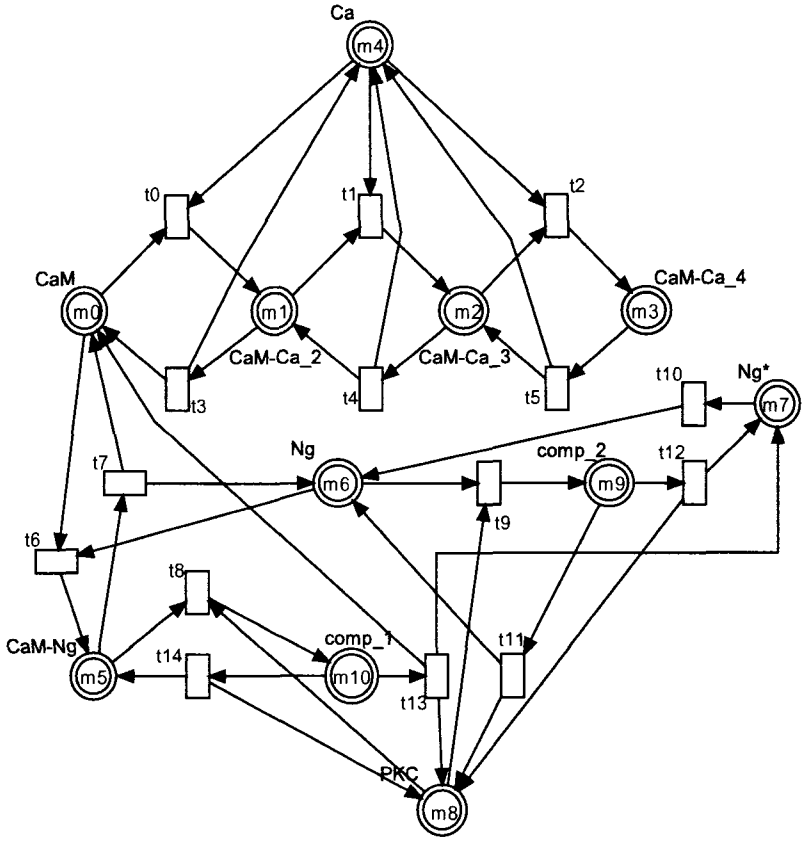


Figure II.2: Petri net model of the calmodulin pathway

circles) corresponds to a molecular substances. Each transition of the Petri net model (drawn as rectangles) corresponds to a chemical reaction. The directed arcs of the model indicate the relationships between the molecular substances and the chemical reactions: an arc from a place to a transition means that the substance modeled by the place is a substrate of the reaction; an arc from a transition to a place means that the substance is a product of the reaction. The places contain a mark, indicating the concentration of the substance. The transitions has a speed equation, used to compute their reaction rate. The tables II.1 and II.2 present the specifications of the places, transitions and arcs of the Petri net model of the calmodulin pathway. These specifications are given so that the reader can follow the steps of the identification of the marking invariants. We will use these invariants for a painting rule.

The calmodulin pathway model will react to an elevation of the calcium concentration. After such an elevation, the fully activated form of this calcium-binding protein (Ca_4.CaM) can then bind to other proteins (not shown in this model) to trigger a specific response. The calmodulin pathway model includes a regulatory protein, neurogranin (Ng), that binds to calmodulin in the absence of calcium. The phosphorylation of Ng by protein kinase C (PKC) lowers its binding ability. The remaining substances, $complex_1$ and $complex_2$, are two intermediary molecular complexes participating in enzymatic reactions caused by the presence of PKC and phosphorylating Ng (Ng* is the phosphorylated form of Ng).

Tableau II.1: Specifications of the places of the Petri net model of the calmodulin pathway

Name	Substance	Variable
P_0	CaM	m_0
P_1	CaM-Ca_2	m_1
P_2	CaM-Ca_3	m_2
P_3	CaM-Ca_4	m_3
P_4	Ca	m_4
P_5	CaM-Ng	m_5
P_6	Ng	m_6
P_7	Ng*	m_7
P_8	PKC	m_8
P_9	$complex_2$	m_9
P_{10}	$complex_1$	m_{10}

II.2.2 Visualization of simulation data with a local relative painting rule

The tools MetVis and SimWiz normalize the raw simulation data of a model into relative values by using the range between minimum and maximum values of each variable. This produces an individual percentage that can be transformed into a color or a scaling factor for visualization. The same data conversion into colors can

Tableau II.2: Specifications of the transitions and arcs of the Petri net model of the camodulin pathway

Name	Speed	From		To	
		variable	factor	variable	factor
t_0	$72 \times m_0 \times {m_4}^2$	m_0	1	m_1	1
		m_4	2		
t_1	$3.6 \times m_1 \times m_4$	m_1	1	m_2	1
		m_4	1		
t_2	$0.465 \times m_2 \times m_4$	m_2	1	m_3	1
		m_4	1		
t_3	$72 \times m_1$	m_1	1	m_0	1
				m_4	2
t_4	$10 \times m_2$	m_2	1	m_1	1
				m_4	1
t_5	$10 \times m_3$	m_3	1	m_2	1
				m_4	1
t_6	$0.3 \times m_0 \times m_6$	m_0	1	m_5	1
		m_6	1		
t_7	m_5	m_5	1	m_0	1
				m_6	1
t_8	$0.0612 \times m_5 \times m_8$	m_5	1	m_{10}	1
		m_8	1		
t_9	$0.101 \times m_6 \times m_8$	m_6	1	m_9	1
		m_8	1		
t_{10}	$0.005 \times m_7$	m_7	1	m_6	1
t_{11}	$2.32 \times m_9$	m_9	1	m_6	1
				m_8	1
t_{12}	$0.58 \times m_9$	m_9	1	m_7	1
				m_8	1
t_{13}	$0.35 \times m_{10}$	m_{10}	1	m_0	1
				m_7	1
				m_8	1
t_{14}	$1.4 \times m_{10}$	m_{10}	1	m_5	1
				m_8	1

be achieved for each place P_i at time t of a painted Petri net with the following painting rule:

$$paint_{P_i,t} = (mark_{P_i,t} - min_{P_i})/(max_{P_i} - min_{P_i}) \quad \text{(II.1)}$$

where $mark_{P_i,t}$ represents the mark of the place P_i at time t and max_{P_i} and min_{P_i} represent the maximum and minimum values reached by each place during the complete simulation. The application of the paints calculated with this painting rule onto a Petri net model produces an animated representation where each frame at time t shows the places of the model with their respective paint at time t.

A first disadvantage of this visualization method is that it is difficult to compare the paints of different places because of their relative nature. The minimum-maximum range can vary greatly from one place to another. A second disadvantage of this painting rule, mainly for signalling network model, is the confusion caused by too many color changes. For some substances, the presence in the pathway of many different intermediate molecular forms with a small minimum-maximum range introduces unwanted color flickering. A global painting rule (the same minimum value and maximum values for every place) instead of a local one would not be a good solution either because of the variation between the different orders of concentration in model. A potential improvement is possible by using some structural properties of the Petri net.

II.2.3 Visualization of simulation data with a painting rule using marking invariants

To visualize the concentration variations of a signaling pathway model without the disadvantages of a local relative method, we suggest to use the Petri net concept of marking invariant. In this subsection, we demonstrate that the sets of places defining

the marking invariants of a biochemical Petri net model, also known as conservative components in Petri net theory, are meaningful biological entities and that their properties justify their use in a painting rule. First, we formaly define these Petri net concepts and we present the common algorithm used to identify the conservative components of a Petri net model. Because Petri nets have a mathematical representation in linear algebra, it is possible to implement an automated procedure. Basic knowledge on Petri net theory is required to understand all the terms in the following paragraphs. Despite the jargon, we provide a comparaison with biochemical concepts that will help any reader to grasp their significance. After this more formal part of the text, we discuss the result of the algorithm for the calmodulin pathway model and present the marking invariant-based painting rule. We invite readers to consult (David et Alla, 2004) for a complete introduction to Petri net theory.

Définition II.1. The set B of places of a Petri net model is a *conservative component* if and only if a weighting vector $\mathbf{x}$ exists such that

$$P(\mathbf{x}) = B \text{ and } \mathbf{x}^T \cdot \mathbf{W} = \mathbf{0} \tag{II.2}$$

where $\mathbf{W}$ is the incidence matrix of the Petri net model. The vector $\mathbf{x}$ is a *P-invariant* ($\mathbf{x}^T$ is its transposed equivalent) and the set of places $P(\mathbf{x})$ is known as the *support* of the P-invariant $\mathbf{x}$.

A consequence of this definition is that for whatever firing sequence from marking $\mathbf{m}_0$, i.e. for every marking $\mathbf{m}_k \in \mathcal{M}(\mathbf{m}_0)$ (set of reachable markings from initial marking), we have

$$\mathbf{x}^T \cdot \mathbf{m}_k = \mathbf{x}^T \cdot \mathbf{m}_0 \tag{II.3}$$

which is a scalar quantity known as a marking invariant. This means that the sum of the marks of the set of places of a conservative component, weighted by vector $\mathbf{x}$,

is constant. This definition of the marking invariant applies in an equivalent way to discrete, continuous and hybrid Petri nets. In a biochemical network model based on chemical reactions and kinetic rates, a conservative component has a physical meaning: the places of a conservative component represent all the different forms of a molecular species for which the total concentration is constant. By forms, we mean the substances that are modified configurations of a molecular species (like Ng and the phosphorylated Ng* are different forms of neurogranin) or the substances that are complexes composed of at least one molecule of a molecular species (like CaM, $CaM.Ca_2$, $CaM.Ca_3$, $CaM.Ca_4$, and CaM.Ng are different forms of calmodulin).

We use algorithm II.1 to find the minimal support P-invariants of the calmodulin pathway Petri net model showed in Figure II.2. This algorithm can be found in (David et Alla, 2004). Minimal support P-invariants is the formal term for the smallest conservative components of a Petri net model.

Algorithm II.1. Search for minimal support P-invariants.

Step 1. Let $\mathbf{A}$ be the dimension n unit matrix (n = number of places) and $\mathbf{B} = \mathbf{W}$ (incidence matrix). Construct matrix $[\mathbf{A}|\mathbf{B}]$.

Step 2. For each index j of transition T_j.

Step 2.1. Add to matrix $[\mathbf{A}|\mathbf{B}]$ as many lines i as there are linear combinations of two lines, with positive integer coefficients, such that element(i,j) is zero.

Step 2.2. Eliminate from matrix $[\mathbf{A}|\mathbf{B}]$ all the line k whose element (k,j) is not zero.

Step 3. Let $l_A \cdot l_B$ denote a line l of the matrix $[\mathbf{A}|\mathbf{B}]$ and $P(l_A)$ the support of l_A (i.e., the set of places for which the weight is not zero). If there are two lines p and q of the matrix such that $P(p_A) \supseteq P(q_A)$, then line p is removed.

Step 4. The minimal support P-invariants correspond to the non zero lines of **A**.

□

Tableau II.3: The four marking invariants of the Petri net model of the calmodulin pathway

Substance	P-invariant Marking invariant
CaM	$\mathbf{x}_1 = [1\ 1\ 1\ 1\ 0\ 1\ 0\ 0\ 0\ 0\ 1]$ $N_{CC_1} = m_0 + m_1 + m_2 + m_3 + m_5 + m_{10}$
Ca	$\mathbf{x}_2 = [0\ 2\ 3\ 4\ 1\ 0\ 0\ 0\ 0\ 0\ 0]$ $N_{CC_2} = 2m_1 + 3m_2 + 4m_3 + m_4$
Ng	$\mathbf{x}_3 = [0\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 0\ 1\ 1]$ $N_{CC_3} = m_5 + m_6 + m_7 + m_9 + m_{10}$
PKC	$\mathbf{x}_4 = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 1\ 1]$ $N_{CC_4} = m_8 + m_9 + m_{10}$

The results of the search for the minimal support P-invariants of the Petri net model of the calmodulin pathway are shown in table II.3. Four conservative components have been found. This means that there are four distinct molecular specie in the model, which are calmodulin, neurogranin, calcium and protein kinase C. These conservative components correspond to the sets of places containing all the different forms of the four molecular species. For example, the strictly positive values of the P-invariant $\mathbf{x}_1$ indicate that the places P_0 (CaM), P_1 (CaM.Ca_2), P_2 (CaM.Ca_3), P_3 (CaM.Ca_4), P_5 (CaM.Ng) and P_{10} ($complex_1$) are associated to the different forms of the calmodulin species. More importantly, the sum of the marks of these places, weighted by the P-invariant values, is the total constant concentration of calmodulin molecules in the model. For example, the sum of $m_0 + m_1 + m_2 + m_3 + m_5 + m_{10}$ is a constant value throughout the simulation indicating the total concentration of all the forms of calmodulin in the system. This result shows that there is a relationship

between mass conservation and marking invariants in biochemical Petri net models. This relationship has already been discussed in previous studies in which the marking invariants were used to verify the soundness of a model (Zevedei-Oancea et Schuster, 2003; Sackmann et al., 2006). Here again, the marking invariants demonstrate that the model specifications and its structural properties respect the mass conservation law. This is not new information, it is the intention of the modeler to assure that mass conservation is respected in the model. The marking invariants validate the modeler's work. The novelty we propose in this paper is to use marking invariants in a painting rule for a better visualization of the dynamical behavior of a Petri net model. The marking invariant-based painting rule is the following:

$$paint_{P_i.CC_x.t} = mark_{P_i.t}/N_{CC_x} \tag{II.4}$$

where the place P_i is part of the conservative component CC_x. This rule computes the ratio between the concentration of one form of a molecular species and the sum of the concentrations of all forms of this molecular species. Since a place can be part of several conservative components, like the place CaM.Ng is part the calmodulin and the neurogranin conservative components in the calmodulin pathway model, a place can have x paint values, where x is the number of conservative component this place is part of.

The hypothesis to be tested next is if this painting rule based on conservative components and marking invariants computes paints that are more significant than the paints computed with the individual relative painting rule. The preliminary animated view of a Petri net painted according to this rule shows color modifications that are relevant to the evolution of the distribution between different forms of a molecular species. For many enzymes, this distribution is linked to their functional activation or deactivation. For example, the distribution of an enzyme in phosphorylated and

dephosphorylated forms can indicate if this enzyme is "on" or "off". This is also the case for other signaling macromolecules involved in the formation of complexes.

II.3 Concluding remarks

The creation of the animated view of the Petri net model of the CaMKII regulation pathway with a painting rule using marking invariants is actually a work in progress. The confirmation or not of our hypothesis – the improvement of the visualization of the simulation data of signaling pathway models– will be possible after a comparison of the animated views of the complex Petri net model painted with the local relative painting rule and painted with the marking invariant-based pating rule. We will also need to define how to deal with some expected special situations that are for the moment left out of the painting rule using marking invariants. One of these situations is how to visualize the places having several paints. This occurs when a place is part of more than one conservative component, like the place P_{10} of the calmodulin module model, which is part of the conservative components of the CaM, Ng and PKC molecular species. A second special situation is how to represent the places without paint. This occurs when a place is not part of any conservative component. In the complete CaMKII regulation pathway model, places without paint are associated to messenger molecules like Ca and cAMP that are produced and degraded or introduced and expelled during the simulation. The total concentration of these substances is not constant during simulation, but the evolution of their concentration is also important to visualize. A second complementary painting rule will probably be developed for places in this situation.

Acknowledgement

This work was supported in part by the grant A-0141 from the National Sciences and Engineering Research Council of Canada and a NSERC Postgraduate Scholarship.