

UNIVERSITÉ DE MONTRÉAL

MODÉLISATION ET SIMULATION DE SYSTÈMES BIOLOGIQUES À
L'AIDE DE RÉSEAUX DE PÉTRI

SIMON HARDY
DÉPARTEMENT DE GÉNIE INFORMATIQUE
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INFORMATIQUE)
OCTOBRE 2004



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-494-01337-0

Our file Notre référence

ISBN: 0-494-01337-0

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

MODÉLISATION ET SIMULATION DE SYSTÈMES BIOLOGIQUES À
L'AIDE DE RÉSEAUX DE PÉTRI

présenté par: HARDY Simon

en vue de l'obtention du diplôme de: Maîtrise ès sciences appliquées

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

Mme BOUCHENEB Hanifa, Doctorat, présidente

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

M. MICHNICK Stephen, Ph.D., membre

REMERCIEMENTS

Je voudrais tout d'abord remercier mon directeur de recherche, Pierre N. Robillard, qui de par sa curiosité intellectuelle et son instinct scientifique, m'a mis sur la piste de la bioinformatique. Sa direction fut remarquable pour m'initier au monde de la recherche. Il a toujours manifesté un intérêt et un enthousiasme grandissants pour mon projet, qui n'ont cessé de me motiver.

J'aimerais aussi remercier Julie Robillard qui m'a, à l'occasion, tiré hors du brouillard biologique par des explications claires et vivantes. C'est elle qui m'a suggéré la plasticité neuronale comme une avenue de modélisation intéressante.

Je ne serais vraisemblablement pas venu à bout de mon modèle moléculaire si ce n'avait été des nombreuses réponses que m'a données Masao Nagasaki. À plusieurs occasions, la correspondance que nous avons entretenue entre Tokyo et Montréal m'a été d'une grande aide.

J'aimerais également remercier Mihaela Dulipovici et Éric Germain pour leur camaraderie au Laboratoire RGL. Merci Mihaela pour tes petites attentions. Merci Éric pour le gabarit L^AT_EX et qui donne à mon mémoire des allures professionnelles. Pour leur soutien financier durant mes études, je remercie le Fonds québécois de la recherche sur la nature et les technologies ainsi que le Conseil de recherches en sciences naturelles et génie du Canada.

Je tiens à remercier mes parents ainsi que mes soeurs Isabelle et Fabienne, qui m'ont toujours encouragé dans mes études comme dans tous mes projets. Finalement, une pensée toute spéciale à Mahité, ma muse, qui aura dispensé mille et un rayons de soleil au cours de ma maîtrise et dont le sourire aura allégé les moments ardu.

RÉSUMÉ

La biologie moléculaire a connu d'importantes avancées au cours des dernières années et la course au séquençage du génome humain a marqué le début d'une révolution dans cette science. De plus, une grande quantité d'informations sur les macromolécules du vivant a été accumulée grâce à de nouvelles technologies. Dans ce contexte, une nouvelle approche scientifique a vu le jour : la biologie des systèmes. Celle-ci prône l'étude du fonctionnement des systèmes biologiques à l'échelle moléculaire, ce qui était peu fait auparavant faute de moyens techniques pour obtenir de l'information sur l'ensemble des molécules impliquées de ces systèmes et leurs interactions.

L'informatique a été appelée à jouer un rôle de plus en plus important dans la recherche en biologie et de nombreux logiciels ont été développés pour résoudre des problèmes complexes posés par la biologie. Ce domaine scientifique fut baptisé bioinformatique. La biologie des systèmes ne fait pas exception à ce besoin d'outils informatiques et on a vu apparaître plusieurs logiciels de modélisation et de simulation destinés à la biologie. Ceux-ci sont basés sur différentes représentations : les équations différentielles, l'analyse logique, les modèles stochastiques, etc. Le formalisme des réseaux de Pétri a aussi été identifié comme outil adéquat pour la modélisation et la simulation de systèmes biologiques et plusieurs équipes de recherche l'ont utilisé.

Dans un premier temps, nous avons étudié les différentes méthodes de modélisation et de simulation basées sur les réseaux de Pétri. Les extensions colorées, stochastiques, hybrides et fonctionnelles de ce formalisme ont été utilisées. Dans ce mémoire, ces méthodes sont toutes présentées et une analyse approfondie est faite pour les catégoriser selon leurs objectifs. Deux catégories d'objectifs de modélisation ont été distinguées : faire soit une étude qualitative pour identifier les propriétés d'un système, soit une étude quantitative pour analyser par simulation

la dynamique d'un système.

Dans un deuxième temps, nous avons modélisé deux vues de la propriété de plasticité neuronale à l'aide de l'extension hybride fonctionnelle des réseaux de Pétri. Celle-ci est à notre avis l'extension des réseaux de Pétri la plus appropriée pour la modélisation et la simulation en biologie. Dans ce mémoire, deux modèles sont présentés : le premier est une modélisation au niveau phénoménologique de la synapse et le deuxième est une modélisation au niveau moléculaire du réseau de régulation de la protéine CaMKII qui est impliqué dans la potentiation à long terme du neurone. Les résultats de simulation que nous présentons concordent avec ceux obtenus antérieurement avec les mêmes modèles, mais basés sur d'autres représentations. À notre connaissance, il s'agit de la première modélisation phénoménologique basée sur les réseaux de Pétri hybrides fonctionnels. De plus, il s'agit aussi du plus large modèle moléculaire représenté avec cette notation rapporté dans la littérature. Ces modélisations à différents niveaux d'abstraction explorent de nouvelles possibilités de représentation avec la notation. Nous proposons une nouvelle avenue de recherche où plusieurs niveaux d'abstraction peuvent être combinés dans un même modèle.

ABSTRACT

Molecular biology has rapidly evolved in the last years and the race for the human genome sequencing marked the beginning of a revolution in this science. An incredible amount of data on the macromolecules of the life was collected with the help of new technologies. In that context, a new scientific approach emerged: systems biology. This approach investigates the structure and the dynamics of biological systems at a molecular scale. This kind of study was not often done before because of the lack of techniques to discover and study all the systems molecules and their interactions.

Computer science played an increasingly important role in biological research and several software tools were designed to solve complex biology problems. This new scientific domain was called bioinformatics. Systems biology is no exception and researchers working with approach have needs for modeling and the simulation tools adapted to biological systems. A few were created using different representations: ordinary differential equations, logical analysis and stochastic models. Petri net formalism has also been identified as an appropriate tool for the modeling and the simulation of biological systems and many research teams have already used it.

In the first part of our research, we studied all the modeling and simulation methods based on Petri nets. Colored, stochastic, hybrid and functional extensions of Petri net formalism have been used for biological modeling. In this master's thesis, we present all these methods and an analysis of their modeling objectives. Two categories of modeling goals were identified: either to make a qualitative study to identify system properties or to make a quantitative study of system dynamics by using simulation.

In the second part of our research, we present models of the neuronal plasticity property represented with hybrid functional Petri net. In our opinion, this Petri net extension is the more appropriate for the modeling and simulation of biological

systems. Two models are presented in this master's thesis. The first one is a phenomenological model of the synapse and the second one is a kinetic model at the molecular level of the CaMKII protein regulation network. This pathway plays a role in neuronal long-term potentiation. We present simulation results similar to the ones described in previous studies on these models. To our knowledge, this is the first phenomenological modeling based on hybrid functional Petri nets and it is also the largest molecular model ever reported in literature. Furthermore, the modeling of different abstraction levels is an exploration of new possibilities of representation with this formalism. As a possible research development, we suggest a new type of modeling where several abstraction levels are combined into the same model.

TABLE DES MATIÈRES

REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vii
TABLE DES MATIÈRES	ix
LISTE DES FIGURES	xi
LISTE DES TABLEAUX	xiii
LISTE DES NOTATIONS ET DES SYMBOLES	xiv
INTRODUCTION	1
CHAPITRE 1 LA BIOINFORMATIQUE ET LES RÉVOLUTIONS DE LA BIOLOGIE MOLÉCULAIRE	5
1.1 La génomique	6
1.2 La protéomique	8
1.3 La biologie des systèmes	11
1.4 La bioinformatique	13
1.5 Impacts sur la biologie et la pharmaceutique	17
CHAPITRE 2 MODELING AND SIMULATION OF MOLECULAR BI- OLOGY SYSTEMS USING PETRI NETS: MODELING GOALS OF VARIOUS APPROACHES	19
2.1 Introduction	19
2.2 Elements of Petri net theory and the earliest attempts at biological modeling	22

2.3	Stochastic Petri nets	28
2.4	Colored Petri nets	31
2.5	Hybrid Petri nets and supplementary extensions	33
2.6	Discussion	38
2.7	Conclusion	43
CHAPITRE 3	PHENOMENOLOGICAL AND MOLECULAR LEVEL PETRI NET MODELING AND SIMULATION OF LONG- TERM POTENTIATION	44
3.1	Introduction	45
3.2	Hybrid functional Petri net modeling	47
3.3	Phenomenological model of the synapse	49
3.3.1	The model	50
3.3.2	Hybrid functional Petri net representation of the synapse model	51
3.3.3	Simulation results	53
3.4	Molecular model of the synapse	54
3.4.1	The model	55
3.4.2	Hybrid functional Petri net representation of chemical and enzymatic reactions	56
3.4.3	Simulation results	61
3.5	Discussion	62
3.6	Conclusion	64
DISCUSSION ET CONCLUSION		66
RÉFÉRENCES		70

LISTE DES FIGURES

Figure 2.1	Petri net transition	24
Figure 2.2	Petri net model of the combined metabolism of the glycolytic and pentose phosphate pathways.	26
Figure 2.3	Colored Petri net model of the glycolysis and pentose phosphate pathways of an erythrocyte cell.	32
Figure 2.4	Hybrid Petri net continuous transition	34
Figure 2.5	Graphical representations of the elements of hybrid functional Petri nets.	35
Figure 2.6	Hybrid functional Petri net model of the glycolytic pathway and lac operon gene regulatory mechanism.	37
Figure 3.1	Petri net transition	48
Figure 3.2	Graphical representations of the elements of hybrid functional Petri nets.	48
Figure 3.3	Hybrid functional Petri net representation of phenomenological synapse model.	52
Figure 3.4	Simulation results of HFPN phenomenological synapse model.	54
Figure 3.5	Block diagram of CaMKII regulation pathway.	56
Figure 3.6	Hybrid functional Petri net representations of a basic chemical reaction.	58
Figure 3.7	Hybrid functional Petri net representations of an enzymatic reaction	58
Figure 3.8	Hybrid functional Petri net representation of the seven modules of the CaMKII regulation pathway.	59
Figure 3.9	Hybrid functional Petri net representation of the calmoduline module (CaM) of the CaMKII regulation pathway.	60

Figure 3.10	Simulation results of enzyme activation from hybrid functional Petri net model of CaMKII regulation pathways. . . .	61
-------------	---	----

LISTE DES TABLEAUX

Tableau 2.1	Mapping between metabolites in the pathway and places in the Petri net models of Figures 2.2, 2.3 and 2.6.	27
Tableau 2.2	Mapping between reactions in the pathway and transitions in the Petri net models of Figures 2.2, 2.3 and 2.6.	28
Tableau 2.3	Biological modeling goals of Petri net extensions.	39

LISTE DES NOTATIONS ET DES SYMBOLES

AC1	Adenylyl cyclase 1
ADN	Acide désoxyribonucléique
ARNm	Acide ribonucléique messenger
ATP	Adenosine triphosphate
BioSPICE	Biological Simulation Program for Intra-Cellular Evaluation
BLAST	Basic Local Alignment Search Tool
CaMKII	Ca ²⁺ /calmodulin-dependant protein kinase II
cAMP	Cyclic adenosine monophosphate
CaN	Calcineurin
CPN	Colored Petri net
DDBJ	DNA Data Bank of Japan
EMBL	European Molecular Biology Laboratory
GON	Genomic Object Net software
HFPN	Hybrid functional Petri net
HPN	Hybrid Petri net
ISI	Irregular inter-stimulus intervals
LTD	Long-term depression
LTP	Long-term potentiation
NMDAR	N-Methyl-D-aspartate Receptor
NSERC	Natural Sciences and Engineering Research Council of Canada
ODE	Ordinary differential equation
PDE	Phosphodiesterase
PKA	Protein kinase A
PP1	Protein phosphatase 1

SPN

Stochastic Petri net

INTRODUCTION

Il se sera écoulé un peu moins de cinquante ans entre la découverte de la structure de l'ADN par Watson et Crick en 1953 et le séquençage complet du génome humain. La première fut une découverte majeure du 20^e siècle puisqu'elle décrivait enfin les molécules responsables de l'hérédité et du stockage de l'information génétique. Le deuxième représentait un travail colossal et ambitieux puisqu'il fallait décrypter les trois milliards de paires de base qui composent la double hélice d'ADN chez l'humain. Ces deux percées scientifiques sont des bornes significatives dans l'histoire la biologie et entre les deux, la recherche en biologie moléculaire a grandement évolué.

Parmi les progrès, une nouvelle discipline scientifique pour soutenir la conduite de la recherche en biologie a été consacrée : la bioinformatique. Ce jumelage entre la biologie et l'informatique a doté les équipes de recherche de nombreux outils de pointe qui mettent à profit les technologies de l'information pour la résolution de problèmes complexes posés par la biologie.

La bioinformatique s'est avérée essentielle pour les nouveaux domaines de recherche de la biologie moléculaire : la génomique, la science qui étudie la structure et le fonctionnement des génomes et en particulier les gènes ; la protéomique, science qui étudie des ensembles de protéines ; et la biologie des systèmes, science qui cherche à élucider les réseaux d'interactions moléculaires.

La naissance de la bioinformatique ainsi que le développement de logiciels destinés à la recherche en biologie moléculaire ont éveillé un intérêt de recherche au Laboratoire de recherche en génie logiciel (RGL) de l'École Polytechnique de Montréal. Un premier exercice exploratoire a été fait à l'été 2001 pour cerner les enjeux de la bioinformatique, pour en identifier les principaux acteurs et pour comprendre les

nouvelles filières de la biologie moléculaire (Robillard et Hardy, 2001). À l'été 2002, ce sont différents outils logiciel de bioinformatique qui ont été étudiés et testés.

Au terme de ces études, l'avenue de recherche qui a été choisie fut la modélisation et la simulation de systèmes biologiques. De plus, le Laboratoire RGL ayant déjà développé un simulateur pour processus d'affaires basé sur les réseaux de Pétri, nous avons alors sélectionné ce formalisme comme outil potentiel de modélisation biologique et les résultats de cette première recherche sont présentés dans le cadre de ce mémoire.

Il faut savoir qu'une approche systémique en biologie moléculaire, comme la biologie des systèmes le préconise, est assez récente. En fait, cette démarche est maintenant possible puisque la génomique et la protéomique ont généré les données expérimentales nécessaires à ce genre d'approche. Ideker et al. (2001a) ont noté un changement de paradigme : la recherche traditionnelle en biologie moléculaire, où une équipe de recherche étudie un ensemble restreint et isolé de molécules, cède sa place à la recherche sur le comportement et les interactions entre tous les éléments moléculaires d'un système biologique. Csete et Doyle (2002) ont invité les biologistes expérimentalistes à établir un dialogue avec les ingénieurs puisque les premiers pouvaient bénéficier de l'expérience des seconds à composer avec des systèmes complexes. Ils ont aussi associé le succès de la biologie des systèmes à l'utilisation d'outils de modélisation et de simulation provenant de l'ingénierie.

Les réseaux de Pétri ont été jugés par Peleg et al. (2002) comme le meilleur outil de modélisation biologique parmi onze types de modèles issus du génie logiciel, des affaires et de la biologie. Depuis les travaux innovateurs de Reddy et al. (1993), plusieurs équipes de recherche ont appliqué les réseaux de Pétri pour une étude des systèmes biologiques. En plus des réseaux de Pétri place/transition, les extensions colorées, stochastiques et hybrides ont été utilisées pour la modélisation et

la simulation des systèmes biologiques. Pour nous familiariser avec ces différentes méthodes, une revue complète de la littérature sur le sujet a été faite (Hardy et Robillard, 2003). Un premier article a été rédigé et nous y avons incorporé cette revue de littérature ainsi qu'une analyse des différents objectifs de modélisation suggérés pour chacune des approches de modélisation biologique basée sur les réseaux de Pétri. Cet article constitue une contribution originale puisqu'il s'agit du premier article de synthèse sur le sujet. Quelques chercheurs avaient déjà souligné l'absence de carte routière pouvant aider les biologistes à se familiariser avec les réseaux de Pétri et leurs applications en biologie moléculaire. Nous avons comblé ce vide. Cet article a été acceptée pour publication dans le *Journal of Bioinformatics and Computational Biology* (Hardy et Robillard, 2004a).

Cette première étude nous a permis d'identifier une extension des réseaux de Pétri qui a été conçue spécialement pour la modélisation biologique : les réseaux de Pétri hybrides fonctionnels (Matsuno et al., 2003c). Cette notation de modélisation est, à notre avis, la plus appropriée pour une étude quantitative des systèmes biologiques. Un outil de modélisation et de simulation biologique utilisant cette notation est disponible. Il s'agit du logiciel Genomic Object Net. Toutefois, nous avons jugé que cette notation était sous-utilisée et nous avons voulu démontrer des possibilités additionnelles de modélisation dans une deuxième étude.

Dans le cadre de cette deuxième étude, nous avons réalisé à l'aide des réseaux de Pétri hybrides fonctionnels deux modélisations de systèmes synaptiques impliqués dans la plasticité neuronale. Cette propriété, qui permet aux neurones de s'adapter lorsqu'ils sont soumis à certains stimuli, a été l'objet de nombreuses études en neurobiologie parce qu'elle est actuellement considérée comme le mécanisme moléculaire responsable de la mémoire et de l'apprentissage à l'échelle moléculaire. En raison de l'attention scientifique qui lui est portée, la plasticité neuronale a été maintes fois modélisée. Deux niveaux d'abstraction ont été utilisés pour cette modélisation :

le niveau phénoménologique, qui se base sur des observations expérimentales faites à partir d'échantillons, et le niveau moléculaire, qui cherche à reproduire des propriétés cellulaires à l'aide de réseaux d'interactions moléculaires. En complétant la modélisation et la simulation de systèmes à deux niveaux d'abstraction différents, notre étude suggère que les réseaux de Pétri hybrides fonctionnels disposent d'une flexibilité de modélisation qui ouvre la porte à la cohabitation de différents niveaux d'abstraction dans un même modèle. Cette étude a été soumise au journal *Biosystems* (Hardy et Robillard, 2004b).

Le premier chapitre de ce mémoire présente une introduction à la bioinformatique et aux nouveaux domaines de la biologie moléculaire. Celle-ci retrace la réflexion que nous avons faite au RGL et elle est inspirée des rapports techniques que nous avons publiés. Elle présente aussi des concepts de biologie de base qui seront utiles aux lecteurs peu familiers avec ce domaine. Le chapitre 2 présente le premier article inclus dans ce mémoire alors que le chapitre 3 reproduit le deuxième. Un chapitre unique de discussion et conclusion clot le mémoire.

CHAPITRE 1

LA BIOINFORMATIQUE ET LES RÉVOLUTIONS DE LA BIOLOGIE MOLÉCULAIRE

La bioinformatique est une discipline reliant les sciences de la vie aux technologies de l'information. Cette science interdisciplinaire est née de l'intégration des technologies de l'information à la biologie et se définit à la base comme l'utilisation de l'informatique pour résoudre des problèmes complexes posés par la biologie. Les efforts de la dernière décennie pour décoder le génome humain et celui de d'autres espèces n'auraient pas porté fruit sans un apport grandissant de la bioinformatique à l'analyse et à l'entreposage des données biologiques.

La bioinformatique est utile, voire même indispensable, aux différents domaines de la biologie, mais c'est en dotant la génomique d'outils de recherche que cette science interdisciplinaire a pris un essor considérable. Alors qu'au départ du projet du génome humain en 1988, on croyait que le séquençage complet se terminerait vers 2025, l'échéance a été maintes fois réévaluée et devancée jusqu'au 15 février 2001, où la carte du génome complétée à 97% a été publiée (International Human Genome Mapping Consortium, 2001). Cette tâche monumentale, qui a été accomplie en coopération internationale, n'aurait pu être terminée si rapidement sans le développement de la bioinformatique et la robotisation des laboratoires.

Pour le lecteur peu familier avec la biologie ainsi qu'à la bioinformatique, ce chapitre présente les questions et certaines réalisations scientifiques qui ont transformé la biologie moléculaire dans la dernière décennie. La mutation est profonde et nous distinguons trois étapes. Elles correspondent à trois domaines de la biologie moléculaire nés au cours de cette révolution : la génomique, la protéomique et la

biologie des systèmes. Ces domaines sont brièvement présentés et commentés dans les sections suivantes. Une autre section porte uniquement sur la bioinformatique et quelques problématiques biologiques qu'elle solutionne. Nous y présentons quelques logiciels à titre d'exemple. Nous terminons en discutant des impacts de cette révolution sur la biologie et sur la pharmaceutique.

1.1 La génomique

Le défi de la génomique est de lire le code de la vie qui est constitué par l'ADN. Langage universel du vivant, les lettres de l'ADN sont un groupe de 4 molécules appelées nucléotides. Ceux-ci sont formés d'une base azotée (adénine, cytosine, guanine ou thymine), d'un sucre et d'un groupement phosphate. Ces nucléotides s'alternent répétitivement pour former la structure spiralée de l'ADN. Dans le cas du génome humain, ce sont 3 milliards de nucléotides qui sont alignés en 23 paires de chromosomes et qui constituent le plan d'architecture de notre organisme. En effet, ce plan situé dans les noyaux de nos 60 000 milliards cellules est le maître d'oeuvre des activités cellulaires. Afin de le lire, il faut décomposer cette gigantesque série de nucléotides associés aux lettres A, C, G et T en mots, c'est-à-dire en gènes.

Repérer les gènes n'est pas une tâche aisée. Les régions codantes de notre ADN sont séparées par une grande quantité de nucléotides non codants, que l'on a longtemps considérés comme n'ayant aucune fonction biologique, mais dont on étudie maintenant les propriétés de régulation des gènes. Dans le noyau des cellules, les gènes fonctionnels activés sont transcrits en ARN messager (ARNm), des molécules chimiquement semblables à l'ADN. Ces brins d'ARNm quittent ensuite le noyau pour participer à la fabrication de protéines. À chaque triplet de nucléotides de l'ARNm est associé un des 20 différents acides aminés. Alors que l'information génétique de l'ARNm est décodée par le ribosome, ce dernier joint les acides aminés

appropriés en une longue chaîne peptidique qui formera une protéine précise. Les acides aminés sont aux protéines ce que les nucléotides sont aux gènes, c'est-à-dire des molécules élémentaires destinées à être liées en chaînes. C'est ainsi que les gènes sont responsables de la synthèse des protéines produites dans la cellule.

Pour identifier les gènes, les scientifiques ont recours à 2 méthodes bioinformatiques (Risch, 2000). Premièrement, ils comparent une séquence d'ADN avec des bases de données contenant des gènes antérieurement séquencés qui ont été découverts chez l'humain ou d'autres espèces. Deuxièmement, ils utilisent des algorithmes informatiques pour prédire l'emplacement de gènes inconnus. Ces gènes doivent ensuite être validés expérimentalement puisque les algorithmes ne sont pas fiables : ils peuvent à l'occasion omettre un gène existant ou en localiser un alors qu'il n'y en a pas. Par ces techniques, on a réduit le nombre de gènes humains, initialement estimé à 100 000, à 30 000.

Pour étudier le génome de différentes espèces, la recherche en génomique se répartit en quelques activités (Tsoka et Ouzounis, 2000). Comme il a été précédemment mentionné, on doit d'abord identifier les gènes à partir des séquences d'ADN. On compare aussi ces séquences entières à d'autres qui ont été recensées dans des bases de données afin d'y repérer des séquences homologues. Ces séquences homologues ont déjà été expérimentalement caractérisées et annotées de façon appropriée. Ceci permet de regrouper les gènes et les protéines en familles. Cette association de gènes permet une validation des observations faites sur le gène à l'étude et donne des pistes sur ses fonctions. Cette analyse complétée, il faut correctement annoter le gène étudié pour des références futures. Cette étape peut être décisive pour d'éventuelles recherches puisque les données annotées servent de base aux prochaines découvertes.

Déterminer correctement les fonctions d'un gène peut s'avérer l'exercice le plus périlleux de l'annotation (Vukmirovic et Tilghman, 2000). En effet, cet aspect de

l'analyse d'un génome n'est pas une analyse classique de séquence. Il faut être en mesure d'intégrer les produits exprimés par les gènes, les protéines, dans des réseaux de processus cellulaires dont font partie les voies métaboliques, les mécanismes de transcription et d'expression génétique ainsi que les cascades de contrôle intracellulaire (Lockhart et Winzeler, 2000).

Au tournant du siècle, le rythme de la recherche en génomique s'est grandement accéléré. L'automatisation du séquençage des génomes a mis à la disposition des chercheurs une très grande quantité de données qui sont depuis analysées. La fin du séquençage du génome humain a marqué la conclusion de la course au décodage des nucléotides de l'ADN. Les biologistes se sont alors tournés vers les protéines dans leur quête du savoir et une nouvelle discipline est née : la protéomique.

1.2 La protéomique

En étudiant l'ADN, les chercheurs ont vite compris que le séquençage du génome ne livrerait pas les secrets de la vie (Pandey et Mann, 2000). La génomique ne fait que lire le génome, elle n'en donne pas la compréhension. Les spécialistes se sont alors entendus pour dire que c'est la "post-génomique" qui apporterait ce savoir. La première discipline de la "post-génomique", la protéomique, s'intéresse aux protéines exprimées par les gènes. Les protéines ont une importance singulière puisqu'elles sont ultimement responsables de presque tous les processus se déroulant dans la cellule. Les applications de la protéomique sont nombreuses tant en médecine, en agrochimie, qu'en environnement et leur commercialisation est intéressante. Pour le moment toutefois, c'est en pharmacutique que la plupart des recherches en protéomique sont effectuées et d'énormes ressources y sont consacrées. Par le fait même, la protéomique a redéfini les méthodes de recherche de médicaments (Roses, 2000).

Si on compare les techniques utilisées, la quantité de données produites et la complexité des connaissances qui en découlent, la protéomique constitue un défi scientifique plus grand que la génomique. Les biologistes travaillant en génomique étaient déjà submergés d'informations, mais ce n'était rien par rapport au flot de données qui est généré par la protéomique. Alors que la génomique n'a que trente mille gènes à étudier chez l'humain, la protéomique s'attaque à plus de six cent mille protéines humaines différentes. De plus, l'analyse des protéines ne peut pas être aussi facilement automatisée que le séquençage du génome.

Les chercheurs qui oeuvrent en protéomique ont présentement plusieurs avenues. La première consiste à prédire ou déterminer la protéine qui est issue d'un gène. Ensuite, on cherche à connaître les conditions particulières qui mènent à l'expression des gènes et à la production de protéines spécifiques. D'autres tentent de caractériser les protéines, c'est-à-dire d'en connaître la structure et les fonctions. Finalement, une intégration des connaissances sur les protéines permet une reconstitution partielle des mécanismes biologiques et de pathologiques complexes se déroulant dans les organismes (Eisenberg et al., 2000).

Jusqu'à il y a une quarantaine d'années, les scientifiques pensaient qu'à un gène correspondait une seule protéine qui à son tour n'avait qu'une seule fonction. On sait maintenant que ce raisonnement est erroné. Désormais, on associe à un gène, n protéines qui ont n fonctions. Les rôles des protéines sont variés et indispensables au bon fonctionnement de la cellule. Ce sont des enzymes accélérant les réactions chimiques, des anticorps qui reconnaissent les molécules étrangères à l'organisme, des récepteurs qui informent la cellule et aussi des produits structurants qui solidifient les parois.

Il faut savoir que ce n'est pas à chaque cellule de notre corps qu'il revient d'exprimer la totalité des gènes. Des cellules de phénotypes différents, par exemple un neurone

ou un leucocyte, ne sollicitent pas les mêmes gènes et produisent donc des protéines distinctes. En conséquence, chaque cellule contient un ensemble protéinique caractéristique et c'est ce qu'on appelle le protéome. La technologie des puces à ADN s'avère utile pour déterminer les conditions d'expression d'un gène (Lockhart et Winzeler, 2000). Ces puces permettent justement de vérifier quel gène est exprimé par une cellule d'un type particulier lorsqu'elle est soumise à des conditions spécifiques. À l'aide de cette technologie, on peut comparer les protéomes de cellules saines et de cellules atteintes d'une pathologie à différents stades d'évolution et repérer des différences au niveau des protéines. Certaines protéines peuvent être en surplus ou en déficit et constituer alors d'éventuelles cibles thérapeutiques.

Lorsque la composition d'une protéine est connue, l'étape suivante consiste à découvrir ses fonctions (Attwood, 2000). Les meilleurs indices menant à l'identification des fonctions d'une protéine sont donnés par sa structure. Celle-ci permet de déduire les macromolécules avec lesquelles la protéine peut interagir et donc certaines réactions biochimiques auxquelles elle prend part. Cependant, il ne suffit pas de connaître la séquence des acides aminés pour déduire son repliement tridimensionnel. La séquence ne constitue que le premier niveau structurel d'une protéine et la biologie moléculaire en a identifié quatre. La modélisation de l'ensemble des interactions dans une protéine pour en connaître l'organisation spatiale est un domaine de recherche qui intéresse de nombreux chercheurs. Toutefois, cette modélisation est une tâche extrêmement complexe nécessitant une très grande puissance de calcul (Baker et Sali, 2001; Schonbrun et al., 2002; Kim, 2002).

Des techniques de corrélation peuvent être employées pour identifier les fonctions des protéines (Spengler, 2000). La méthode des arbres phylogéniques associe les fonctions de certaines protéines en vérifiant leur présence dans différentes espèces. Par exemple, si deux protéines sont présentes ou absentes simultanément dans le génome de deux espèces, on peut émettre l'hypothèse que ces protéines sont

impliquées dans le même complexe fonctionnel ou la même voie métabolique. Une autre méthode, la Pierre de Rosette, permet d'identifier des réactions entre paire de protéines. Si dans un organisme donné, on retrouve des gènes fusionnés, alors que dans un autre organisme, ces mêmes gènes sont séparés, alors il est fort probable que les protéines provenant des deux gènes séparés interagissent.

En résumé, la protéomique est dédiée à la caractérisation des protéines constituant différents protéomes. Les scientifiques cherchent à connaître la séquence, la structure et les fonctions de ces macromolécules. Toutefois, malgré que les avancées de la biologie moléculaire aient permis l'accumulation de nombreux faits biologiques, comme les séquences d'ADN ou les propriétés de protéines, la génomique et la protéomique ne suffisent pas pour comprendre le fonctionnement des systèmes biologiques. Cette compréhension peut provenir, entre autres, de l'analyse des mécanismes de régulation et des interactions gènes-protéines, de l'identification des cascades de réactions biochimiques et des circuits biologiques. En d'autres mots, comprendre la dynamique des systèmes biologiques.

1.3 La biologie des systèmes

Les biologistes organisent la compréhension du vivant par système. Il est habituel de diviser le fonctionnement des organismes en différents modules à l'intérieur duquel différents éléments coopèrent pour accomplir une fonction commune. Par exemple, le système digestif à l'échelle de l'individu ou alors le système de synthèse des protéines à l'échelle cellulaire. Toutefois, devant le manque d'informations et la complexité des systèmes, qui sont entre autres dus au grand nombre de composés et d'interactions, cette compréhension était impossible à l'échelle moléculaire. Cependant, la génomique et la protéomique ont permis l'acquisition de données qui permettent maintenant de s'attarder au comportement et aux relations entre tous

les éléments d'un système biologique. L'étude des gènes et des protéines isolés a cédé sa place à cette nouvelle approche que l'on nomme biologie des systèmes, ou *systems biology* (Ideker et al., 2001a).

Kitano (2002b) a identifié quatre propriétés clés permettant une compréhension des systèmes. Premièrement, la structure des systèmes, qui est définie par les réseaux de régulation des gènes et les voies biochimiques. En second, la dynamique des systèmes, c'est-à-dire l'étude du comportement du système lorsqu'il est soumis à différentes conditions. Ensuite, les mécanismes de contrôle qui modulent l'état du système. Et finalement, des méthodes de design de systèmes pour modifier et construire des systèmes biologiques ayant les propriétés recherchées autrement que par un essais-erreur effectué à l'aveugle.

Avec la naissance de cette nouvelle façon de voir en biologie moléculaire, de nouvelles méthodes biologiques de recherche et d'analyse ont été proposées (Ideker et al., 2001b). Puisqu'il est maintenant possible de mesurer le niveau d'activation de tous les gènes grâce aux puces à ADN et le niveau d'expression et d'interaction des protéines, ces méthodes prônent l'intégration des analyses génomiques et protéomiques dans le but de bâtir des modèles biologiques. Ces modèles permettent aux chercheurs d'émettre des hypothèses et de prédire le comportement cellulaire. Ces hypothèses peuvent ensuite être vérifiées expérimentalement.

Une approche par système en biologie moléculaire introduit dans cette science des concepts de la théorie de la complexité. Csete et Doyle (2002) suggèrent que les propriétés, de robustesse, de fragilité, de modularité et de rétroaction, qui sont bien connues des ingénieurs, sont présentes dans les systèmes biologiques et ils avancent que les nouveaux projets de recherche devraient en tenir compte.

1.4 La bioinformatique

La bioinformatique est composée de deux branches distinctes (Kitano, 2002a). La première consiste essentiellement en de la fouille de données. Il s'agit de concevoir des logiciels qui permettent l'extraction de connaissances à partir de données expérimentales. La deuxième branche consiste à faire des analyses basées sur la simulation, donc tester des hypothèses *in silico* pour obtenir des prédictions et ensuite les vérifier *in vivo* ou *in vitro*. Dans les prochains paragraphes, nous présentons les principales bases de données en biologie moléculaire et certains outils manipulant les séquences génétiques et les structures protéiques. Aussi, nous présentons différents environnements de modélisation développés pour la biologie de systèmes.

La révolution en biologie moléculaire n'aurait pu avoir lieu sans la bioinformatique. La quantité de données générées nécessitait une grande capacité de stockage et une bonne organisation de l'information. Par exemple, l'archivage de la séquence complète des nucléotides du génome d'une seule personne nécessite 50 teraoctets (Butte, 2001). Ainsi, on a vu se mettre sur pied plusieurs grandes bases de données publiques, qui servent de référence à l'ensemble de la communauté scientifique : GenBank pour le continent américain (Benson et al., 2004), la banque du European Molecular Biology Laboratory (EMBL) à Heidelberg en Allemagne pour l'Europe (Kulikova et al., 2004) et la DNA Data Bank of Japan (DDBJ) pour l'Asie-Pacifique (Miyazaki et al., 2004). Ces banques regroupent principalement des données sur des séquences génomiques. D'autres bases de données concernent les protéines : par exemple UniProt (Apweiler et al., 2004) qui fait partie du projet du protéome humain (O'Donovan et al., 2001). Accessibles par Internet, ces banques riches en informations sont parcourues quotidiennement par différentes équipes de recherche à travers le monde. Avec l'avènement de la biologie des systèmes, des banques contenant d'autres types de données sont apparues : la Database of Interacting

Proteins (Salwinski et al., 2004) et BIND (Bader et al., 2003) contiennent des informations sur les interactions protéine-protéine ; TRANSFAC (Wingender et al., 2001) répertorie les interactions protéine-ADN ; KEGG (Kanehisa et al., 2004) et EcoCyc (Karp et al., 2002) cataloguent les voies métaboliques.

Les découvertes doivent être cataloguées afin d'être réutilisées, donc les séquences et les structures contenues dans les bases de données doivent être annotées. Par annotation, on signifie rattacher à une séquence ou une structure les diverses informations reliées à l'état des connaissances sur celles-ci. On peut vouloir connaître l'organisme dont elles proviennent, les fonctions qui leur sont attribuées, les pathologies qui leur sont associées, etc. Le méta-logiciel de recherche Entrez (Schuler et al., 1996) du National Computational Biology Institute des États-Unis est un exemple d'intégration d'informations. La plupart des bases de données de l'Institut sont interreliées, incluant celles des structures tridimensionnelles des protéines, des génomes complets de plusieurs organismes et des références bibliographiques de publications scientifiques.

Comme il a été mentionné précédemment, une grande partie des analyses des nouvelles séquences d'ADN se fait en les comparant avec d'autres séquences qui proviennent de ces banques de données. Ceci permet de trouver des séquences similaires menant à l'identification de gènes. La même méthode peut être employée avec les séquences de protéines pour potentiellement obtenir des informations sur des motifs structuraux si des protéines similaires ont déjà été répertoriées et ainsi disposer de pistes sur leurs fonctions. Certaines compagnies offrent des logiciels gratuits en ligne permettant d'accomplir ces analyses. Parmi ceux-ci, l'outil Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) utilise la méthode des mots pour rechercher une séquence de nucléotides à travers différents génomes. D'autres logiciels ont aussi été développés pour l'analyse de séquences. Certains font une estimation de la similarité entre deux séquences, d'autres détectent les répétitions

ou des motifs particuliers.

Au-delà de la séquence, nous savons aussi que la structure des protéines est riche d'information. C'est pourquoi les bioinformaticiens ont développé des logiciels qui effectuent des analyses tridimensionnelles des protéines. Parmi ceux-ci, notons le logiciel DALI (Holm et Park, 2000) qui permet d'aligner deux structures protéiques, en d'autres mots les superposer, ce qui peut révéler d'intéressantes similarités architecturales qui ne peuvent être détectées lors de la comparaison de leur séquence. D'autres logiciels d'analyse de structure moléculaire peuvent calculer la géométrie interne et les propriétés physicochimiques des protéines. L'étude des liaisons hydrogène dans une structure, le calcul des surfaces accessibles au solvant et la détermination de forces électromagnétiques sont des exemples de propriétés évaluées par certains logiciels.

L'avènement de la biologie des systèmes a aussi suscité le développement de logiciels spécialisés pour faire la modélisation et la simulation de systèmes biologiques. De tels modèles, lorsqu'ils sont suffisamment détaillés, peuvent servir de référence et de guide pour interpréter les résultats expérimentaux ainsi que constituer un outil pour l'élaboration de nouvelles hypothèses de recherche (Takahashi et al., 2002). La simulation de modèles de larges réseaux biochimiques a aussi permis d'identifier des propriétés systémiques inapparentes lors de l'observation de réactions isolées (Bhalla et Iyengar, 1999). Ideker et al. (2001a) sont d'avis que l'utilisation de modèles basés sur des données quantitatives amorce une transformation de la biologie. D'une science descriptive qu'elle était, elle devient une science prédictive.

Inspiré de la façon traditionnelle en biochimie de représenter et de simuler les réactions chimiques, il était naturel que les premiers logiciels de modélisation emploient les équations différentielles. Le logiciel Gepasi (Mendes, 1997) est basé sur cette approche et il permet de modéliser de petits systèmes biologiques. E-Cell

(Tomita et al., 1999; Takahashi et al., 2003) est un logiciel qui peut modéliser et simuler de plus larges systèmes, en compilant des substances et des réactions avec des équations différentielles. E-Cell offre aussi la possibilité à l'utilisateur d'intégrer des programmes en C++ pour les réactions impossibles à représenter avec des équations différentielles.

Afin d'aplanir certaines difficultés d'utilisation reliées à une approche mathématique, des chercheurs ont développé le logiciel de modélisation et de simulation biologique Genomic Object Net (GON) (Nagasaki et al., 2003). Cet outil propose une modélisation basée sur les réseaux de Pétri hybrides fonctionnels (Matsuno et al., 2003c). Ces concepteurs ont considéré deux critères pour que leur outil soit accepté par les biologistes qui font de la modélisation de systèmes : (1) Retirer tout ce qui ne concerne pas la biologie afin d'éviter que les usagers n'aient à être formés en mathématique, en physique et/ou en informatique. (2) Que la représentation des systèmes soit intuitive et naturelle et que les mécanismes de simulation soient faciles à comprendre. Ce logiciel a été utilisé pour les modélisations et les simulations rapportées au chapitre 3.

Il existe aussi un mouvement de développement de logiciels libres nommé Biological Simulation Program for Intra-Cellular Evaluation (BioSPICE) (Kumar et Feidler, 2003) qui encadre la création de modèles informatiques, d'outils et d'infrastructures destinés à des problèmes biologiques de grande complexité. La communauté de BioSPICE regroupe des chercheurs provenant de l'informatique, de la biologie cellulaire, des mathématiques, de la biologie moléculaire, de l'intelligence artificielle et de la neuroscience et leur objectif est d'offrir aux biologistes des outils combinant des techniques informatiques, mathématiques et du génie pour caractériser, prédire et contrôler les dynamiques cellulaires.

Parmi d'autres outils disponibles, notons le logiciel Virtual Cell (Loew et Schaff,

2002) qui permet la modélisation et la simulation de modèles en deux ou trois dimensions où l'aspect de compartimentalisation spatiale est important ainsi que CADLIVE (Kurata et al., 2003) qui est un logiciel pour éditer de cartes à grande échelle de voies complexes de signalisation cellulaire.

1.5 Impacts sur la biologie et la pharmaceutique

La bioinformatique et les dernières révolutions de la biologie moléculaire ont donné aux chercheurs la possibilité de modifier leur approche réductionniste de recherche en biologie. Auparavant, les études en biologie se faisaient en observant et en analysant les parties d'un tout : on a décortiqué les organes pour comprendre le fonctionnement du corps humain comme on a subdivisé la cellule en organites. Avec la compréhension des mécanismes de l'ADN, il est impossible de réduire davantage, du moins du point de vue des sciences biologiques. À présent, ce sont des relations entre les découvertes, les molécules, les mécanismes et les interactions qu'il faut faire. On cherche à comprendre exactement comment, à partir de l'ADN, un organisme se bâtit, évolue et réagit.

De la même façon, les techniques de recherche de médicaments pourront en partie délaisser la méthode empirique d'essais-erreurs pour adopter une méthode plus scientifique, basée sur une meilleure compréhension des interactions moléculaires. La démarche des compagnies pharmaceutiques sera plus prédictive, les cibles thérapeutiques clairement identifiées et les coûts de développement moindres. De plus, les effets secondaires d'une thérapie pourront être diminués. Présentement, pour un médicament mis en marché, 10 000 substances chimiques ont été testées en laboratoire. Ce taux de succès est appelé à grandement s'améliorer.

Un autre domaine de recherche découlant des récentes découvertes attire l'atten-

tion : la pharmacogénomique (Roses, 2000). Ce domaine veut miser sur l'avancement de la génomique et de la protéomique pour développer des traitements individualisés. Le protéome de chacun est unique puisqu'il est à la fois influencé par le génome et l'environnement de l'individu. Pour le moment, on développe des médicaments généraux pour l'ensemble de la population. Toutefois, si on considérait leur protéome, des traitements pourraient s'avérer plus efficaces pour certains individus alors que ces mêmes médicaments pourraient s'avérer toxiques pour d'autres personnes ayant un protéome différent. Actuellement, les médecins disposent de plusieurs médicaments pour une même maladie. Dans certains cas, les patients doivent essayer différents remèdes avant de sentir les effets escomptés. Ceci est dû au fait que certaines protéines de l'organisme qui sont impliquées dans une maladie peuvent être légèrement différentes entre individus. En conséquence, certaines thérapies seront inopérantes (et parfois nocives) alors que d'autres mèneront à la guérison à cause de ces petites différences. La pharmacogénomique permettra donc de prescrire avec précision le traitement le plus prometteur pour une personne selon ce qui la distingue des autres, son protéome.

CHAPITRE 2

MODELING AND SIMULATION OF MOLECULAR BIOLOGY SYSTEMS USING PETRI NETS : MODELING GOALS OF VARIOUS APPROACHES

Abstract

Petri nets¹ are a discrete event simulation approach developed for system representation, in particular for their concurrency and synchronization properties. Various extensions to the original theory of Petri nets have been used for modeling molecular biology systems and metabolic networks. These extensions are stochastic, colored, hybrid and functional. This paper carries out an initial review of the various modeling approaches based on Petri net found in the literature, and of the biological systems that have been successfully modeled with these approaches. Moreover, the modeling goals and possibilities of qualitative analysis and system simulation of each approach are discussed.

2.1 Introduction

The completion of the human genome sequencing project, the rapid development of bioinformatics and the phenomenal accumulation of biological data have made the understanding of biological processes and cellular functions a growing research interest. This post-genomic wave is called "systems biology", and with it has arisen

¹Cet article a été accepté par la revue *Journal of Bioinformatics and Computational Biology* et sera publié dans le volume 2, numéro 4.

a greater interest in the modeling and simulation of biological systems (Kitano, 2002a). Many formalisms from the fields of biology, mathematics and the computer sciences are used to integrate, represent and analyze the vast amount of biological data.

A traditional representation uses ordinary differential equations (ODEs) to model biological systems. It is widely used and many tools are based on this approach (Heinrich et Schuster, 1998). Appropriate for representing and simulating the kinetic equations of biochemical reactions, ODEs are mostly used to study the dynamics of a metabolic process. The software packages Gepasi (Mendes, 1993) and E-CELL (Tomita et al., 1999) have been developed to support modeling with this analytical representation. Boolean logic and state machines are also used for modeling biological systems. Even though boolean models are simple representations, they can be used to express characteristics of biological phenomena (Kauffman, 1993). For example, a logic model of the *Endo16* gene of *Strongylocentrotus purpuratus* predicted an internal *cis*-regulatory switch (Yuh et al., 2001). Stochastic models are another approach used in molecular biology modeling. Processes like genetic expression and regulation, signal transduction and cellular reproduction exhibit a stochastic behavior (Goss et Peccoud, 1998). A stochastic kinetic analysis of the phage λ lysis-lysogeny decision circuit resulted in statistics on regulatory outcomes (Arkin et al., 1998). Many literature reviews presenting these modeling approaches and others have been published, and some books introducing the subject are available (de Jong, 2002; Endy et Brent, 2001; Bower et Bolouri, 2001; Collado-Vides et Hofestadt, 2002).

Discrete event simulation is another interesting approach for molecular biology modeling. Categorized in this last family of approaches, Petri nets can serve to model, analyze and simulate biological processes. The use of Petri nets in biology was suggested for the first time by Reddy et al. (1993), who qualitatively analyzed

metabolic pathways. Since then, several types of biological processes have been modeled and simulated with Petri nets, mainly molecular biology systems, but also in epidemic and ecologic modeling (Bahi-Jaber et Pontier, 2003; Gronewold et Sonnenschein, 1998; Sharov, 1991). Peleg et al. (2002) assessed Petri nets and ten other types of model from the fields of software engineering, business and biology to evaluate their appropriateness for representing and simulating biological processes. Their conclusions were that the combination of two of the assessed types of model, workflow and Petri net models, was the most suitable notation. Aptness of Petri nets for biological research is also demonstrated in recent articles (Chen et Hofestädt, 2003; Matsuno et al., 2003a,c; Oliveira et al., 2003; Voss et al., 2003; Zevedei-Oancea et Schuster, 2003). Furthermore, software tools for molecular biology modeling and simulation based on a Petri net architecture are being developed (Matsuno et al., 2001; Nagasaki et al., 2003; Doi et al., 2003b).

Despite the various works with different Petri net approaches, Chen et al. (2002) emphasized that "they lack unity in their concepts, notations and terminologies. This makes it very difficult for new scientists to understand the potential applications of Petri nets due to the various interpretations presented by different authors." The aim of this paper is to analyze the modeling of biological systems with the various types of Petri nets. For the complete methodology underlying each approach, the referenced articles should be consulted in their entirety. In the next section, elements from Petri net theory and the earliest attempts at using them for modeling are introduced. Then, the modeling and simulation of biological process with stochastic, colored and hybrid Petri nets are presented and the characteristics of each approach are discussed. The glycolysis pathway, modeled with three different types of Petri net, is illustrated. The formal definition of each Petri net type and the definitions of some of their properties are included with each presentation. Petri nets have also been used to analyze metabolic pathway databases (Küffner

et al., 2000). In this kind of application, Petri nets are useful for comparing data from various sources, but this is not within the scope of this paper.

2.2 Elements of Petri net theory and the earliest attempts at biological modeling

Petri nets were introduced at the beginning of the sixties by Prof. Carl A. Petri as a mathematical modeling tool to express system properties like concurrency, indeterminism, communication and synchronization. The basic Petri net is also called a place/transition net. It is founded on a mathematical formalism of oriented graphs. Petri nets and their changes of state can respectively be transposed into matrices and matricial operations. Petri nets are a network where tokens located in places will initiate transitions according to given conditions that will result in the generation of new tokens in the output places. Petri nets and their various elements have a standardized graphical representation as shown in Figure 2.1: places p_1 , p_2 and p_3 are circles, transition t_1 is a full rectangle, individual tokens are full dots and arcs are drawn as arrows where a positive integer indicates their weight. The number of tokens in each place is usually associated with variables, which are m_1 , m_2 and m_3 in the figure. Places can contain tokens which move from place to place via transitions. Tokens are not distinguishable from one another. An arc always binds a place to a transition. Definition 1 is the formal definition of a Petri net.

Definition 1. The Petri net N is defined by the n-tuple $(P, T, Pre, Post, M)$ where:

$P = \{p_1, p_2, \dots, p_u\}$, a finite set of places where $u > 0$;

$T = \{t_1, t_2, \dots, t_v\}$, a finite set of transitions where $v > 0$;

$P \cap T = \emptyset$, meaning that the sets P and T are disjoint;

$Pre = P \times T \rightarrow \mathbb{N}$, is the input incidence mapping (weights of the arcs going from places to transitions) and where $\mathbb{N}$ is the set of natural numbers;

$Post = P \times T \rightarrow \mathbb{N}$, is the output incidence mapping (weights of the arcs going from transitions to places);

$M = P \rightarrow \mathbb{N}$, is the marking of the net which is a vector of u components $(m_1, m_2, \dots, m_u)$, where m_i is the number of tokens contained in the place p_i . M_0 is the initial marking.

The marking M of the network gives the state of the Petri net. It is a vector indicating the number of tokens $M(p)$ at each place p . When a transition is fired, there is a change in the state of the net, and consequently a modification of the marking. A firing can occur when all the input places of a transition contain the minimal number of tokens defined by the Pre relation. In other words, when the number of tokens of all input places is greater than or equal to the weight of the arc linking them to a transition, this transition can fire. Then, the tokens are consumed by the transition and withdrawn from the input places, just as other tokens are created and added to the output places of the same transition. The number of tokens created is specified by the $Post$ relation.

Figure 2.1 illustrates a Petri net discrete transition. In this transition, places p_1 and p_2 are input places and place p_3 is an output place of the transition t_1 . The token contents of places p_1 , p_2 and p_3 are m_1 , m_2 and m_3 respectively. The weight constants 1 and 3 on the arcs going out of places p_1 and p_2 and the value 1.0 attached to transition t_1 mean that t_1 can fire if $m_1 \geq 1$ and $m_2 \geq 3$, and that the firing delay is 1 time unit (in the case of a timed net). When t_1 is fired, one token is removed from p_1 , three tokens are removed from p_2 and two tokens are added to p_3

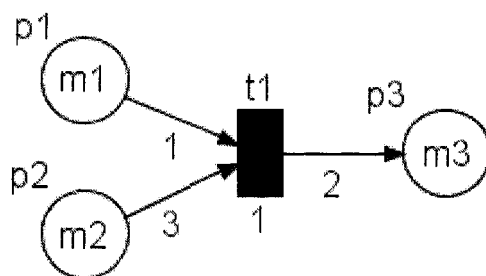


Figure 2.1 Petri net transition

For a more formal and complete coverage of traditional Petri nets and an analysis of their structural properties, consult Reisig's introduction on the subject (Reisig, 1985).

Many extensions have been added to the initial model, the purpose of which can be to transform models into a more compact form, to elevate the abstraction level or to give Petri nets new capabilities. Some of them have been used for modeling and simulation in biology, and these will be briefly presented in the following sections. The similarities between modeling in molecular biology and Petri net theory are thoroughly discussed (Zevedei-Oancea et Schuster, 2003).

Traditional Petri nets were originally suggested for biological pathway modeling by Reddy et al. (1993), and the bridging of molecular species and chemical reactions with Petri net places and transitions was achieved for the first time by them. The association of places with molecular species and transitions with chemical reactions is used for all types of Petri net model presented in this review. However, special situations necessitate more than one place for one species, for example, when distinguishing between an enzyme in an activated or a deactivated state, or a metabolite in various sites of the cell. The number of tokens indicates the quantity of substance and it corresponds to a predefined measure unit according to the scale of the model, such as the exact number of molecules, mole, millimole, etc. Reddy

et al. (1993) demonstrated that the Petri net approach was an appropriate tool for a preliminary qualitative analysis of biopathways. Behavioral and structural properties of Petri nets, like liveness, boundedness and invariants were used to identify some characteristics of models (see Definitions 2 to 6).

Definition 2. *Reachability.* A marking M is reachable if it can be reached from the current marking M_i in a finite firing sequence.

Definition 3. *Boundedness.* A place is bounded with bound k , if the token count does not exceed k for any reachable marking M of the net. A Petri net is k -bounded if each place is k -bounded.

Definition 4. *Liveness.* A transition is potentially firable if there exists a sequence of transition firings leading to a marking in which the transition is enabled. A transition is live if it is potentially firable for all reachable marking. A transition is dead if it is not potentially firable at the marking M ; so if the Petri net enters marking M , the dead transition cannot fire any more.

Definition 5. *S-invariant.* If C is the incidence matrix corresponding to the result of $Post - Pre$, then S-invariants are the solutions to the equation $Cy = 0$. The non-zero entries in the vector y constitute the set of places whose total token count does not change with any firing sequence. It is a conservation rule.

Definition 6. *T-invariant.* If C is the incidence matrix corresponding to the result of $Post - Pre$, then T-invariants are the solutions to the equation $C^T x = 0$, $x \geq 0$. The solution vector x is the set of transitions that have to fire, from some marking M , to return the Petri net to the same marking M . It is a regenerative rule.

This analysis approach was applied to the erythrocyte pentose phosphate pathway and to the main glycolytic pathway (see Figure 2.2 for model and Tables 2.1 and 2.2 for symbols definition) (Reddy et al., 1996). In Figure 2.2, the pentose phosphate

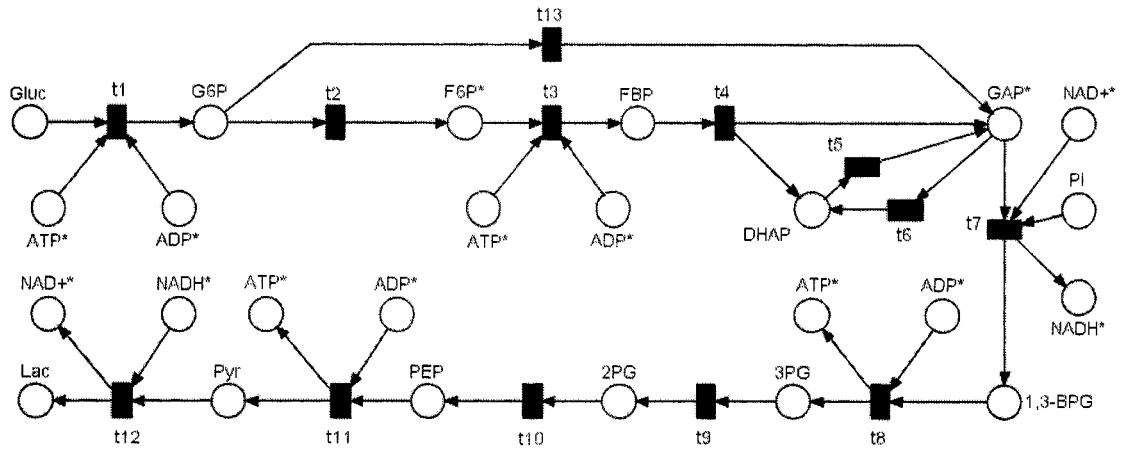


Figure 2.2 Petri net model of the combined metabolism of the glycolytic and pentose phosphate pathways.

pathway is abstracted by transition t_{13} and places labeled with an asterisk are, in fact, one place only, which has been divided for reasons of clarity. ATP, for example, has only one place. This model is inspired after Figure 8 of Reddy et al. (1996). The analysis of these pathways showed boundaries for certain molecular species, conservation properties, regenerative reactions and situations leading to a deadlocking of the system.

Hofestädt (1994) modified Petri net formalism in order to adapt it to metabolic network modeling. The objective was to model metabolic processes in a natural way by developing a new formal graphical representation based on Petri nets. Places and transitions were specialized: places are either metabolite or enzyme, and transitions are either biosynthesis, cellular communication or protein biosynthesis. Unfortunately, the advantage given by this increase in modeling power is of no use for solving the simulation issues that arose at that moment. Later work of Hofestädt et Thelen (1998) suggested a solution to this problem: the application of a Petri net extension corresponding more to the biological context. This extension is the self-modified Petri net, first defined by Valk (2003) and now known as the functional Petri net (see Definition 7). The main feature of this augmented formalism

Tableau 2.1 Mapping between metabolites in the pathway and places in the Petri net models of Figures 2.2, 2.3 and 2.6.

Abbreviation	Metabolite/compound	
	Marking variable associated to concentration	Name
Gluc	m_1	Glucose
G6P	m_2	Glucose-6-phosphate
F6P	m_3	Fructose-6-phosphate
FBP	m_4	Fructose biphosphate
DHAP	m_5	Dihydroxyacetone phosphate
GAP	m_6	Glyceraldehyde-3-phosphate
1,3-BPG	m_7	1,3-Biphosphoglycerate
3PG	m_8	3-Phosphoglycerate
2PG	m_9	2-Phosphoglycerate
PEP	m_{10}	Phosphoenolpyruvate
Pyr	m_{11}	Pyruvate
Lac	m_{12}	Lactate
ATP	m_{13}	Adenosine triphosphate
ADP	m_{14}	Adenosine diphosphate
NAD+	m_{15}	Nicotinamide adenine dinucleotide, oxidized form
NADH	m_{16}	Nicotinamide adenine dinucleotide, reduced form
Pi	m_{17}	Orthophosphate, ionic form

is the possibility of assigning to Petri net arcs an equation using marking variables instead of a positive integer. The result is that network marking dynamically modifies the weight of the arcs. In the case of a biochemical process model, this feature is particularly useful for simulating the influence of the variation in concentration on the kinetic rate of biocatalytic reactions. In other words, concentrations, represented by the number of tokens in the net, are variables for the functions that define the weight of the arcs. Thus, the reaction rate of a transition is modified according to the concentration of the various substances involved. Hofestädt and Thelen suggested that quantitative simulations with self-modified Petri nets would

Tableau 2.2 Mapping between reactions in the pathway and transitions in the Petri net models of Figures 2.2, 2.3 and 2.6.

Index	Enzyme/Reaction
t_1	Hexokinase
t_2	Phosphoglucose isomerase (forward reaction)
t_3	Phosphofructokinase
t_4	Aldolase
t_5	Triosephosphate isomerase (forward reaction)
t_6	Triosephosphate isomerase (reverse reaction)
t_7	Glyceraldehyde-3-phosphate dehydrogenase
t_8	Phosphoglycerate kinase
t_9	Phosphoglycerate mutase
t_{10}	Enolase
t_{11}	Pyruvate kinase
t_{12}	Lactate dehydrogenase
t_{13}	Pentose phosphate pathway abstraction
t_{14}	Phosphoglucose isomerase (reverse reaction)

help detect metabolic bottlenecks in some defective processes.

Definition 7. The functional Petri net N is defined by the n-tuple $(P, T, Pre, Post, V, M)$ where:

$(P, T, Pre, Post, M)$ is a Petri net as described in Definition 1;

$V = \{g_a(m_1, \dots, m_u), a \in Pre \cup Post \mid g : p_1 \times \dots \times p_u \rightarrow \mathbb{N}\}$, a set of functions assigned to arcs of the net using its marking $(m_1, \dots, m_u)$ as parameters.

2.3 Stochastic Petri nets

The random nature of molecular interactions at low concentration has been observed in several experiments. However, the Kolmogorov equations of the stochastic model corresponding to a biological system rapidly become impossible to solve ana-

lytically. Goss et Peccoud (1998) used stochastic Petri nets (SPN)(Ajmone Marsan et al., 1991) as a tool for biological modeling of stochastic models. They implied that the Petri net formalism and its modeling power can reduce model implementation delays. With their model, they successfully analyzed the stabilizing effect of the ROM protein on the genetic network controlling the replication of ColE1 plasmid replication (Goss et Peccoud, 1999). More recently, the response of transcription factor σ^{32} to a heat shock and the intracellular kinetics of a viral invasion have been studied through simulation of SPN models (Srivastava et al., 2001, 2002).

In the SPN model of a system composed of molecular interactions, each place corresponds to a particular molecular species. Tokens represent molecules and transitions between places are chemical reactions involving reactants (input places) and products (output places). At any time, the marking of the system indicates the number of molecules of each species involved. The values of arcs originating from input places and ending at output places are the equivalent of stoichiometric coefficients. As in traditional place/transition nets, if the number of tokens at input places is higher than the weight of all the input arcs of a transition, this transition can fire. In molecular terms, the firing of a transition means that a chemical reaction is occurring. The particularity of SPN is that the firing of a transition is not instantaneous. There is a delay following a probabilistic distribution, thus the delay is a random variable (see Definition 8). In SPN biological models, this delay is interpreted as the reaction rate, and it is given by the weight function of the corresponding transition. The delay mean time is obtained by the transition reaction rate, which is a function of a stochastic rate constant and the quantity of each molecular species involved as a reactant or a catalyst. This constant takes into account volume, temperature, pH and other environmental factors. It is also related to the deterministic rate of the reaction. Several types of SPN exist, but, in the type discussed here, the same type that is used by Goss and Peccoud for modeling

a biological process, the weight function will take into account the marking of the net in order to correctly calculate reaction rates. When the number of molecules is sufficiently large, the stochastic constant of the reaction rate is equal to the deterministic rate.

Definition 8. The stochastic Petri net N is defined by the n-tuple $(P, T, Pre, Post, F, \lambda, M)$ where:

$(P, T, Pre, Post, M)$ is a Petri net as described in Definition 1;

$F = \{F_t, t \in T \mid F_t : [0, \infty) \rightarrow [0, 1]\}$, a set of probability density functions for the net firing delays. Their average is 1 and they are independent of the marking;

$\lambda = \{\lambda_t, t \in T \mid \lambda_t : \mathbb{N} \rightarrow \mathbb{R}^+\}$, a set of firing rates, which are function of the marking (a set of natural integers) and where each element is associated with a transition t . This rate, a positive real number from the set $\mathbb{R}^+$, is used to calculate the probability density function for the transition t .

Stochastic models are applicable when molecules are considered as a discrete amount. Then, a deterministic change in concentrations, quantified by reaction kinetic rates of the incessant flux of transformation of reactants in products, becomes a random event where reactions are ruled by probabilistic laws. SPN can help build these models from their reaction equations and simulate them. It consequently becomes possible to study a system with the simulation results.

The software Möbius (or UltraSAN in its earlier version) has been used for all the systems modeled with SPN mentioned in this section (Performability Engineering Research Group, 2004). This SPN simulation tool – not exclusive to biology – also has a model numerical resolution option. With this tool, the molecular species distribution can be studied and the occurrence probability of certain events can

be calculated. For example, in the analysis of the stabilizing effect of the ROM protein on a genetic network, the probability that a cell will divide without having replicated its plasmid was estimated.

2.4 Colored Petri nets

The differentiation between categories of tokens when modeling large systems with Petri nets was considered in order to reduce the size of models. Thus, Petri nets were enhanced with this new feature by adding colors. The resulting high-level net, colored Petri net (CPN), is composed of tokens identified by a color (see Definition 9). With this augmentation of the formalism, it is possible to represent, in the same model, different dynamic behaviors modeled by different token colors (Jensen, 1992). Two research teams modeled and simulated biological processes with CPN. In all projects, Design/CPN software, consisting of a set of edition and simulation tools for CPN, has been used (CPN group at the University of Aarhus, 2004). However, each team had a different modeling goal and the conceptual meaning of the token colors was not the same.

Definition 9. The colored Petri net N is defined by the n-tuple $(P, T, Pre, Post, C, M)$ where:

$(P, T, Pre, Post, M)$ is a Petri net as described in Definition 1 and the tokens of M are identified by a color;

$C = \{C_1, C_2, \dots\}$, a set of colors. The incidence mappings Pre and $Post$ are functions of the token colors.

Genrich et al. (2001) modeled an enzymatic reaction with a colored Petri net transition. This transition is connected to places representing substrates like reactant,

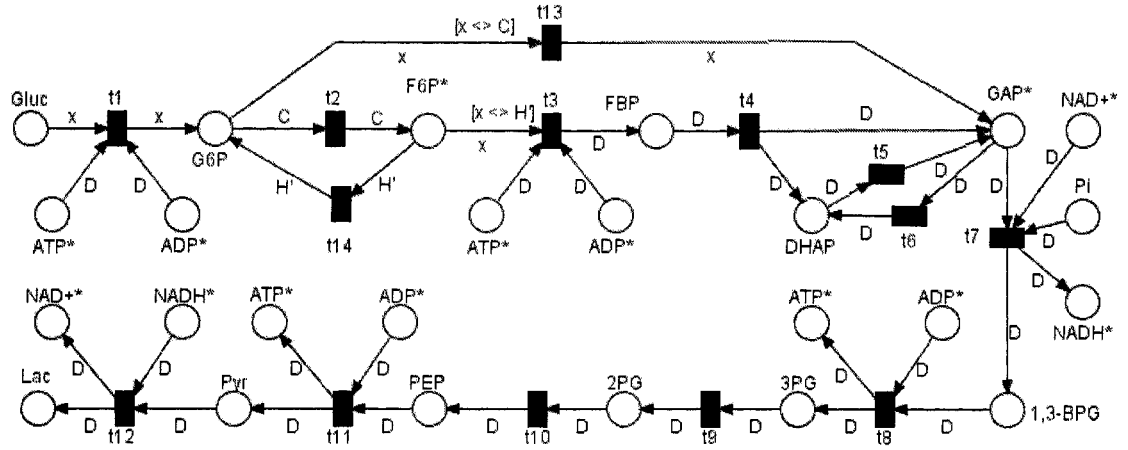


Figure 2.3 Colored Petri net model of the glycolysis and pentose phosphate pathways of an erythrocyte cell.

product, enzyme and inhibitor. In this model, tokens are identified by two colors, one associated with the substance name and the other with its concentration. The CPN used for this modeling also has functional features because an execution model is called upon, after every firing of the transition, to calculate and modify substrate concentration. These reaction rate calculations are performed according to the Michaelis-Menten biochemical equation, augmented by an additional term for the free reaction energy. The specific constants associated with each enzyme needed for these calculations are extracted from the BRENDA biochemical database (Schomburg et al., 2004). This transition is, in fact, a sub-model integrated into the glycolysis and citric acid metabolic models. A chain of enzymatic reactions constitutes the metabolic network to be quantitatively simulated. Another interesting part of the Genrich et al. paper is to propose rules for automatic pathway identification from databases, after which the pathways are modeled as Petri nets for simulation purposes.

Although the CPN was used in their work to model a metabolic system, Voss et al. (2003) proceeded differently to accomplish a qualitative analysis of steady

states in pathways. The objectives of this approach are compatible with the work of Reddy, but the modeling power and the communication capacity of the model are enhanced by the addition of colors. They refined the initial Reddy model of glycolysis and pentose phosphate pathways in an erythrocyte cell by the inclusion of reversible reactions and flux modes (see Figure 2.3 for model). Unlike to Genrich, the intention in using colors was to separate branches of a metabolic pathway and to differentiate molecules of the same species (thus, tokens of the same place) according to their origin and destination reaction. The analysis of the invariants of the CPN model found a preservation law for the amounts of all metabolites in the system and confirmed regenerative reactions and their partial order.

In the model shown in Figure 2.3, the colors used in this model are C, D and H'. The variable X means any color. Three flux modes are represented in the original model. The first mode is glycolysis, associated with the guard $[X \neq C]$ of t_{13} , which means that tokens of color C cannot travel through the pentose phosphate pathway. The two other modes are combined in the pentose phosphate pathway (all details are not shown), one of them is associated with the guard $[X \neq H']$. The model is inspired after Figure 3 of Heiner et al. (2001).

2.5 Hybrid Petri nets and supplementary extensions

An intuitive way for representing a molecular species concentration is with tokens of a continuous nature instead of a discrete nature. The hybrid Petri net (HPN) offers this possibility with a new continuous type of places and transitions (see Definition 5) (Alla et David, 1998) In HPN, discrete places and transitions, with their number of tokens represented by integers and their possible firing delay, are unchanged. But, in the new continuous places, tokens are replaced by a non-negative real number called a mark, and a variable called speed is assigned to the new continuous

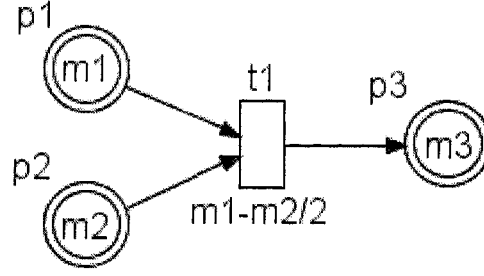


Figure 2.4 Hybrid Petri net continuous transition

transitions. The continuous transition speed is a rate of quantity transformation from input places to output places. Thus, the modeling of metabolic reactions and genetic regulation, usually performed with ODEs, can now be accomplished with hybrid Petri nets. Matsuno et al. (2000) and Chen et Hofestädt (2003) have demonstrated the feasibility of modeling biological systems with HPN.

Figure 2.4 shows a continuous HPN transition. In this transition, places p_1 , p_2 and p_3 are continuous places having content m_1 , m_2 and m_3 respectively. The function $m_1 - m_2/2$ is assigned to the continuous transition t_1 as its firing speed. t_1 can be fired if $m_1 > 0$ and $m_2 > 0$. The contents of p_1 and p_2 are consumed with the speed $m_1 - m_2/2$, and the content of p_3 increases with the same speed when the transition t_1 is fired.

Definition 10. The hybrid Petri net N is defined by the n-tuple $(P, T, Pre, Post, h, M)$ where:

$(P, T, Pre, Post, M)$ is a Petri net as described in Definition 1, where M is a combination of integers for the number of tokens in discrete places and of real numbers for the mark of continuous places;

$h : P \cup T \rightarrow \{D, C\}$, called a hybrid function, indicates for each place and transition, if it is discrete ($h(p_i) = D$ and $h(t_j) = D$) or continuous ($h(p_k) = C$ and $h(t_l) = C$);

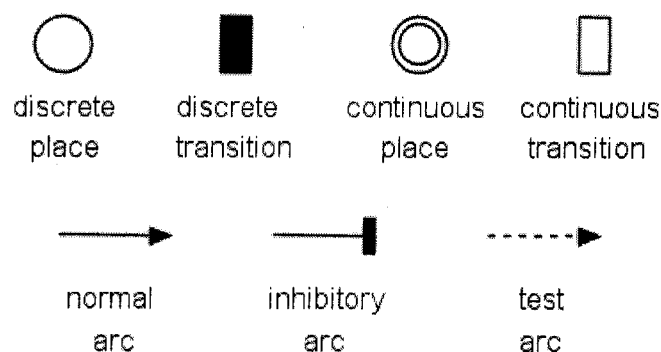


Figure 2.5 Graphical representations of the elements of hybrid functional Petri nets.

A delay d_t is assigned to all discrete transitions and a speed v_t is assigned to all continuous transitions.

Matsuno et al. (2003b) modeled the genetic switching mechanism of the λ phage with HPN by using the Visual Object Net++ tool (Drath, 1998). Part of their work demonstrated that the HPN approach constitutes a more intuitive modeling tool than the ODE, because of its graphical notation as much as for its power as a simulation tool. The variations in expressed protein concentration, resulting from the simulations, correspond to the biological data on that system.

To be able to adapt the HPN modeling approach to biological processes more accurately, functional Petri net properties have been added to create a new extension: the hybrid functional Petri net (HFPN) (Matsuno et al., 2003c). Figure 2.5 shows the HFPN elements graphically. This twinning allows a dynamic adaptation during the execution of the Petri net like in the work of Genrich et al. (2001). The attribution of a value to the net arcs makes it possible to model biochemical reaction rate equations like the Michaelis-Menten equation. Moreover, two new arc types, different from "normal" arcs, are included in the HFPN to model biological aspects. Firstly, inhibitory arcs model the inhibition function of molecules in some reactions.

An inhibitory arc with weight r enables the transition to fire only if the content of the place at the source of the arc is less than or equal to r . Secondly, test arcs verify the presence of content at its source place when the related transition fires without consuming anything. Unlike inhibitory arcs, some content is necessary in a place connected with a test arc to a transition about to fire. With the inhibitory arc, the repressing function of an operator on gene transcription can be represented. With the test arc, the action of an enzyme in a metabolic reaction, where the enzyme is required but not consumed by the reaction, can be modeled. From the HFPN architecture, the Matsuno team has developed biosimulation software for biologists called Genomic Object Net (GON) (Nagasaki et al., 2003; Doi et al., 2003b; Miyano Laboratory, University of Tokyo, 2003).

Many biological processes and systems have already been modeled and simulated using HFPN. Some processes are related to genetic regulation: the λ phage genetic control mechanism (Matsuno et al., 2000), circadian rhythms in *Drosophila* (Matsuno et al., 2003c) and the control mechanism of the lac operon of *E. coli*. (Matsuno et al., 2003a) Others are metabolic networks: the glycolytic pathway (Matsuno et al., 2003a) (see Figure 2.6) and the urea cycle (Chen et Hofestädt, 2003). The transduction signal system of apoptosis induced by the Fas ligand was also modeled with HFPN (Matsuno et al., 2003c). The pattern formation by a multicellular system due to interactions between cells with Delta-Notch signaling was simulated (Matsuno et al., 2003b). In this last experiment, a cellular boundary formation in *Drosophila* and other abnormal patterns could be analyzed with simulation results, corroborating observations from laboratory experiments.

In Figure 2.6, transitions $t_{12} - t_{23}$ are the natural degradation of substrates and their firing speed is given by the formula $m_X/10000$ where $X = 1, 2, \dots, 10, 12, 13$ for the different places. Production rate of enzymes ($t_{24}, t_{26}, \dots, t_{42}$) has a speed set to 1 and their degradation rate ($t_{25}, t_{27}, \dots, t_{43}$) has a speed given by the formula

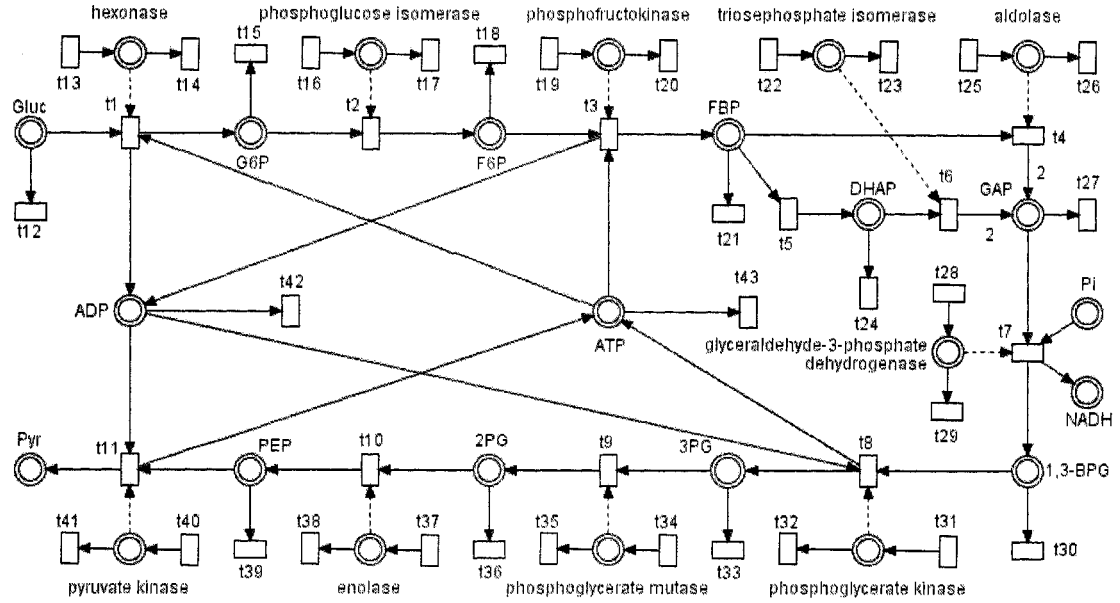


Figure 2.6 Hybrid functional Petri net model of the glycolytic pathway and lac operon gene regulatory mechanism.

(*enzyme concentration*)/10. The reactions in the main pathway ($t_1, t_2, \dots, t_{11}$) have the Michaelis-Menton equation for speed: $\frac{V_{max}m_X}{K_m + m_X}$ where $X = 1, 2, \dots, 10$, V_{max} is the maximum reaction speed and K_m is a Michaelis constant. The model is inspired after Figure 9 of Matsuno et al. (2003c).

The scientists who originated the HFPN architecture are continuing its elaboration by planning the incorporation of extensions to enhance its modeling power and simulation precision (Matsuno et al., 2003a). To achieve this, more complex information like the localization of biological objects or intercellular interactions at the molecular level must be incorporated into the models. One of these extensions makes it possible for a Petri net place to have various data types, like integers or real numbers, vectors and strings. Also, a conversion module implementing the procedure will be capable of converting biological models represented with ODEs in the E-CELL software format to HFPN models (Matsui et al., 2001).

2.6 Discussion

To make the modeling power of Petri nets richer and to adapt them to diverse problems, several extensions have been added to Professor Petri's theory. However, enhancements given by the extensions presented in this review do not offer the same modeling and simulation possibilities. Colored Petri nets (CPN) were created to diminish model size and so allow the models to manage more information without rendering its structure too complex. Stochastic Petri nets (SPN) are a specialized member of the timed Petri net family where firing delays are random variables. The need to represent discrete and continuous quantities in the same model motivated the development of hybrid Petri nets. Finally, the dynamic modification of arc weight through net marking is possible with functional Petri nets. Other extensions were developed to solve other problems. With the variety of Petri net types available, the context of use of the various extensions for biological process modeling and simulation can be called into question.

To analyze each approach and its biological modeling possibilities, it is pertinent to recall the compromise discussed by Reddy et al. (1996) between the modeling and simulation power and the decision-making power of Petri net. According to this compromise, the addition of modeling power to a Petri net with extensions will decrease its analytic capabilities. Indeed, augmenting modeling power amplifies the complexity of the determination of some properties, even to the point of indecision. For example, inhibitor arcs expand the richness of concepts expressed by a model, but greatly complicate its mathematical analysis. Thus, it is important to choose the Petri net extension that will be used for modeling judiciously, in accordance with the objectives to be attained.

In the literature, there are two categories of goals of Petri net biological modeling: **qualitative** and **quantitative** analysis. One can either learn more about the

Tableau 2.3 Biological modeling goals of Petri net extensions.

Petri net type	Modeling goal	Analysis type	Process types	Available software
CPN	Analysis of biological system properties	Qualitative	Metabolic	Design/CPN
SPN	Simulation of biological systems with low concentrations	Quantitative	Any biological stochastic process	Möbius
HFPN	Simulation of biological systems	Quantitative	Metabolic, Regulatory networks, Signal transduction	Genomic Object Net

properties of the system under study with a qualitative approach or study the system dynamics with simulation. When one wants to analyze a complex system of biochemical reactions, for example by identifying invariants, the presence of boundaries or the liveness in the system model, the Petri net extension chosen must enable those properties to be determined. By contrast, one may want to study the model's behavior by simulating it and thus obtain concentration graphs, and/or observe the achievability of a steady state, without any concern for property decidability. In that case, modeling power is more important. Thus, it is possible to identify modeling goals for each Petri net extension (see Table 2.3).

As was the case for the earliest modeling attempts with Petri net by Reddy et al. (1993), where a qualitative analysis was performed using place/transition nets, CPN can give some insights into a biological system (Heiner et al., 2001). Thus, a rigorous preliminary analysis can guide the elaboration of experiments when quantitative data is missing. Token color does not reduce the decidability power of Petri nets because it is always possible to convert each CPN mode into a traditional net. One advantage of using CPN is the possibility of discriminating metabolites on the basis of their chain of reactions in a model. However, the size of the model

must not exceed a certain limit because if the model complexity is too great, the state space will explode and its complete exploration and analysis will be impossible. Consequently, the analysis of complex systems is still a difficult task. Until now, this method has only been used on classic metabolic systems (like the Krebs cycle or glycolysis) to demonstrate its potential, but has also been incorporated in algorithms to find new metabolic pathways between two compounds (Küffner et al., 2000). Little information is needed to perform a qualitative analysis: the stoichiometry and the reversibility of the system reactions.

Hybrid and stochastic net attributes are intended for simulation, and token activity in the model is the main aspect considered in reproducing the behavior of a system. Wanting to model a system in order to quantitatively simulate it is in accordance with the modeling goals of Hofestädt et Thelen (1998). The criterion of choice between the stochastic and the hybrid extensions is the nature of the system to be modeled. If, for example, a model deals with a small number of molecules, such that their individuality has to be taken into account, its stochastic nature has to be represented and SPN are appropriate. By contrast, for models where the number of molecules is high enough to be represented in a satisfying way as a continuous quantity or as a concentration, HPN is the appropriate modeling approach. It is interesting to note, however, that, if the discrete transition delay in HPN is modified to become a random variable, then it is possible to fuse HPN with SPN. Matsuno et al. (2000) mentioned that their architecture could be easily modified to allow this blending. It is also important to recall that the twinning of the functional and hybrid extensions, as in the HFPN, facilitates the modeling of metabolic reactions.

To accomplish a quantitative analysis of a biological system, kinetic parameters like reaction rates have to be available. This kind of analysis provides means to test new hypothesis and to evaluate the impact of the variation of conditions in which a system evolves. It has proven its usefulness in several biology research projects.

As mentioned above, the stabilizing effect of the ROM protein on the genetic network controlling the replication of ColE1 plasmid replication was successfully analyzed with a SPN model (Goss et Peccoud, 1999) and a cellular boundary formation in *Drosophila* could be analyzed with simulation results of a HFPN model, corroborating observations from laboratory experiments (Matsuno et al., 2003b).

The work of Genrich et al. (2001) is an exception to the classification shown in Table 2.3. In their model, the color of the tokens is associated with the concentration of the molecular species in order to make a quantitative analysis. This demonstrates that the classification is not absolute. However, in addition to the colored extension, the Petri net type used in that model also includes an executable component which is called upon at every transition firing to calculate the concentration variations according to the Michaelis-Menten equation. Despite the validity of this approach, it is less intuitive and harder to implement than an approach using HFPN.

Using the Petri net for biological modeling offers many advantages. First, two Petri net properties, identified by Reddy et al. (1996), are pertinent to the modeling of biochemical systems: extendibility and abstraction. These two features are related to model hierarchization. Extendibility is the property of adding new sections to a net – for example, when supplementary information becomes available to complete a model, or when one wants to combine two complementary models – without having to considerably modify the structure of the resulting model. Abstraction is the property of neglecting the modeling of some aspects which do not concern the system under study by representing the sub-model by a transition. An example of abstraction is the transition t_{13} of the figures 2.2 and 2.3 representing the pentose phosphate pathway.

Second, many theoretical elements of Petri nets with a mathematical basis are useful as a preliminary analysis tool for biological pathways. Zevedei-Oancea et

Schuster (2003) thoroughly discuss this. Oliveira et al. (2001, 2003) developed and defined rigorously a computational approach based on Petri nets to identify interesting sub-circuit pathways in biochemical networks and applied their methodology to the Krebs cycle. Petri net invariants can be associated with flux modes and conservation relations. Furthermore, special sets of places known as siphons and traps can be identified. The places constituting a siphon stay empty once they have no tokens. At the other extreme, places forming a trap cannot lose one token when they reach a certain marking. Traps and siphons are of interest in biochemical modeling because these notions can be associated with the storage and consumption of system resources. Finally, the liveness of a net or its deadlocked condition are also properties which provide information about the biological system.

Third, biologists can easily model a biological system with the Petri net, and study it with the simulation capabilities of Petri net tools. The graphical aspects of the Petri net are quite similar to biochemical network representation, and this gives superior communication ability to models and facilitates their design. Moreover, the Petri net is readily comprehensible and necessitates little related knowledge on the part of biologists. Its mathematical basis makes it possible to accomplish complex simulations and to visualize results. The development of software based on the Petri net and specific to biology, like the Genomic Object Net tool (Nagasaki et al., 2003; Doi et al., 2003b; Miyano Laboratory, University of Tokyo, 2003), and the proposal of a data exchange format for models, the "Biology Petri Net Markup Language" (BioPNML) (Chen et al., 2002; Chen, 2002) leave few obstacles for the adoption of a Petri net approach by biologists. A powerful analysis and simulation environment can be implemented from this modeling technique to study hundreds, even thousands, of interconnections formed by the various genetic and metabolic networks in the cell.

Several types of formal representation other than Petri nets are required to model

biological processes. It was demonstrated that ordinary differential equations can be substituted by HFPN, but other biological phenomena involving spatial modeling, like diffusion, or molecular motions modeling, like molecular motors, do not have an equivalent in Petri net modeling. Because of the potential extensibility of Petri net formalism, it is possible to think that scientists will go beyond these modeling limits. Thus, we could also say that these biological phenomena do not have an equivalent in Petri net modeling *yet*.

2.7 Conclusion

In this paper, analysis, modeling and simulation of molecular biology systems using Petri nets have been presented and an overview of various approaches using colored, stochastic, hybrid and functional Petri nets was made. The modeling goal of each approach was identified, thus providing a starting point to interested new users.

Petri net is a formalism with many advantages for biologists. It has analytical and simulation capabilities which provide means to test hypotheses and gather information that might help the elaboration of experiments. As we learn more about metabolic pathways, gene regulatory networks and signalling pathways, powerful modeling tools like Petri net will be needed to understand the complexity of living systems.

Acknowledgments

We would like to thank Prof. Hanifa Boucheneb for her insightful comments and the two anonymous reviewers for their useful remarks about this article. This work was supported in part by NSERC grant A-0141 and a Canada Graduate Scholarship.

CHAPITRE 3

PHENOMENOLOGICAL AND MOLECULAR LEVEL PETRI NET MODELING AND SIMULATION OF LONG-TERM POTENTIATION

Abstract

Petri net¹ based modeling methods have been used in many research projects to represent biological systems. Among those methods, hybrid functional Petri net (HFPN) was developed especially for biological modeling in order to have a more intuitive Petri net based method for biologists. In literature, HFPN are used to represent kinetic models at the molecular level. We present two models of long-term potentiation previously represented by differential equations that we transformed into HFPN models: a phenomenological synapse model and a molecular level model of the CaMKII regulation pathway. By simulation, we obtained results similar to previous studies using these models. These results open the door to a new type of modeling where HFPN is used to combine different levels of abstraction inside one model. This approach can be useful when kinetic data is missing to fully model a system at the molecular level or when it is not in the research scope to fully study a system at the molecular level.

¹Cet article a été soumis en octobre 2004 à la revue *Biosystems*.

3.1 Introduction

Different biopathway representation and simulation methods are being developed and used to perform computational analysis of biological systems. de Jong (2002) provides a good introduction on modeling and simulation of genetic regulatory systems. He reviewed ten biological formalisms, including Boolean networks and ordinary differential equations, and he discussed modeling and simulation techniques appropriate for each one of them. One method that was not presented in de Jong's paper and that since has been recognized as promising to represent biopathways is Petri net. Petri net is a graphical and mathematical formalism mainly used in computer science to model systems with concurrent properties (Reisig, 1985). It has been applied for modeling in several different disciplines, from manufacturing systems to networking and communication protocols.

Reddy et al. (1993) were the first to use Petri nets to model biological systems. They related various mathematical properties of Petri net to biopathways biological properties. Shortly afterwards, Hofestädt (1994) tried to adapt the Petri net formalism to represent more adequately a biological system and he eventually designed a Petri net biological model with simulation capabilities (Hofestädt et Thelen, 1998). Still today, their respective work reflects the two main goals of Petri net based modeling methods: to qualitatively study the structure and topology of metabolic pathways or to quantitatively evaluate the dynamical behavior of biological systems. Specific extensions of Petri net theory such as stochastic to include randomness, colored to represent in the same model many distinct systems, hybrid to merge digital and continuous process and functional to dynamicaly modify the net parameters, have been used to model biological systems. We wrote a review explaining every Petri net approaches in biological modeling, where we discussed whether an approach is more appropriate for a qualitative or quantitative analysis (Hardy et Robillard,

2004a).

Matsuno et al. (2003c) developed the hybrid functional Petri net formalism (HFPN), "an intuitive and natural modeling method for biological pathways" and a step-by-step HFPN biological model construction process has been described (Doi et al., 2003a). HFPN models can be simulated with the Genomic Object Net (GON) software package (Nagasaki et al., 2003) to obtain metabolite and protein concentration graphs and to reproduce biological system behaviors. Until now, GON is the only Petri net modeling and simulation tool created specifically for biologists. An unknown biological phenomenon in a multicellular system, a mechanism of Notch-dependent boundary formation in the *Drosophila* large intestine, was discovered using HFPN and GON (Matsuno et al., 2003b).

In this paper, we transformed two neurobiology models, originally based on differential equations, of neuronal long-term potentiation (LTP) of the synaptic strength into hybrid functional Petri net models: one phenomenological model from Migliore et Lansky (1999b) and one molecular model from Bhalla et Iyengar (1999). The original models are briefly presented and one should refer to the referenced paper for more information. Only basic HFPN elements useful to the understanding of our results are presented.

We believe that HFPN is appropriate to neurobiology modeling and our purpose in this paper is to emphasize two useful aspects of HFPN modeling. First, it is possible to represent biological systems at different levels of abstraction because of the flexibility provided by the extendibility and abstraction properties of Petri net. Second, Petri nets are a powerful communication tool and ease model design and modification.

3.2 Hybrid functional Petri net modeling

The structure of a Petri net is composed of places and transitions linked by directed arcs. It is a network where moving objects are represented by tokens that are located in places. The amount of tokens in every places of the Petri net describes the state of the network and it is called marking. When the number of tokens in the input places of a transition meets certain conditions, this transition can be initiated. This event is called transition firing. The firing of a transition causes the consumption of tokens in its input places and/or the generation of tokens in its output places. The number of tokens implicated in transition firing is specified by a value assigned to each arc, known as weight. Thus, the Petri net dynamical aspect is the flow of tokens moving through the network.

Figure 3.1 illustrates a Petri net discrete transition (in contrast to continuous transition, which will be explained later). Places p_1 and p_2 are the input places and place p_3 and p_4 are the output places of the transition t_1 . The token content of places p_1 , p_2 , p_3 and p_4 are given by the variables m_1 , m_2 , m_3 and m_4 respectively. The weight constants 1 and 3 on the arcs going out of places p_1 et p_2 mean that t_1 can fire if $m_1 \geq 1$ and $m_2 \geq 3$, and the value 1 attached to transition t_1 means that the firing delay is 1 time unit (in the case of a timed net). These are the preconditions of the firing of transition t_1 . Generally, the absence of any value attached to an arc means that its weight is the default value of 1. When transition t_1 is fired, one token is removed from p_1 , three tokens are removed from p_2 , two tokens are added to p_3 and 1 token is added to p_4 . For more information on Petri net and their structural and dynamical properties, consult Reisig's introduction on the subject (Reisig, 1985). In Reddy et al. (1993) and other Petri net models of molecular biology systems, places are associated with molecular species, transitions to chemical reactions, tokens to molecules and arcs have reaction stoichiometric values.

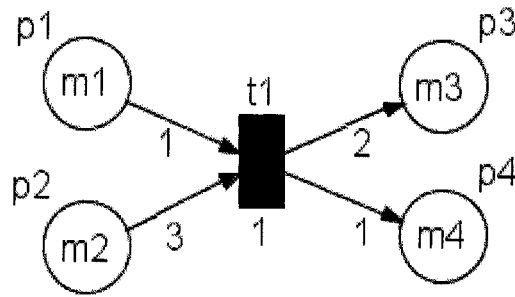


Figure 3.1 Petri net transition

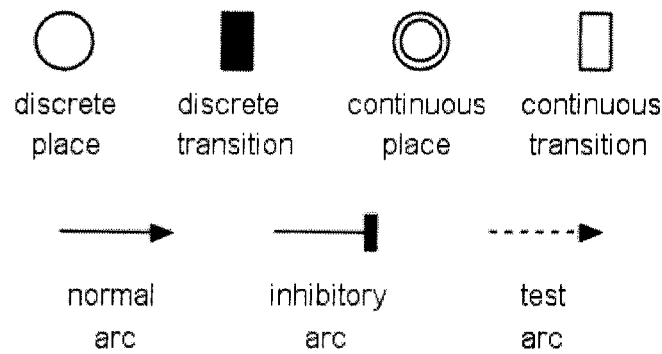


Figure 3.2 Graphical representations of the elements of hybrid functional Petri nets.

Hybrid functional Petri nets (HFPN) differ from the traditional place/transition nets by combining discrete and continuous quantities and events in the same representation (Matsuno et al., 2003c). In addition to discrete places and transitions, HFPN has continuous places and transitions. A continuous place contains a positive real value instead of a discrete amount of tokens like in a discrete place. The graphical notations of HFPN places, transitions and arcs are shown in Figure 3.2. A continuous transition fires continuously at a specified transition speed as long as preconditions are true, in opposition to the firing of discrete transition which can have a delay between events. A model based on differential equations can easily be represented by HFPN continuous places and transitions. There are also three types of arcs. Normal arcs are identical to the ones described previously, they allow the consumption of tokens once transition preconditions are met and the genera-

tion of tokens according to transition postconditions. The two other types of arcs, inhibitory and test arcs, do not participate in the flow of tokens. Instead, they can turn transitions on and off. Inhibitory arcs will disable the firing of a transition when the content of the arc source place is more or equal than the weight of the arc. At contrast, test arcs will enable the firing of a transition when the content of the arc source place is more or equal to the arc's weight.

3.3 Phenomenological model of the synapse

The long-term potentiation (LTP) and the long-term depression (LTD) of synaptic strength, which are neuronal properties known as synaptic plasticity, are the most extensively studied among the modifications that a synapse can undergo after presynaptic stimulation. Hippocampal LTP (Bliss et Lømo, 1979) and cerebellar LTD (Ito et Kano, 1982) are fairly well studied examples. The main reason for this interest is that synaptic plasticity is currently considered a potential neuronal mechanism involved in information storage and learning. Many computational synapse models have been developed, some of them based on phenomenological evidence and others for kinetic simulations at the molecular level. Among several phenomenological synapse models (Tsodyks et al., 1998; Murzina et Sil'kis, 1998; Migliore et Lansky, 1999a; Abarbanel et al., 2002; Sargsyan et al., 2003), we chose Migliore and Lansky's for its relatively simple mathematical representation with six differential equations, which was appropriate for our research. This model was originally designed to test whether stochastic conditioning, in other words different temporal distributions of irregular inter-stimulus intervals, could result in different amounts of LTP and LTD.

3.3.1 The model

The synapse model is formed of three differential equations for the presynaptic side (PRE) and three others for the postsynaptic side (POST) (see equations 3.1 to 3.6). The variables have the following meaning: x , y , z are respectively the fractions of presynaptic resources in recovered, active and inactive state (the sum of the three variables is always 1), v is the postsynaptic membrane potential in volts, C is an amount of second messenger molecules, N_P and N_D represent potentiation (P) and depression (D) synaptic processes involved with LTP and LTD maintenance. The time unit of these equations is the second. The variable I is the input stimulus, generated accordingly to the chosen irregular inter-stimulus intervals (ISI) distribution. The value of I is 300 for 5 ms when a synapse happens and 0 for the remaining time. The other symbols of the equations are constants.

PRE:

$$\frac{dx}{dt} = \frac{z}{\tau_{rec}} - U_{SE}xI \quad (3.1)$$

$$\frac{dy}{dt} = \frac{y}{\tau_{in}} + U_{SE}xI \quad (3.2)$$

$$z = 1 - x - y \quad (3.3)$$

POST:

$$\frac{dv}{dt} = \frac{v}{\tau_m} - R_{in}I_{syn} \left(\frac{1}{\tau_m} + f\delta(N_P - N_D) \right) \quad (3.4)$$

$$\frac{dC}{dt} = \gamma v - \eta C \quad (3.5)$$

$$\frac{dN_s}{dt} = vC - (\rho_s + R_{in}I_{syn}g\delta)N_s + \frac{M_s N_s^2}{A_s + N_s^2} - I\delta N_s \quad (3.6)$$

where $s = \{P, D\}$ (P for the potentiation equation and D for the depression equation), and the synaptic current $I_{syn} = yA_{SE}$.

When the input stimulus I is generated, a fraction of presynaptic resources is activated (corresponding to y), which in turn generates the synaptic current $I_{syn} = yA_{SE}$, where the constant A_{SE} is the absolute synaptic strength. In response, the membrane potential v elevates and this depolarization produces a second messenger C . This messenger catalyses the production of N_P and N_D which will modulate future synaptic responses.

3.3.2 Hybrid functional Petri net representation of the synapse model

The Petri net representation of this model was quite straightforward to complete. The seven variables (x , y , z , v , C , N_P and N_D) are represented in the model by the continuous places p_1 to p_7 in the PRE and POST subnets (see Figure 3.3). The third subnet is composed of discrete places and transitions configured in such a way to generate stimuli according to a specified distribution. Every place has an identifier and a variable (for place p_1 , variable is x). Every discrete transition has a delay (the delay of transition t_0 is 0.02 s) and every continuous transition has a speed that is computed from an equation (the speed of transition t_2 is $0.5 * x * I$). For each differential equation, positive factors correspond to continuous input transitions and negatives factors, to continuous output transitions of the places p_1 to p_7 . For example, equation 3.5 gives the rate for C (place p_5), thus the positive factor of this equation, $200 * v$, is the speed of the transition t_7 and the negative factor, $2 * C$, is the speed of transition t_8 .

The structure of this Petri net is not usual and it seems fractured in several disconnected parts. However, it is important to notice the role of the different variables in the rate equation of input transitions. For example, the variable y is used in the rate equation of transition t_5 , which is an input transition of variable v . It is also the case for C , N_P and N_D , that are connected to v , and for I , that is connected

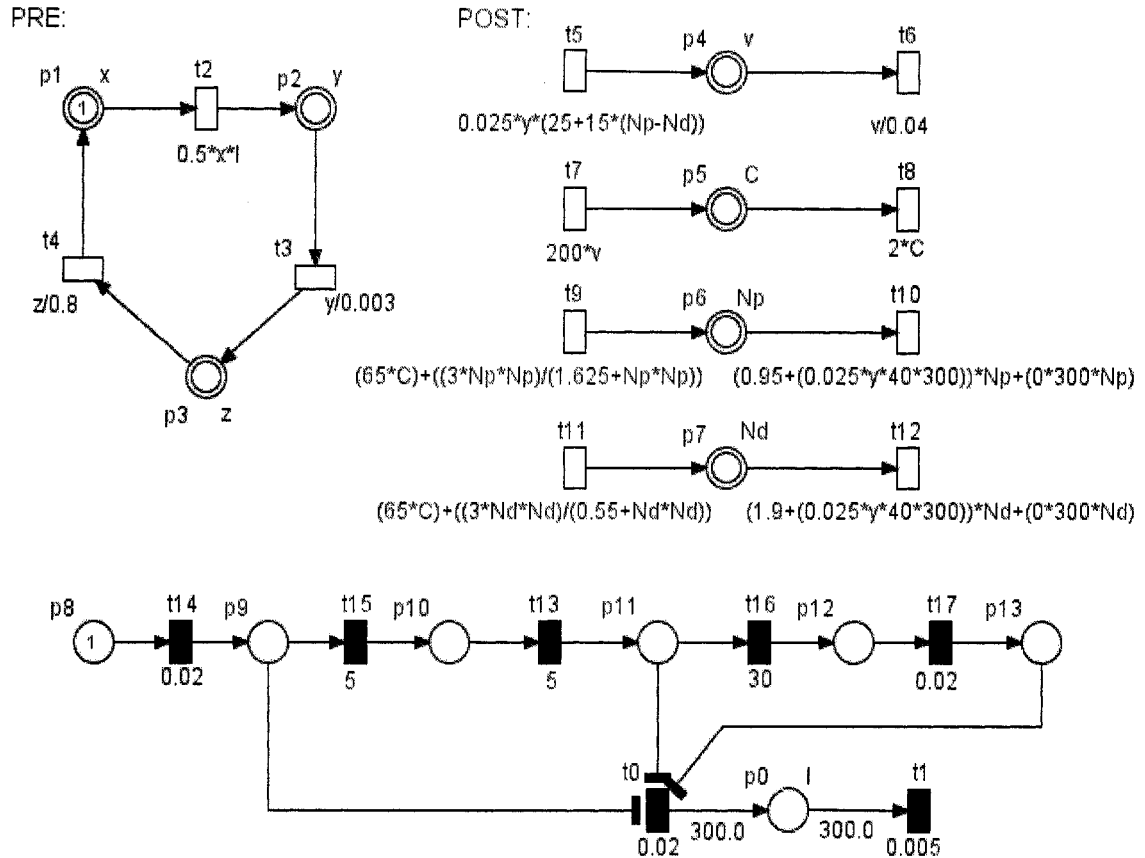


Figure 3.3 Hybrid functional Petri net representation of phenomenological synapse model.

to x and y . There is an absence of direct connectivity in the model between some variables because no conservation law applies for most of the variables. This is not the case for variables x , y and z . Their places are connected through arcs and transitions because presynaptic resources are always in either one state or another.

Input stimulus I is modeled by the discrete place p_0 . Several discrete places and transitions generate stimuli correspondingly to a specified conditioning. In the model shown in Figure 3.3, input stimuli are generated at a constant frequency of 50 Hz. This information is given by the delay of 20 ms of transition t_0 . This transition can be considered as the input stimulus generator. At each stimulus generated, 300 tokens are added in place p_0 . After 5 ms, they are all consumed by transition t_1 .

This part of the model has also three inhibitory arcs that will alternatively disable the generator. At the beginning of the simulation, the generator is on while there is one token in place p_8 . The first impulse is produced and it is considered the initial stimulus. After a delay of 20 ms, the token is consumed by transition t_{14} and another one is added in place p_9 . Because of an inhibitory arc linking place p_9 to the generator, the generator is then disabled and it will remain disabled as long as there is a token in place p_9 . This 5 second pause is to allow the system to return to its basal state. At the end of the delay, the token is absorbed by transition t_{15} and one is generated in place p_{10} where it stays for another 5 s while the generator is enabled. This is the conditioning period where inputs are generated at a constant frequency of 50 Hz. After the conditioning, the generator is disabled for 30 s because of the second inhibitory arc linking place p_{11} to the generator. One last input is generated at the end of that period. This last impulse is to be compared with the one before conditioning. Finally, a token is generated in place p_{13} where the last inhibitory arc disables the generator until the end of the simulation.

3.3.3 Simulation results

The results obtained by simulation of this model are identical to the ones achieved by Migliore and Lansky. The induction of LTP at a frequency of 50 Hz and of LTD at a frequency of 5 Hz is similar. Furthermore, every ISI distribution used in their paper (exponential and gamma distributions and bursting Poisson process) can be incorporated in the HFPN model. In HFPN, the delay of a discrete transition can not only be determined by a value or an equation, but also by a script. Thus, it is possible to program a random delay by using GON's random number generator, therefore creating a stochastic process. Simulation results for constant ISI of 5 and 50 Hz are shown in Figure 3.4. LTD is induced with a conditioning stimulation at a constant frequency of 5 Hz (on the left) and LTP is induced with a 50 Hz constant

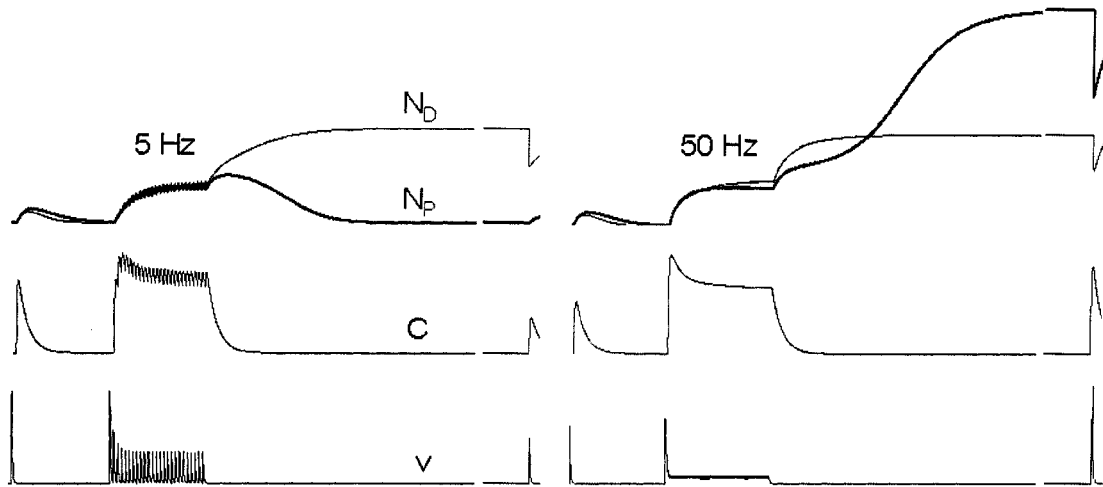


Figure 3.4 Simulation results of HFPN phenomenological synapse model.

frequency (on the right). One can observe this fact by noticing the membrane potential v before and after the 5 s conditioning. A drop of the membrane potential response to synapse indicates LTD and a rise indicated LTP.

3.4 Molecular model of the synapse

Kinetic simulations at the molecular level to study mechanisms responsible for synaptic plasticity have been developed in the last few years as more and more molecular kinetic data become available (Yang et al., 1999; Okamoto et Ichikawa, 2000; Kubota et Bower, 2001; Kuroda et al., 2001; Kikuchi et al., 2003). Bhalla and Iyengar model was interesting because these researchers defined a system-level approach: they tried to figure out if biological properties can emerge from the structure of networks of signaling pathways (Bhalla et Iyengar, 1999). This kinetic model was originally developed and simulated using the neuroscience simulator GENESIS (Bhalla, 2002). Kikuchi et al. (2003) proposed a modified version of this model by modeling LTP with dynamic reactions of protein phosphatase 2A

(PP2A) and it was simulated using E-Cell simulation environment (Tomita et al., 1999). Because it was sufficient for the purpose of our research, we restricted our HFPN model to the CaMKII regulation pathway, which is composed of 7 of the 15 modules from Bhalla and Iyengar's paper.

3.4.1 The model

Ca^{2+} /calmodulin-dependant protein kinase II (CaMKII) is known to function in long-term potentiation of synaptic responses in the hippocampus: a persistent activation of CaMKII leads to increased synaptic responses (Lledo et al., 2002). Also, it has been reported that the cAMP pathway is required in certain forms of LTP (Nguyen et al., 1994) and that CaMKII is activated by autophosphorylation (Meyer et al., 1992). On the basis of these experimental observations and others linking Ca^{2+} /calmodulin (CaM) and calcineurin (CaN) to cAMP pathway, Bhalla et Iyengar (1999) proposed that the cAMP pathway gates CaMKII signaling through the regulation of protein phosphatases and they tested if interactions between CaMKII, the cAMP pathway, and calcineurin (CaN) were sufficient to produce prolonged activation of CaMKII even after the Ca^{2+} signal is terminated.

In their model (see schematic representation in Figure 3.5 where rounded rectangles represent enzymes, circles represent messenger molecules, regular arrows represent activation and circle-ended arrows represent inhibition. Original figure in Bhalla et Iyengar (1999)), the inflow of intracellular Ca^{2+} entering through the NMDA receptor (NMDAR) and caused by synaptic stimuli known to induce LTP, results in an activation of CaMKII, AC1 and CaN via CaM binding. Consequently, there is an activation of PKA through the augmentation in cAMP level produced due to activation of AC1-AC8. PKA activation leads to the phosphorylation of Inhibitor-1 proteins and therefore to the inhibition of protein phosphatases (PP1). This inhi-

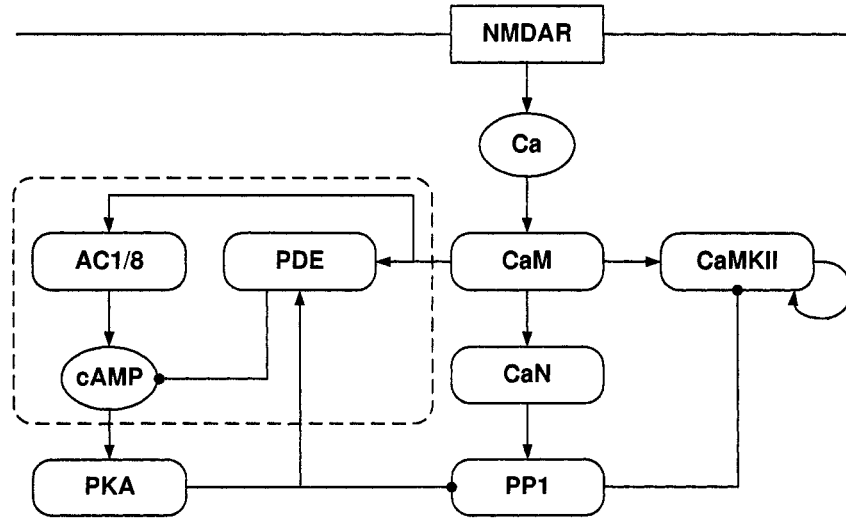


Figure 3.5 Block diagram of CaMKII regulation pathway.

bition of PP1 reduces dephosphorylation of CaMKII and allows activated CaMKII level to rise.

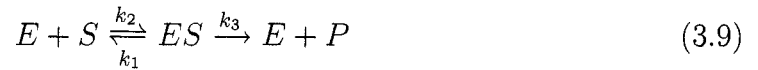
3.4.2 Hybrid functional Petri net representation of chemical and enzymatic reactions

The building blocks of signaling pathways are basic chemical reactions. Bhalla et Iyengar (1999) used the formulation of equation 3.7 to represent reactions. The rate constants (k_b and k_f) and the concentration of reactants A , B , C and D specify reactions dynamic. Differential equations are used to compute concentration variation (see equation 3.8).



$$d[A]/dt = k_b[C][D] - k_f[A][B] \quad (3.8)$$

Enzymatic reactions are usually modeled with the Michaelis-Menten formulation (Michaelis et Menten, 1913), which is suitable for a system in a quasi-steady state. But to be able to study dynamic changes, the enzyme-substrate complex is included and the enzymatic reaction is modeled by three mass action equations. Consequently, an enzymatic reaction is composed of two reactions in sequence, where the final step is assumed to be irreversible (see equation 3.9). In this equation, the symbols E , S , P and ES stand for enzyme, substrate, product and enzyme-substrate complex. k_1 , k_2 and k_3 are rate constants.



Chemical and enzymatic reactions are necessary building blocks to model basic signal transduction mechanisms such as protein-protein interactions and the production of intracellular messenger. Thus, HFPN equivalents of these blocks are essential to complete a molecular level model (Figure 3.6 and 3.7). In HFPN representations of equations 3.7 and 3.9, every place relates to a molecular species and its content is the molecular species quantity, expressed in concentration units or in number of molecules. Transitions represent unidirectional molecular reactions. For bidirectional reactions, two transitions must be defined. Each transition has a speed, which corresponds to the reaction rate. In Figure 3.6, the speed of transition t_1 is the product of constant k_f and the concentrations of substrates A and B . Transition t_1 represents the forward part of bidirectional reaction described in equation 3.7 and transition t_2 , its backward counterpart. Reaction substrates are associated to input places of a transition, and reaction products, to its output places.

To represent every molecular elements and interactions, the HFPN model of the CaMKII regulation networks is composed of 102 places and 188 transitions (see Figure 3.8). In this figure, seven modules are identified: A is the calcium pump

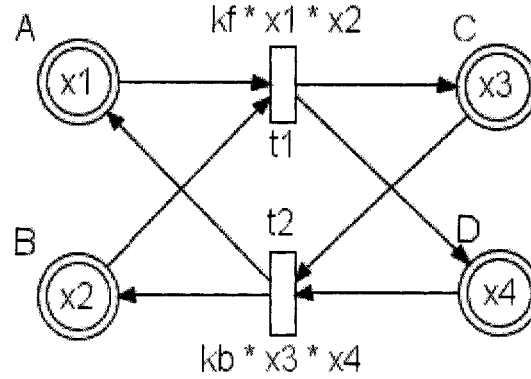


Figure 3.6 Hybrid functional Petri net representations of a basic chemical reaction.

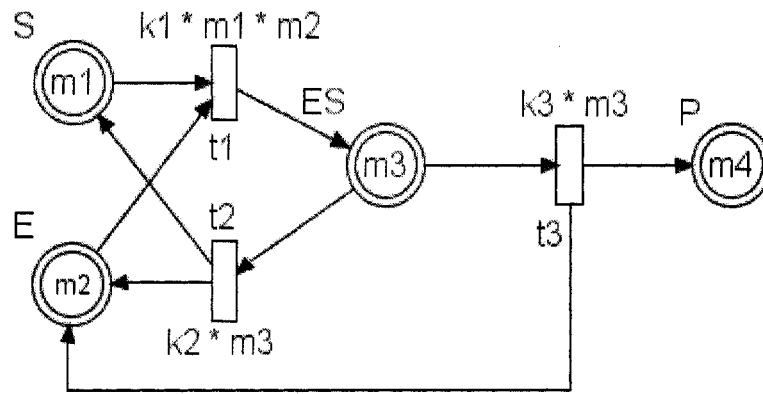


Figure 3.7 Hybrid functional Petri net representations of an enzymatic reaction

module, B is the CaM module (see Figure 3.9), C is the CaN module, D is the cAMP module with AC1-8 and PDE interactions, E is the PKA module, F is the PP1 module and G is the CaMKII module. The spatial organization of this model is similar to the block diagram organization in Figure 3.5. Reaction rate constants used to specify transitions speed were obtained from Bhalla and Iyengar's paper and a database called DOQCS (Bhalla et Iyengar, 1999; Sivakuraman et al., 2003).

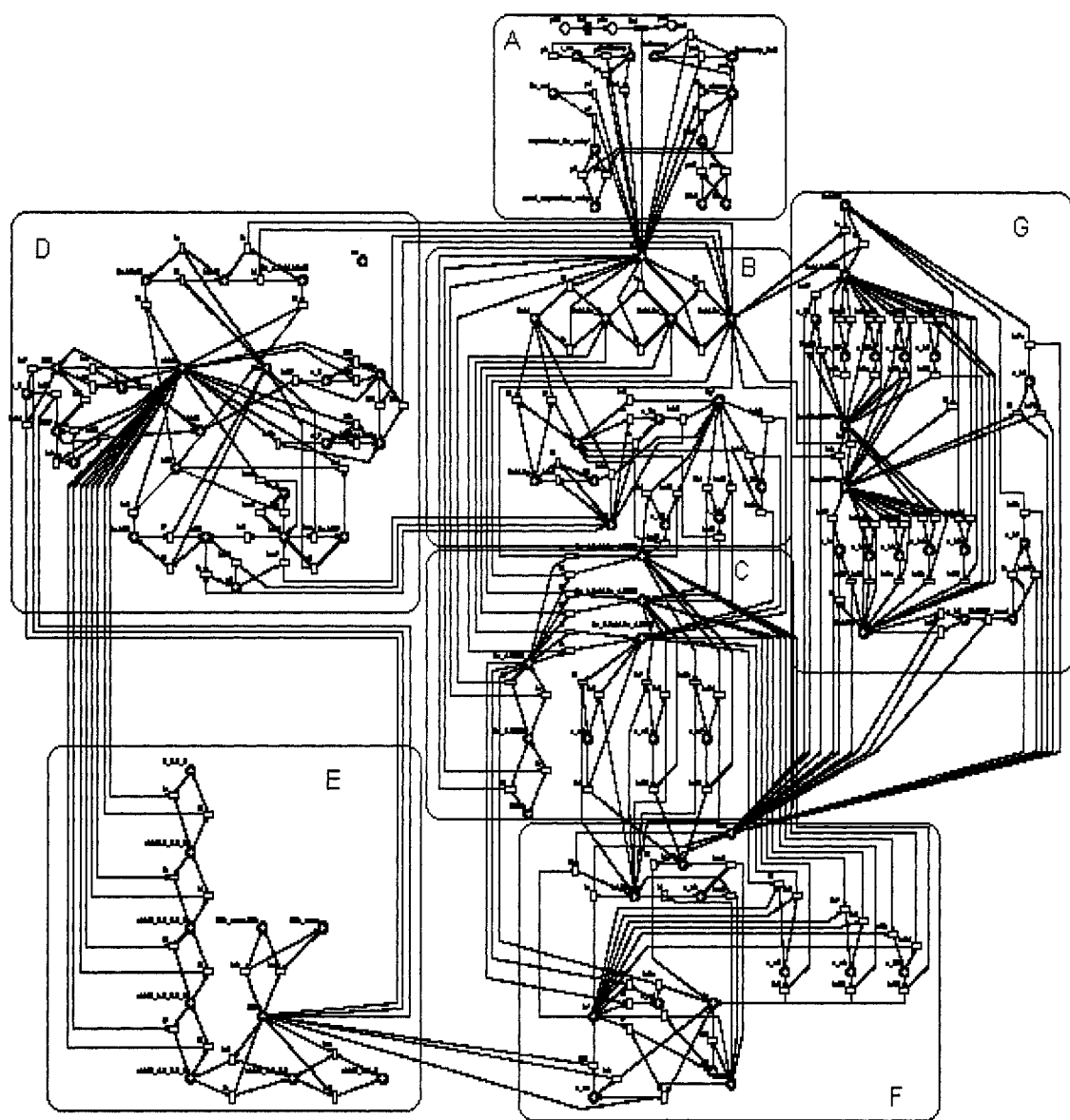


Figure 3.8 Hybrid functional Petri net representation of the seven modules of the CaMKII regulation pathway.

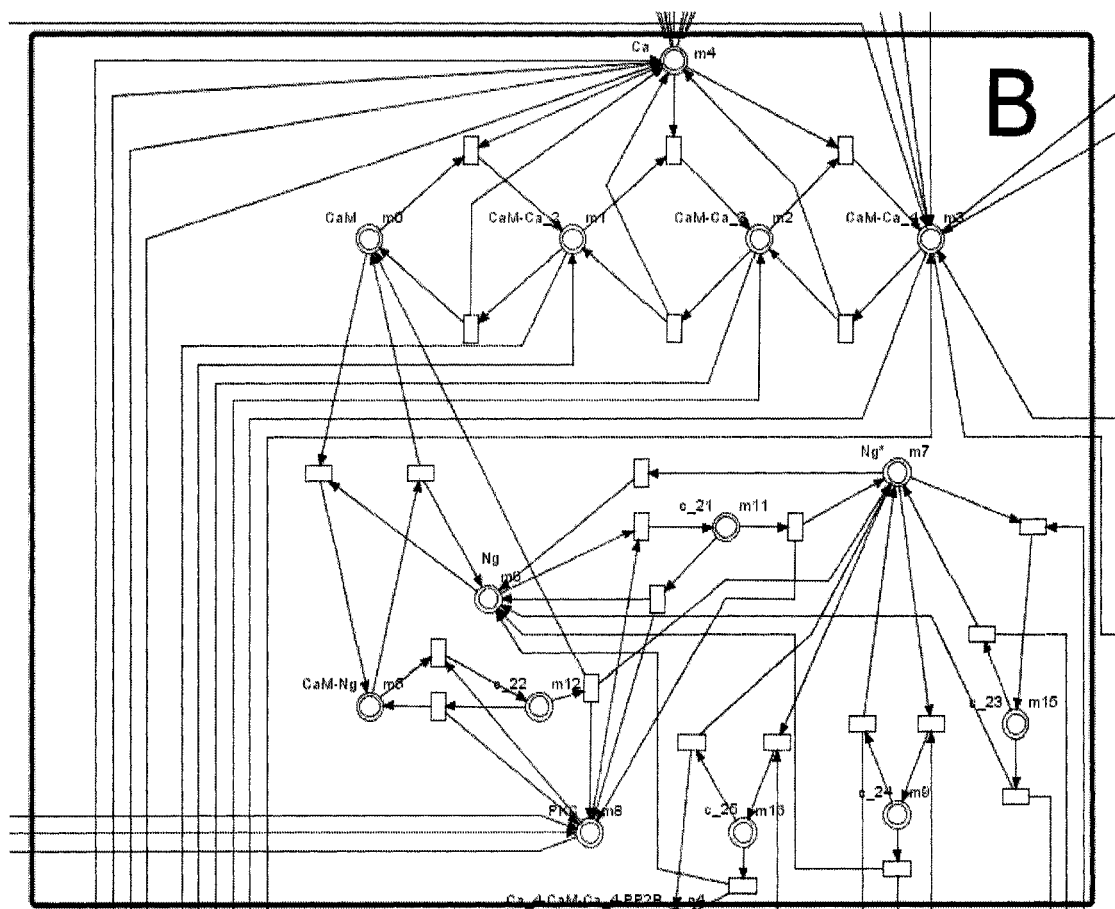


Figure 3.9 Hybrid functional Petri net representation of the calmoduline module (CaM) of the CaMKII regulation pathway.

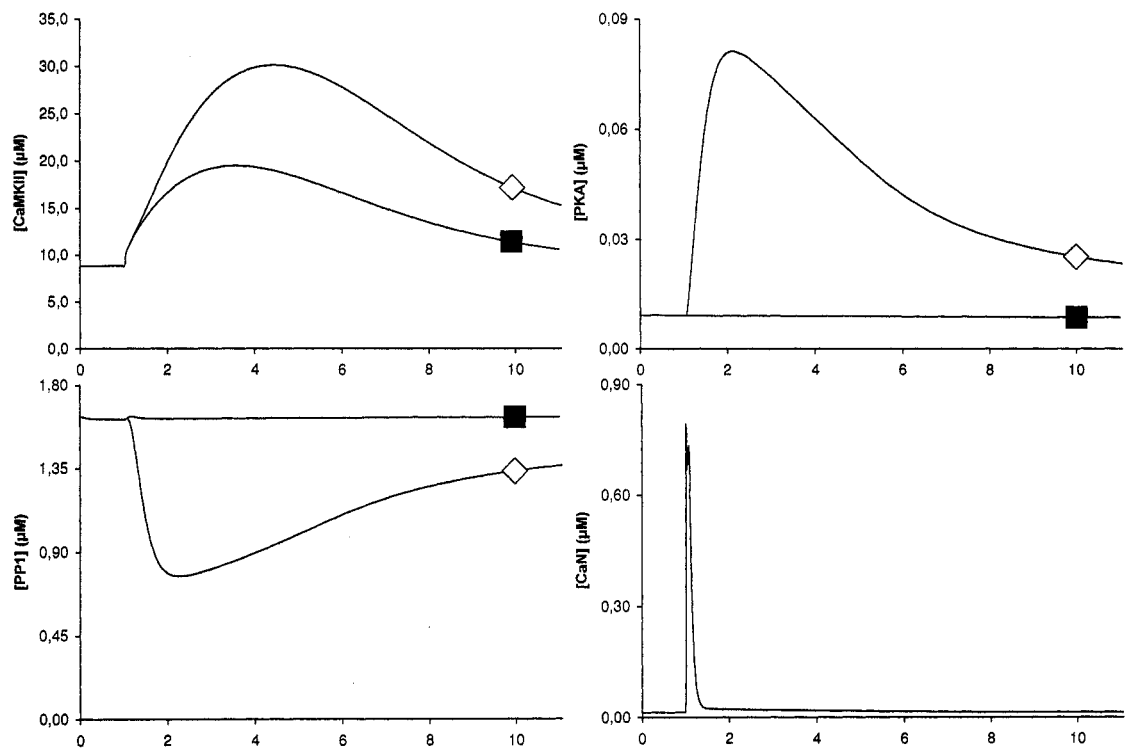


Figure 3.10 Simulation results of enzyme activation from hybrid functional Petri net model of CaMKII regulation pathways.

3.4.3 Simulation results

Simulation of the molecular level HFPN model of CaMKII regulation produced similar results to the original differential equation model (Bhalla et Iyengar, 1999) (see Figure 3.10, where filled squares identify the full model curves and open diamonds, the model curves with cAMP level held fixed at resting concentrations). A calcium stimulus is induced after 1 minute of simulation. The presence of cAMP creates a large increase in the amplitude of CaMKII response and prolongs its activity. However, this effect is not persistent and CaMKII levels return to basal state after 20 minutes of the final stimuli. With cAMP concentration held fixed, PKA stayed inactive, thus preventing inhibition of PP1. As a result, CaMKII activity was decreased and shortened. CaN activity remained unchanged.

3.5 Discussion

It was not a surprise that we were able to reproduce the results obtained by the authors of the phenomenological and the molecular level models. HFPN representations are already known to be equivalent to differential equations systems (Matsuno et al., 2000). Our goal, by completing these modelings and these simulations, was to show the effectiveness of a HFPN method to represent different views of a system and to exemplify a novel way to use HFPN to accomplish biological modeling. Until now, published HFPN models were only representations of biological systems at the molecular level.

Our results suggest that with the HFPN modeling method, a modeler can navigate through different levels of abstraction in the same model. The simulation of phenomenological models represented with HFPN offers the same advantages as other mathematical tools, already numerous, plus the added value of its graphical representation. Moreover, one of the great modeling possibilities of any Petri net based approach is the possibility to easily combine different levels of abstraction in the same representation. By this we mean that a synapse model, or any other biological model, could bring together a phenomenological part and a kinetic part into the same representation.

In the Petri net theory, the property of abstraction is defined as a way to collapse certain sections of a net, e.g. by reducing several transitions and places to a few, without losing any behavioral characteristics. Petri net theory also defines its opposite: extendibility, which is the property to extend a net by adding or modifying sections of a net without changing its whole structure (Reddy et al., 1996). These features have already been recognized to be useful for the modeling of biological phenomena in cases where "the present knowledge is incomplete and an extension or a greater precision could eventually be added to a model" (Reddy, 1994).

We would like to take this idea one step further. The Petri net abstraction property can be viewed as a way to choose the appropriate level of details for modeling different parts of a biological system. This is useful in two situations: first, when kinetic data is missing to fully model a system at the molecular level, then some parts of the model can be represented at the phenomenological level; and second, when it is not in the research scope to fully study a system at the molecular level. A modular model is suitable for this last choice of representation: some modules of molecular reactions can be modeled with the minimum HFPN elements to have expected behavioral characteristics and the other modules that are generating simulation data can include every interaction and be modeled with kinetic parameters.

Many aspects of modeling are facilitated by the graphical nature of Petri net. It is easier to visualize complex biological systems as network components rather than formal and complex equations. Also, HFPN representation is similar to the pathway representations that biologists are used to. Thus, models are easier to understand, to convey to others, to design and to modify. A diagrammatic representation is helpful to think clearly about the model. This is particularly important for combinations of different levels of abstraction in a single model as we suggested. Furthermore, different modules can be delimited and therefore are made straightforward to identify.

A HFPN method can be useful when a model needs frequent changes. We believe HFPN is a computer representation that can be incorporated into a systems biology experimental process where a computational model is developed by reverse engineering (one alters the model to best fit the biological model) inside an iterative experimental framework (Ideker et al., 2001a). With HFPN, different hypotheses can be modeled and then tested by simulation to see the outcome of structure changes or parameter variations. Feedback loops, largely present in large and complex biological systems, can also be identified and studied.

In the models presented in this paper, discrete places and transitions have solely been used to generate different distributions of stimuli. However, the primary function of HFPN discrete elements, as intended by its designers, is to model genetic switch controls which have a discrete on/off behavior (Matsuno et al., 2003c).

The HFPN method, as it is for the moment, has some limitations similarly to other modeling methods. To simulate events of different time frames is computationally expensive. Some biological phenomena involving spatial modeling and molecular motions modeling, like diffusion and molecular motors, do not have an equivalent in Petri net modeling. Other types of formal representation are required to model these biological processes. However, Petri net is an extendable notation and it is possible to think that extensions will overcome these limits.

3.6 Conclusion

In this paper, we used hybrid functional Petri net to represent long-term potentiation models: a phenomenological synapse model from Migliore et Lansky (1999b) and a molecular level model of the CaMKII regulation pathway from Bhalla et Iyengar (1999). We showed how chemical and enzymatic reactions can be represented by the HFPN elements and we obtained the same simulation results as the ones published in these previous studies.

We also suggested that HFPN is a method that can combine different levels of abstraction in the same model, thus providing many advantages to biologists. The graphical nature of HFPN makes it a powerful method for communication and for model design and implementation.

The combination of different levels of abstraction is not an easy modeling task and new research questions will arise from this problem. Structural reduction is

a well documented property of traditional Petri net, but not for HFPN. We will investigate the means to reduce complex modules to few HFPN elements.

Acknowledgments

We would like to thank Masao Nagasaki for his help with GON and the CaMKII regulation model. This work was supported in part by NSERC grant A-0141 and a Canada Graduate Scholarship.

DISCUSSION ET CONCLUSION

Ce travail de recherche nous a permis de développer une aptitude de modélisation et de simulation des systèmes biologiques à l'aide des réseaux de Pétri. La première étude rapportée dans ce mémoire fait la recension des différentes méthodes de modélisation basées sur ce formalisme ainsi que l'analyse des objectifs de recherche de chacune. Cette étude est le premier article de synthèse sur ce sujet et constitue en soi une contribution originale. Lors de cette réflexion, nous avons alors identifié une méthode, les réseaux de Pétri hybrides fonctionnels, qui semblait la plus intéressante pour effectuer des évaluations quantitatives. Nous avons ensuite expérimenté et appliqué cette méthode pour réaliser que de nouveaux types de modélisation étaient possibles. Nous l'avons utilisée dans notre deuxième étude, qui est notre seconde contribution originale et dans laquelle nous avons présenté deux modèles de systèmes biologiques impliqués dans la plasticité neuronale.

Ce qui était au départ un travail exploratoire nous a permis de proposer d'autres possibilités de modélisation avec les réseaux de Pétri hybrides fonctionnels. En effet, tous les modèles basés sur cette approche et qui ont fait l'objet d'articles publiés dans la littérature ont représenté des systèmes au niveau moléculaire. Pourtant, ce type de modèle peut représenter plusieurs niveaux d'abstraction, ce que nous avons démontré avec le modèle phénoménologique de la synapse présenté dans la deuxième étude.

Parmi les difficultés rencontrées lors de la réalisation des études, plusieurs ont été causées par l'utilisation du logiciel Genomic Object Net. Celui-ci en est à sa première version et certaines fonctionnalités étaient déficientes. D'ailleurs, la prochaine version du logiciel viendra corriger certains problèmes que nous avons rencontrés pendant la recherche. Il était bien souvent frustrant de constater que la plupart

des problèmes que nous communiquions à l'équipe de développement du logiciel ne pouvaient être solutionnés dans l'immédiat. Cette dernière nous annonçait alors que des correctifs étaient présentement apportés pour régler ces problèmes et que la prochaine version en serait exempte. Nous détectons donc des lacunes de la première édition de ce logiciel.

Parmi ces inconvénients, mentionnons la sauvegarde des données de simulation. Dans la version du logiciel que nous avons utilisée, la fréquence de sauvegarde doit correspondre au temps d'échantillonnage de la simulation. On doit comprendre de ceci que pour la simulation de larges systèmes sur une période relativement longue, énormément de données sont obligatoirement sauvegardées. Malheureusement, le logiciel n'arrive pas à gérer correctement les simulations qui génèrent beaucoup de données et ceci nous a obligés à réévaluer la durée des simulations initialement prévues. Une autre remarque, plus théorique celle-là, concerne la tirabilité de transitions dont les ressources sont en compétition (en d'autres mots, lorsqu'une place est la place d'entrée de plusieurs transitions). Avec l'utilisation de réseaux discrets, cette tirabilité est indéterministe. Ainsi, les jetons seront alternativement consommés par différentes transitions. Toutefois, avec les réseaux hybrides, surtout pour faire la simulation de systèmes biologiques, ceci ne peut être le cas et la tirabilité des transitions ne peut pas être indéterministe. L'algorithme de tirabilité tel qu'il est actuellement implanté dans GON nous a causé quelques problèmes et nous avons dû ajuster certains paramètres en conséquence. Malheureusement les développeurs n'ont pas voulu dévoiler l'algorithme utilisé, mais ils nous ont affirmé que celui-ci avait été refait pour la prochaine version.

Ces remarques peuvent paraître bien critiques, mais nous voulons saluer le travail effectué par l'équipe du Genome Human Center de l'Université de Tokyo qui développe le logiciel GON. Il demeure jusqu'à aujourd'hui le seul outil basé sur les réseaux de Pétri qui soit dédié à la biologie. Les prochaines versions de GON

en feront un outil efficace, agréable à utiliser, qui offrira aux biologistes toute la puissance des réseaux de Pétri.

Une des conclusions de la deuxième étude est d'envisager un modèle où sont combinés plusieurs niveaux d'abstraction. Par exemple, on peut imaginer un modèle où certaines sections critiques sont très détaillées alors que d'autres sections moins importantes, mais dont le comportement est essentiel au système, sont modélisées à un niveau d'abstraction plus haut. Ce type de modélisation ne doit pas être considéré comme un substitut dont la conception est plus facile. Le choix du niveau de détail peut s'avérer une tâche délicate et davantage de réflexion devra être consacrée à ce sujet. La propriété de réduction structurale est bien documentée pour les réseaux de Pétri traditionnels, mais l'équivalent pour les réseaux de Pétri hybrides fonctionnels n'a pas encore été fait. Il s'agit d'une des principales avenues de recherche future qui sont issues de notre projet.

Nous suggérons aussi l'intégration de notre méthode de modélisation à l'intérieur d'un processus expérimental comme le propose Ideker et al. (2001a). À notre avis, les avantages des réseaux de Pétri hybrides fonctionnels en font un excellent outil informatique pour les biologistes. Néanmoins, un travail de réflexion est aussi nécessaire pour adapter la méthode aux besoins des expérimentalistes. Chaque projet a ses propres particularités qui doivent influencer l'approche de modélisation. Par exemple, les paramètres cinétiques d'une voie biochimique ne sont pas toujours disponibles, on doit alors s'interroger sur les approximations à faire. De plus, la présence de certains éléments et de certaines interactions d'un système dans un modèle doivent aussi être questionnées. Nous considérons les réseaux de Pétri comme un outil très intéressant pour les biologistes, mais une collaboration étroite entre modélisateurs et expérimentalistes sera nécessaire pour que l'intégration de cette méthode de modélisation dans un processus expérimental soit réussie. Surtout que c'est avec la découverte de nouveaux faits scientifiques prédits par ou découlant

de notre méthode de modélisation que celle-ci obtiendra ses lettres de noblesse et une réelle reconnaissance dans la communauté scientifique.

RÉFÉRENCES

- ABARBANEL, Henry D. I., HUERTA, R., RABINOVICH, M. I. 2002. «Dynamical model of long-term synaptic plasticity». *Proceedings of the National Academy of Sciences of the United States of America*. 99:15. 10132–10137.
- AJMONE MARSAN, Marco, BALBO, Gianfranco, CHIOLA, Giovanni, CONTE, Gianni, DONATELLI, Susanna, FRANSCHESCHINIS, Giuliana. 1991. «An introduction to generalized stochastic Petri nets». *Microelectronics and Reliability*. 31:4. 699–725.
- ALLA, Hassane, DAVID, René. 1998. «Continuous and hybrid Petri nets». *Journal of Circuits, Systems and Computer Science*. 8:1. 159–188.
- ALTSCHUL, Stephen F., GISH, Warren, MILLER, Webb, MEYERS, Eugene W., LIPMAN, David J. 1990. «Basic local alignment search tool». *Journal of Molecular Biology*. 215:3. 403–410.
- APWEILER, Rolf, BAIROCH, Amos, WU, Cathy H., BARKER, Winona C., BOECKMANN, Brigitte, FERRO, Serenella, GASTEIGER, Elisabeth, HUANG, Hongzhan, LOPEZ, Rodrigo, MAGRANE, Michele abd Martin, Maria J., NATALE, Darren A., O'DONOVAN, Claire, REDASCHI, Nicole, YEH, Lai-Su L. 2004. «UniProt : the universal protein knowledgebase». *Nucleic Acids Research*. 32: *Database Issue*. D115–D119.
- ARKIN, Adam, ROSS, John, MCADAMS, Harley H. 1998. «Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected *Escherichia coli* cells». *Genetics*. 149. 1633–1648.
- ATTWOOD, Teresa K. 2000. «The babel of bioinformatics». *Science*. 290. 471–473.

- BADER, Gary D., BETEL, Doron, HOGUE, Christopher W. V. 2003. «BIND : the biomolecular interaction network database». *Nucleic Acids Research*. 31:1. 248–250.
- BAHI-JABER, Narges, PONTIER, Dominique. 2003. «Modeling transmission of directly transmitted infectious diseases using colored stochastic Petri nets». *Mathematical Biosciences*. 185:1. 1–13.
- BAKER, David, SALI, Andrej. 2001. «Protein structure prediction and structural genomics». *Science*. 294. 93–96.
- BENSON, Dennis A., KARSCH-MIZRACHI, Ilene, LIPMAN, David J., OSTELL, James, WHEELER, David L. 2004. «Genbank : update». *Nucleic Acids Research*. 32: *Database Issue*. D23–D26.
- BHALLA, Upinder S. 2002. «Use of kinetikit and GENESIS for modeling signaling pathways». *Methods in Enzymology*. 345. 3–23.
- BHALLA, Upinder S., IYENGAR, Ravi. 1999. «Emergent properties of networks of biological signaling pathways». *Science*. 283. 381–387.
- BLISS, T.V.P., LØMO, T. 1979. «Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path». *Journal of Physiology*. 232:2. 1979.
- BOWER, James M., BOLOURI, Hamid. 2001. *Computational modeling of genetic and biochemical networks*. Cambridge, MA. : MIT Press. 390 p.
- BUTTE, Atul J. 2001. «Challenges in bioinformatics : infrastructure, models and analytics». *TRENDS in biotechnology*. 19:5. 159–160.
- CHEN, Ming. 2002. «Modelling and simulation of metabolic networks : Petri net approach and perspective». *Proceedings of the European Simulation Multiconference*. Sous la direction de AMORSKI, K. et al. Modelling and Simulation 2002 P. 441–444.

- CHEN, Ming, FREIER, Andreas, KÖHLER, Jacob, RÜEGG, Alexander. 2002. «The biology Petri net markup language». *Proceedings of Promise'2002*. Sous la direction de DESEL, J. et al. t. 21 de *Lecture Notes in Informatics*. P. 150–161.
- CHEN, Ming, HOFESTÄDT, Ralf. 2003. «Quantitative Petri net model of gene regulated metabolic networks in the cell». *In Silico Biology*. 3: 3. [En ligne]. [http ://www.bioinfo.de/isb/2003/03/0030/](http://www.bioinfo.de/isb/2003/03/0030/) (Page consultée le 6 août 2004).
- COLLADO-VIDES, Julio, HOFESTÄDT, Ralf. 2002. *Gene Regulation and Metabolism – Postgenomic Computational Approaches*. Cambridge, MA. : MIT Press. 310 p.
- CPN group at the University of Aarhus. 2004. «Design/CPN Online». [En ligne]. [http ://www.daimi.au.dk/designCPN/](http://www.daimi.au.dk/designCPN/) (Page consultée le 8 août 2004).
- CSETE, Marie E., DOYLE, John C. 2002. «Reverse engineering of biological complexity». *Science*. 295. 1664–1669.
- DE JONG, Hidde. 2002. «Modeling and simulation of genetic regulatory systems : a literature review». *Journal of Computational Biology*. 9:1. 67–103.
- DOI, Atsushi, FUJITA, Sachie, MATSUNO, Hiroshi, NAGASAKI, Masao, MIYANO, Satoru. 2003a. «Constructing biological pathway models with hybrid functional Petri nets». *In Silico Biology*. 4: 2. [En ligne]. [http ://www.bioinfo.de/isb/2004/04/0023/](http://www.bioinfo.de/isb/2004/04/0023/) (Page consultée le 7 août 2004).
- DOI, Atsushi, NAGASAKI, Masao, FUJITA, Sachie, MATSUNO, Hiroshi, MIYANO, Satoru. 2003b. «Genomic Object Net : II. Modeling biopathways by hybrid functional Petri net with extension». *Applied Bioinformatics*. 2: 3. 185–188.
- DRATH, Rainer. 1998. «Visual object net ++». [En ligne]. [http ://www.systemtechnik.tu-ilmenau.de/ drath/visual_E.htm](http://www.systemtechnik.tu-ilmenau.de/drath/visual_E.htm) (Page consultée le 8 août 2004).

- EISENBERG, David, MARCOTTE, Edward M., XENARIOS, Ioannis, YEATES, Todd O. 2000. «Protein function in the post-genomic era». *Nature*. 405. 823–826.
- ENDY, Drew, BRENT, Roger. 2001. «Modelling cellular behavior». *Nature*. 409. 391–395.
- GENRICH, Hartmann J., KÜFFNER, Robert, VOSS, Klaus. 2001. «Executable Petri net models for the analysis of metabolic pathways». *International Journal on Software Tools for Technology Transfer (STTT)*. 3:4. 394–404.
- GOSS, Peter J. E., PECCOUD, Jean. 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, Peter J. E., PECCOUD, Jean. 1999. «Analysis of the stabilizing effect of Rom on the genetic network controlling ColE1 plasmid replication». *Proceedings of the Pacific Symposium on Biocomputing (PSB'99)* t. 4. P. 65–76.
- GRONEWOLD, Anja, SONNENSCHNIG, Michael. 1998. «Event-based modeling of ecological systems with asynchronous cellular automata». *Ecological Modelling*. 108:1. 37–52.
- HARDY, Simon, ROBILLARD, Pierre N. 2003. *Modélisation et simulation de processus de biologie moléculaire basées sur les réseaux de Pétri : une revue de littérature*. Montréal : École Polytechnique de Montréal. 19 p. EPM/RT-2003-10.
- HARDY, Simon, ROBILLARD, Pierre N. 2004a. «Modeling and simulation of molecular biology systems using petri nets : modeling goals of various approaches». *Journal of Bioinformatics and Computational Biology*. Accepté pour publication.
- HARDY, Simon, ROBILLARD, Pierre N. 2004b. «Phenomenological and molecular level petri net modeling and simulation of long-term potentiation». *Biosystems*. Soumis.
- HEINER, Monika, KOCH, Ina, VOSS, Klaus. 2001. «Analysis and simulation of steady states in metabolic pathways with Petri nets.» *3rd Workshop and Tutorial*

on *Practical Use of Coloured Petri Nets and the CPN Tools (CPN'01)*. Sous la direction de JENSEN, Kurt. DAIMI PB-554, Aarhus University. P. 15–34.

HEINRICH, Reinhart, SCHUSTER, Stefan. 1998. «The modelling of metabolic systems. structure, control and optimality». *Biosystems*. 47:1. 61–77.

HOFESTÄDT, Ralf. 1994. «A Petri net application of metabolic processes». *Journal of System Analysis, Modeling and Simulation*. 16. 113–122.

HOFESTÄDT, Ralf, THELEN, S. 1998. «Quantitative modeling of biochemical networks». *In Silico Biology*. 1: 1. [En ligne]. <http://www.bioinfo.de/isb/1998/01/0006/> (Page consultée le 7 août 2004).

HOLM, Liisa, PARK, Jong. 2000. «DaliLite workbench for protein structure comparison». *Bioinformatics*. 16:6. 566–567.

IDEKER, Trey, GALITSKI, Timothy, HOOD, Leroy. 2001a. «A new approach to decoding life : Systems biology». *Annual Review on Genomics and Human Genetics*. 2. 343–372.

IDEKER, Trey, THORSSON, Vesteinn, RANISH, Jeffrey A., CHRISTMAS, Rowan, BUHLER, Jeremy, ENG, Jimmy K., BUMGARNER, Roger, GOODLETT, David R., AEBERSOLD, Ruedi, HOOD, Leroy. 2001b. «Integrated genomic and proteomic analyses of a systematically perturbed metabolic network». *Science*. 292. 929–934.

International Human Genome Mapping Consortium. 2001. «A physical map of the human genome». *Nature*. 409. 934–941.

ITO, Masao, KANO, Masanobu. 1982. «Long-lasting depression of parallel fiber-purkinje cell transmission induced by conjunctive stimulation of parallel fibers and climbing fibers in the cerebellar cortex». *Neuroscience Letters*. 33:3. 253–258.

JENSEN, Kurt. 1992. *Coloured Petri nets : basic concepts, analysis methods, and practical use*. Monographs on Theoretical Computer Science. New York : Springer-Verlag. 3 volumes.

- KANEHISA, Minoru, GOTO, Susumu, KAWASHIMA, Shuichi, OKUNO, Yasushi, HATTORI, Masahiro. 2004. «The KEGG resource for deciphering the genome». *Nucleic Acids Research*. 32: *Database Issue*. D277–D280.
- KARP, Peter D., RILEY, Monica, SAIER, Milton, PAULSEN, Ian T., COLLADO-VIDES, Julio, PALEY, Suzanne M., PELLEGRINI-TOOLE, Alida, BONAVIDES, César, GAMA-CASTRO, Socorro. 2002. «The EcoCyc database». *Nucleic Acids Research*. 30:1. 59–61.
- KAUFFMAN, Stuart A. 1993. *The Origins of order : Self-organization and selection in evolution*. New York : Oxford University Press. 734 p.
- KIKUCHI, Shinichi, FUJIMOTO, Kenji, KITAGAWA, Moriyuki, FUCHIKAWA, Taro, ABE, Michiko, OKA, Kotaro, TAKEI, Kohtaro, TOMITA, Masaru. 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.
- KIM, Junhyong. 2002. «Computers are from mars, organisms are from venus». *IEEE Computer*. 35:7. 25–32.
- KITANO, Hiroaki. 2002a. «Computational systems biology». *Nature*. 420. 206–210.
- KITANO, Hiroaki. 2002b. «Systems biology : A brief overview». *Science*. 295. 1662–1664.
- KUBOTA, Yoshihisa, BOWER, James M. 2001. «Transient versus asymptotic dynamics of CaM kinase II : possible roles of phosphatase». *Journal of Computational Neuroscience*. 11:3. 263–279.
- KULIKOVA, Tamara, ALDEBERT, Philippe, ALTHORPE, Nicola, BAKER, Wendy, BATES, Kirsty, BROWNE, Paul, VAN DEN BROEK, Alexandra, COCHRANE, Guy, DUGGAN, Karyn, EBERHARDT, Ruth, FARUQUE, Na-deem, GARCIA-PASTOR, Maria, HARTE, Nicola, KANZ, Carola, LEINONEN,

Rasko, LIN, Quan, LOMBARD, Vincent, LOPEZ, Rodrigo, MANCUSO, Renato, MCHALE, Michelle, NARDON, Francesco, SILVENTOINEN, Ville, STOEHR, Peter, STOESSER, Guenter, TULI, Mary Ann, TZOUVARA, Katerina, VAUGHAN, Robert, WU, Dan, ZHU, Weimin, APWEILER, Rolf. 2004. «The EMBL nucleotide sequence database». *Nucleic Acids Research*. 32: Database Issue. D27–D30.

KUMAR, Srikanta P., FEIDLER, Jordan C. 2003. «BioSPICE : A computational infrastructure for systems biology». *OMICS*. 7:3. 225.

KURATA, Hiroyuki, MATOBA, Nana, SHIMIZU, Natsumi. 2003. «CADLIVE for constructing a large-scale biochemical network based on a simulation-directed notation and its application to yeast cell cycle». *Nucleic Acids Research*. 31:14. 4071–4084.

KURODA, Shinya, SCHWEIGHOFER, Nicolas, KAWATO, Mitsuo. 2001. «Exploration of signal transduction pathways in cerebellar long-term depression by kinetic simulation». *Journal of Neuroscience*. 21:15. 5693–5702.

KüFFNER, Robert, ZIMMER, Ralf, LENGAUER, Thomas. 2000. «Pathway analysis in metabolic databases via differential metabolic display (DMD)». *Bioinformatics*. 16:9. 825–836.

LLEDO, Pierre-Marie, HJELMSTAD, Gregoy O., MUKHERJI, Sucheta, SODERLING, Thomas R., MALENKA, Robert C., NICOLL, Roger, A. 2002. «Calcium/calmodulin-dependent kinase II and long-term potentiation enhance synaptic transmission by the same mechanism». *Proceedings of the National Academy of Sciences of the United States of America*. 92:24. 11175–11179.

LOCKHART, David J., WINZELER, Elizabeth A. 2000. «Genomics, gene expression and dna arrays». *Nature*. 405. 827–835.

LOEW, Leslie M., SCHAFF, James C. 2002. «The Virtual Cell : a software environment for computational cell biology». *Trends in Biotechnology*. 19: 10.

401–406.

MATSUI, Mika, DOI, Atsushi, MATSUNO, Hiroshi, HIRATA, Yuichi, MIYANO, Satoru. 2001. «Biopathways model conversion from E-CELL to Genomic Object Net». *Genome Informatics*. 12. 290–291.

MATSUNO, Hiroshi, DOI, Atsushi, HIRATA, Yuichi, MIYANO, Satoru. 2001. «XML documentation of biopathways and their simulations in Genomic Object Net». *Genome Informatics*. 12. 54–62.

MATSUNO, Hiroshi, DOI, Atsushi, NAGASAKI, Masao, MIYANO, Satoru. 2000. «Hybrid Petri net representation of gene regulation network». *Proceedings of the Pacific Symposium on Biocomputing* t. 5. P. 338–349.

MATSUNO, Hiroshi, FUJITA, Sachie, DOI, Atsushi, NAGASAKI, Masao, MIYANO, Satoru. 2003a. «Towards biopathways modeling and simulation». *Proceedings of the 24th Conference on Applications and Theory of Petri Nets (ICATPN 2003)*. Sous la direction de VAN DER AALST, W. M. P., BEST, E. t. 2679 de *Lecture Notes in Computer Science*. Springer-Verlag. P. 3–22.

MATSUNO, Hiroshi, MURAKAMI, Ryutaro, YAMANE, Rie, YAMASAKI, Naoyuki, FUJITA, Sachie, HARUKA, Yoshimori, MIYANO, Satoru. 2003b. «Boundary formation by notch signaling in drosophila multicellular systems : Experimental observations and gene network modeling by Genomic Object Net». *Proceedings of the Pacific Symposium on Biocomputing* t. 8. P. 152–163.

MATSUNO, Hiroshi, TANAKA, Yukiko, AOSHIMA, Hitoshi, DOI, Atsushi, MATSUI, Mika, MIYANO, Satoru. 2003c. «Biopathways representation and simulation on hybrid functional Petri net». *In Silico Biology*. 3: 3. [En ligne]. [http ://www.bioinfo.de/isb/2003/03/0032/](http://www.bioinfo.de/isb/2003/03/0032/) (Page consultée le 6 août 2004).

MENDES, Pedro. 1993. «GEPASI : a software package for modelling the dynamics, steady states and control of biochemical and other systems». *Computer Applications in the Biosciences*. 9:5. 563–571.

- MENDES, Pedro. 1997. «Biochemistry by numbers : simulation of biochemical pathways with Gepasi 3». *Trends in Biochemical Science*. 22:9. 361–363.
- MEYER, T., HANSON, P. I., STRYER, L., SCHULMAN, H. 1992. «Calmodulin trapping by calcium-calmodulin-dependent protein kinase». *Science*. 256:5060. 1199–1202.
- MICHAELIS, Leonor, MENTEN, Maude Leonora. 1913. «Die kinetik der invertinwirkung». *Biochemische Zeitschrift*. 49. 333–369.
- MIGLIORE, Michele, LANSKY, Petr. 1999a. «Computational model of the effects of stochastic conditioning on the induction of long-term potentiation and depression». *Biological Cybernetics*. 81:4. 291–298.
- MIGLIORE, Michele, LANSKY, Petr. 1999b. «Long-term potentiation and depression induced by a stochastic conditioning of a model synapse». *Biophysical Journal*. 77:3. 1234–1243.
- Miyano Laboratory, University of Tokyo. 2003. «Genomic object net project». [En ligne]. <http://www.genomicobject.net> (Page consultée le 8 août 2004).
- MIYAZAKI, S., SUGAWARA, H., IKEO, K., GOJOBORI, T., TATENO, Y. 2004. «DDBJ in the stream of various biological data, nucleotide sequence database». *Nucleic Acids Research*. 32:DatabaseIssue. D31–D34.
- MURZINA, G. B., SIL'KIS, I. G. 1998. «Studies of long-term potentiation and depression of inhibitory transmission by mathematical modeling of post-synaptic processes». *Neuroscience and Behavioral Physiology*. 28:2. 121–129.
- NAGASAKI, Masao, DOI, Atsushi, MATSUNO, Hiroshi, MIYANO, Satoru. 2003. «Genomic Object Net : I. A platform for modeling and simulating biopathways,». *Applied Bioinformatics*. 2:3. 181–184.
- NGUYEN, Peter V., ABEL, Ted, KANDEL, Eric R. 1994. «Requirement of a critical period of transcription or induction of a late phase of LTP». *Science*. 265:5175. 1104–1107.

- O'DONOVAN, Claire, APWEILER, Rolf, BAIROCH, Amos. 2001. «The human proteomics initiative (HPI)». *TRENDS in biotechnology*. 19:5. 178–181.
- OKAMOTO, Hiroshi, ICHIKAWA, Kazuhisa. 2000. «Switching characteristics of a model for biochemical-reaction networks describing autophosphorylation versus dephosphorylation of Ca^{2+} /calmodulin-dependent protein kinase II». *Biological Cybernetics*. 82:1. 35–47.
- OLIVEIRA, Joseph S., BAILEY, Colin G., JONES-OLIVEIRA, Janet B., DIXON, David A. 2001. «An algebraic-combinatorial model for the identification and mapping of biochemical pathways». *Bulletin of Mathematical Biology*. 63:6. 1163–1196.
- OLIVEIRA, Joseph S., BAILEY, Colin G., JONES-OLIVEIRA, Janet B., DIXON, David A., GULL, Dean W., CHANDLER, Mary L. 2003. «A computational model for the identification of biochemical pathways in the Krebs cycle». *Journal of Computational Biology*. 10:1. 57–82.
- PANDEY, Akhilesh, MANN, Matthias. 2000. «Proteomics to study genes and genomes». *Nature*. 405. 837–845.
- PELEG, Mor, YEH, Iwei, ALTMAN, Russ B. 2002. «Modelling biological processes using workflow and Petri net models». *Bioinformatics*. 18:6. 825–837.
- Performability Engineering Research Group. 2004. «The Möbius tool». [En ligne]. <http://www.mobius.uiuc.edu/> (Page consultée le 8 août 2004).
- REDDY, Venkatramana N. 1994. *Modeling biological pathways : A discrete event systems approach*. 82 p. Mémoire de maîtrise en génie chimique, University of Maryland.
- REDDY, Venkatramana N., LIEBMAN, Michael N., MAVROVOUNIOTIS, Michael L. 1996. «Qualitative analysis of biochemical reaction systems». *Computers in biology and medicine*. 26:1. 9–24.

- REDDY, Venkatramana N., MAVROVOUNIOTIS, Michael L., LIEBMAN, Michael N. 1993. «Petri net representation in metabolic pathways». *Proceedings of the First International Conference on Intelligent Systems for Molecular Biology (ISMB)*. Sous la direction de HUNTER, L., SEARLS, D., SHAVLIK, J. AIII Press. P. 328–336.
- REISIG, Wolfgang. 1985. *Petri nets : an introduction*. Monographs on Theoretical Computer Science. New York : Springer-Verlag. 161 p.
- RISCH, Neil J. 2000. «Searching for genetic determinants in the new millennium». *Nature*. 405. 847–856.
- ROBILLARD, Pierre N., HARDY, Simon. 2001. *Les protéines : des molécules de génie, Rapport sur la recherche en bioinformatique, en protéomique et en génie protéique*. Montréal : École Polytechnique de Montréal. 105 p. EPM/RT-2001-04.
- ROSES, Allen D. 2000. «Pharmacogenetics and the practice of medicine». *Nature*. 405. 857–865.
- SALWINSKI, Lukasz, MILLER, Christopher S., SMITH, Adam J., PETTIT, Frank K., Bowie, James U., EISENBERG, David. 2004. «The database of interacting proteins : Update 2004». *Nucleic Acids Research*. 32: Database Issue. D449–D451.
- SARGSYAN, Armen R., MELKONYAN, Albert, A., PAPATHEODOROPOULOS, Costas, MKRTCHIAN, Hovhannes H., KOSTOPOULOS, George K. 2003. «A model synapse that incorporates the properties of short- and long-term synaptic plasticity». *Neural Networks*. 16:8. 1161–1177.
- SCHOMBURG, Ida, CHANG, Antje, EBELING, Christian, GREMSE, Marion, HELDT, Christian, GREGOR, Huhn, SCHOMBURG, Dietmar. 2004. «BRENDA, the enzyme database : updates and major new developments». *Nucleic Acids Research*. 32: Database Issue. D431–D433.

- SCHONBRUN, Jack, WEDEMEYER, William J., BAKER, David. 2002. «Protein structure prediction in 2002». *Current Opinion Structural Biology*. 12: 3. 348–354.
- SCHULER, Gregory D., EPSTEIN, Jonathan A., OHKAWA, Hitomi, KANS, Jonathan A. 1996. «Entrez : Molecular biology database and retrieval system». *Methods in Enzymology*. 266. 141–162.
- SHAROV, Andrei A. 1991. «Self-reproducing systems : structure, niche relations and evolution». *Biosystems*. 25:4. 237–249.
- SIVAKURAMAN, Sudhir, HARIHARAPUTRAN, Sridhar, MISHRA, Jyoti, BHALLA, Upinder S. 2003. «The database of quantitative cellular signaling : management and analysis of chemical kinetic models of signaling networks». *Bioinformatics*. 19:3. 408–415.
- SPENGLER, Sylvia J. 2000. «Bioinformatics in the information». *Science*. 287. 1221–1223.
- SRIVASTAVA, R., PETERSON, M. S., BENTLEY, W. E. 2001. «Stochastic kinetic analysis of escherichia coli stress circuit using σ^{32} -targeted antisense.» *Biotechnology and Bioengineering*. 75:1. 120–129.
- SRIVASTAVA, R., YOU, L., SUMMERS, J., YIN, J. 2002. «Stochastic versus deterministic modeling of intracellular viral kinetics.» *Journal of Theoretical Biology*. 218:3. 309–321.
- TAKAHASHI, K., ISHIKAWA, N., SADAMOTO, Y., SASAMOTO, H., OHTA, S., SHIOZAWA, A., MIYOSHI, F., NAITO, Y., NAKAYAMA, Y., TOMITA, M. 2003. «E-CELL 2 : Multi-platform E-CELL simulation system». *Bioinformatics*. 19:13. 1727–1729.
- TAKAHASHI, Kouichi, YUGI, Katsuyuki, HASHIMOTO, Kenta, YAMADA, Yohei, PICKETT, Christopher J. F., TOMITA, Masaru. 2002. «Computational

challenges in cell simulation : A software engineering approach». *IEEE Intelligent Systems*. 17:5. 64–71.

TOMITA, Masaru, HASHIMOTO, Kenta, TAKAHASHI, Kouichi, SHIMIZU, Thomas Simon, MATSUZAKI, Yuri, MIYOSHI, Fumihiki, SAITO, Kanako, TANIDA, Sakura, YUGI, Katsuyuki, VENTER, J. Craig, HUTCHISON III, Clyde A. 1999. «E-CELL : software environment for whole-cell simulation». *Bioinformatics*. 15:1. 72–84.

TSODYKS, Misha, PAWELZIK, Klaus, MARKRAM, Henry. 1998. «Neural networks with dynamic synapses». *Neural Computation*. 10:4. 821–835.

TSOKA, Sophia, OUZOUNIS, Christos A. 2000. «Recent developments and future directions in computational genomics». *FEBS Letters*. 480. 42–48.

VALK, Rüdiger. 2003. «Self-modifying nets, a natural extension of Petri nets». *Proceedings of the 5th Conference on Automata, Languages and Programming (ICALP 1978)*. Sous la direction de AUSIELLO, Giorgio, BÖHM, Corrado t. 62 de *Lecture Notes in Computer Science*. Springer-Verlag. P. 464–476.

VOSS, Klaus, HEINER, Monika, KOCH, Ina. 2003. «Steady states analysis of metabolic networks based on Petri nets». *In Silico Biology*. 3: 3. [En ligne]. [http ://www.bioinfo.de/isb/2003/03/0031/](http://www.bioinfo.de/isb/2003/03/0031/) (Page consultée le 6 août 2004).

VUKMIROVIC, Ognjenka Goga, TILGHMAN, Shirley M. 2000. «Exploring genome space». *Nature*. 405. 820–822.

WINGENDER, E., CHEN, X., FRICKE, E., GEFFERS, R., HEHL, R., LIEBICH, I., KRULL, M., MATYS, V., MICHAEL, H., OHNHÄUSER, R., PRÜß, M., SCHACHERER, F., THIELE, S., URBACH, S. 2001. «The TRANSFAC system on gene expression regulation». *Nucleic Acids Research*. 29:1. 281–283.

YANG, Shao-Nian, TANG, Yun-Gui, ZUCKER, Robert S. 1999. «Selective induction of LTP and LTD by postsynaptic $[Ca^{2+}]_i$ elevation». *Journal of Neurophysiology*. 81:2. 781–787.

- YUH, Chiou-Hwa, BOLOURI, Hamid, DAVIDSON, Eric H. 2001. «*Cis*-regulatory logic in the *endo16* gene : switching from a specification to a differentiation mode of control». *Development*. 128:5. 617–629.
- ZEVEDEI-OANCEA, Ionela, SCHUSTER, Stefan. 2003. «Topological analysis of metabolic networks based on Petri net theory». *In Silico Biology*. 3:3. [En ligne]. [http ://www.bioinfo.de/isb/2003/03/0029/](http://www.bioinfo.de/isb/2003/03/0029/) (Page consultée le 6 août 2004).