

Modélisation et simulation qualitative de réseaux de régulation génique

Mémoire présenté pour l'obtention de l'habilitation
à diriger des recherches, délivrée par
l'Université Joseph Fourier à Grenoble

Hidde de Jong

Institut National de Recherche en Informatique et en Automatique (INRIA)
Unité de recherche Rhône-Alpes
655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier Cedex
Courriel : Hidde.de-Jong@inrialpes.fr
Toile : <http://www-helix.inrialpes.fr/article165.html>

6 septembre 2004

Résumé

L'analyse des réseaux de régulation génique responsables du fonctionnement et du développement des organismes vivants est fortement stimulée par la mise à l'échelle génomique des méthodes expérimentales en biologie moléculaire. Outre ces méthodes à haut débit, des outils de modélisation et de simulation sont indispensables pour analyser la dynamique des réseaux de régulation génique. En effet, à cause du grand nombre de composants connectés par des boucles de rétroaction complexes, une compréhension intuitive du système est difficile à obtenir.

Une grande variété d'approches pour la modélisation et la simulation de réseaux de régulation génique a été proposée dans les quarante dernières années. L'application de ces méthodes est toutefois soumise à deux contraintes importantes. Premièrement, les mécanismes des réactions biochimiques impliquées dans les interactions du réseau ne sont souvent pas bien connus. Ceci implique que des modèles cinétiques détaillés ne peuvent en général pas être construits. Une deuxième contrainte vient de l'absence quasi totale d'informations quantitatives sur les paramètres cinétiques et les concentrations moléculaires. Par conséquent, les méthodes traditionnelles d'analyse numérique sont difficiles à appliquer.

Dans ce mémoire, je décris une méthode de modélisation et de simulation qualitative de réseaux de régulation génique, basée sur une classe d'équations différentielles linéaires par morceaux bien étudiée en biologie mathématique. Les modèles linéaires par morceaux permettent une description approchée des réseaux de régulation génique, particulièrement bien adaptée à l'état de l'art des techniques génomiques. Ces modèles possèdent de plus des propriétés mathématiques qui facilitent l'analyse qualitative du comportement des systèmes de régulation. La méthode a été implémentée en Java dans l'outil informatique Genetic Network Analyzer (GNA). L'application de GNA à l'analyse des processus de régulation génique bactériens – comme l'initiation de la sporulation chez *Bacillus subtilis* – a montré que la méthode est capable de reproduire le comportement qualitatif de réseaux de régulation complexes, mais déjà bien connus. Cette validation de la méthode nous a permis de continuer avec l'étude de systèmes expérimentaux dont le fonctionnement est moins bien compris, comme la régulation globale de la transcription chez *Escherichia coli* et *Synechocystis* PCC 6803.

Les travaux réalisés à ce jour forment une première étape dans le développement d'un environnement informatique intégré pour l'analyse de réseaux de régulation génique. Cet environnement permettrait non seulement la simulation du comportement d'un réseau à partir d'un modèle linéaire par morceaux donné, mais il soutiendrait également la construction et la révision de modèles, la sélection d'expériences discriminatoires et la validation de modèles par des données expérimentales. À plus long terme, il est même envisageable de généraliser cette approche à des modèles intégrés de réseaux d'interactions impliqués dans la régulation génique, le métabolisme et la transduction de signaux, afin d'obtenir une compréhension globale du fonctionnement de la cellule.

Abstract

The analysis of the genetic regulatory networks underlying the functioning and development of living organisms will much benefit from the recent upscaling to the genomic level of experimental methods in molecular biology. In addition to high-throughput experimental methods, computer tools for the modeling and simulation of gene regulation will be indispensable. As most networks of interest involve many genes connected through interlocking positive and negative feedback loops, an intuitive understanding of their dynamics is difficult to obtain.

A large variety of approaches for modeling and simulating genetic regulatory networks have been proposed in the past forty years. However, the application of these methods is submitted to two major constraints. First of all, the biochemical reaction mechanisms underlying the interactions are usually not or incompletely known. This generally prevents the formulation of detailed kinetic models. A second constraint arises from the almost complete absence of quantitative information on the kinetic parameters and molecular concentrations. As a consequence, traditional methods for numerical analysis are difficult to apply.

In this document, I describe a method for the qualitative modeling and simulation of genetic regulatory networks, based on a class of piecewise-linear differential equations that has been well-studied in mathematical biology. The piecewise-linear models provide an approximate description of genetic regulatory networks, particularly well-adapted to state-of-the-art techniques in genomics. In addition, these models have mathematical properties that facilitate the qualitative analysis of the behavior of regulatory systems. The method has been implemented in Java in the computer tool Genetic Network Analyzer (GNA). The application of GNA to the analysis of bacterial genetic regulatory processes – like the initiation of sporulation in *Bacillus subtilis* – has shown that the method is capable of reproducing the qualitative behavior of complex, but well-understood regulatory networks. This validation of the method has allowed us to continue with the study of experimental systems whose functioning is less understood, like the global regulation of transcription in *Escherichia coli* and *Synechocystis* PCC 6803.

The work completed up until now constitutes the first stage in the development of an integrated computer environment for the analysis of genetic regulatory networks. This environment would not only allow the simulation of the behavior of a network from a given piecewise-linear model, but also support the construction and revision of models, the selection of discriminatory experiments, and the validation of models by experimental data. In the long term, it might even be possible to generalize the approach to integrated models of the networks of interactions involved in gene regulation, metabolism, and signal transduction, in order to obtain a global comprehension of the functioning of the cell.

Table des matières

Préface	9
1 Introduction	11
1.1 Réseaux de régulation génique	11
1.2 Modélisation et simulation de réseaux	12
1.3 Modélisation et simulation qualitative	15
1.4 Plan	17
2 États de l'art	19
H. de Jong (2002), Modeling and simulation of genetic regulatory systems : A literature review, <i>Journal of Computational Biology</i> , 9(1) :69-105	21
H. de Jong, J. Geiselmann (2004), Modélisation et simulation de réseaux de régulation génique par des équations différentielles ordinaires, J.-F. Boulicaut, O. Gandrillon (dir.), <i>Informatique pour l'analyse du transcriptome</i> , Hermès, Paris, sous presse	23
H. de Jong (2003), Simulation qualitative, L. Travé-Massuyès, P. Dague (dir.), <i>Modèles et raisonnements qualitatifs</i> , Hermès, Paris, 269-329	25
3 Méthode de modélisation et de simulation	27
H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, J. Geiselmann (2004), Qualitative simulation of genetic regulatory networks using piecewise-linear models, <i>Bulletin of Mathematical Biology</i> , 66(2) :301-340	29
4 Outil informatique	31
H. de Jong, J. Geiselmann, C. Hernandez, M. Page (2003), Genetic Network Analyzer : Qualitative simulation of genetic regulatory networks, <i>Bioinformatics</i> , 19(3) :336-344	33
5 Applications	35
H. de Jong, J. Geiselmann, G. Batt, C. Hernandez, M. Page (2004), Qualitative simulation of the initiation of sporulation in <i>Bacillus subtilis</i> , <i>Bulletin of Mathematical Biology</i> , 66(2) :261-300	37

6 Discussion	39
6.1 Récapitulatif des résultats	39
6.2 Extensions de la méthode	40
6.3 Vers des cellules virtuelles	43
A Curriculum vitae	55
B Publications depuis 1998	61

Préface

La rédaction de ce mémoire m'a permis de revenir sur les six dernières années de ma vie professionnelle, depuis mon arrivée à l'INRIA fin 1998. Le sentiment dominant de ces années est celui d'un grand plaisir. En écrivant les chapitres du mémoire et en faisant le tri des articles à inclure, j'ai réalisé encore une fois à quel point il est passionnant de travailler à l'interface de l'informatique, des mathématiques et de la biologie. Si l'on peut espérer trouver de l'élégance abstraite dans les équations, les algorithmes, les théorèmes et les schémas conceptuels en informatique et en mathématiques, la biologie par contre est bien ancrée dans le concret avec ses méthodes et matériaux, ses résultats d'expériences et ses exceptions à la règle. Ainsi la position entre plusieurs domaines scientifiques m'a permis de satisfaire des plaisirs intellectuels divers et complémentaires.

Ceci ne devrait pas occulter le fait que mes efforts se sont parfois transformés en labeurs. Décidément, il n'est pas toujours facile de trouver un chemin allant des solutions de Filippov aux gènes rapporteurs, en passant par les écouteurs de Java! Chacun de ces domaines réserve bien des surprises désagréables, tout spécialement aux nouveaux arrivants. Heureusement j'ai eu la chance de rencontrer des collègues spécialistes en informatique, en mathématiques et en biologie prêts à m'aider. En rédigeant ce mémoire, il m'est devenu encore plus clair que les travaux accomplis sont le fruit d'un effort commun, au point qu'il m'était parfois difficile d'utiliser la première personne du singulier au lieu de la première personne du pluriel. Mis à part le fait que leur aide a été précieuse, les interactions avec mes collègues ont aussi ajouté un plaisir humain aux plaisirs intellectuels.

Je voudrais remercier Hans Geiselmann, Jean-Luc Gouzé, Michel Page et nombre d'autres collègues – dont notamment Grégory Batt, Richard Casey, Hanna Härdin, Céline Hernandez, Jean Houmard, Eva Laget, Delphine Ropers, Tewfik Sari et Dominique Schneider – pour leurs contributions aux travaux présentés dans ce mémoire. Je remercie également Olivier Bernard, Claudine Chaouiya, Damien Eveillard, Giancarlo Ferrari-Trecate, Hubert Garavel, Radu Mateescu, Nicolaas Mars, Anne Morgat, Sébastien Provencher, François Rechenmann, Marie-France Sagot, Denis Thieffry, Ivayla Vatcheva et de nombreuses autres personnes pour des collaborations et des discussions autour et dans la suite du sujet de ce mémoire. Je dois beaucoup à François Rechenmann, à Françoise de Coninck et aux autres membres du projet HELIX pour avoir créé une ambiance de travail ouverte et conviviale. Finalement, je remercie François Rechenmann d'avoir accepté d'être mon directeur des travaux pour cette habilitation à diriger des recherches (HDR), Jean Della Dora d'avoir voulu assumer le rôle de président de mon jury d'HDR et Leon Glass, Albert Goldbeter,

Nicolaas Mars et Jan van Schuppen d'avoir accepté de faire partie du jury en tant que membre ou rapporteur.

Ce mémoire est dédié à mes parents, en signe de reconnaissance pour leur amour et leur soutien infailibles.

Grenoble, le 23 août 2004

Chapitre 1

Introduction

1.1 Réseaux de régulation génique

Une des caractéristiques les plus remarquables de la biologie moléculaire aujourd'hui est la mise à l'échelle génomique de ses méthodes expérimentales. Le séquençage de génomes complets, impensable il y a vingt ans, est devenu un travail de routine, hautement automatisé et réalisé dans un environnement quasi-industriel. La miniaturisation des technologies d'hybridation a produit les puces à ADN, des outils pour mesurer le niveau d'expression des gènes [38], tandis que le développement des méthodes protéomiques – basées sur l'électrophorèse bidimensionnelle, la spectrométrie de masse et le système double-hybride – permet l'identification des protéines et leurs interactions à l'échelle de génomes entiers [49].

Ces nouvelles méthodes génomiques produisent d'énormes quantités de données sur différents aspects de la cellule. D'une part, elles identifient les interactions entre les gènes, les protéines, les métabolites et d'autres molécules, dressant ainsi la structure des réseaux d'interactions. D'autre part, elles détectent l'évolution de l'état de la cellule, c'est-à-dire, les variations de la concentration et de la localisation des différents composants moléculaires au cours du temps, en fonction de changements dans l'environnement. Le grand défi de la *biologie fonctionnelle* ou la *biologie des systèmes* (« *systems biology* ») consiste à mettre en relation ces données structurales et fonctionnelles afin d'arriver à une interprétation globale du fonctionnement d'un organisme [32, 36]. Il s'agit de prédire et comprendre comment le comportement observé de l'organisme – l'adaptation à son milieu, la différenciation cellulaire lors de son développement, voire son évolution sur des périodes plus longues – peut émerger d'un réseau d'interactions moléculaires.

Ces interactions sont de nature très hétérogène. Elles interviennent aussi bien au niveau de la transcription et de la traduction d'un gène, qu'au niveau de la dégradation enzymatique d'un métabolite et la phosphorylation d'une protéine régulatrice. Dans l'étude d'un processus cellulaire particulier, il suffit souvent, au moins dans un premier temps, de se focaliser sur une partie du réseau d'interactions, caractérisée par un type donné d'interaction. Ainsi, les biologistes ont pris l'habitude de distinguer, entre autres, les *réseaux métaboliques*, les *réseaux de transduction des signaux* et les *réseaux de régulation génique*.

Les réseaux métaboliques assurent la connection entre les petites molécules de la cellule à travers des réactions enzymatiques, tandis que la transduction des signaux passe en grande partie par la modification post-traductionnelle des protéines. Les réseaux de régulation génique concernent surtout les interactions entre protéines et acides nucléiques, contrôlant la transcription et la traduction des gènes.

Dans les travaux présentés dans ce mémoire, je me suis focalisé sur les réseaux de régulation génique. Ces réseaux sont responsables du contrôle d'un grand nombre de processus importants dans le développement et le fonctionnement d'un organisme. Par exemple, ils permettent de changer le programme génétique de la cellule bactérienne, l'ensemble des gènes exprimés et leur niveau d'expression, en réponse à un stress imposé par les conditions du milieu. Aussi, une grande partie des données expérimentales disponibles aujourd'hui, par exemple les données du transcriptome, portent sur les réseaux de régulation génique. Toutefois, il est important de ne pas oublier que les réseaux de régulation génique sont imbriqués dans d'autres réseaux, parfois à un point tel qu'il est difficile de les distinguer.

1.2 Modélisation et simulation de réseaux de régulation génique

Outre des méthodes expérimentales à haut débit, des approches mathématiques et informatiques sont indispensables pour analyser les réseaux de régulation génique. À cause du grand nombre de composantes et d'interactions, connectées par des boucles de rétroaction positive et négative, une compréhension intuitive de la dynamique du système est souvent difficile, voire impossible à obtenir. La *modélisation* et la *simulation* par ordinateur peuvent contribuer à l'élucidation d'un réseau de régulation génique en permettant au biologiste de tester un nombre plus restreint d'hypothèses plausibles. La formulation de modèles mathématiques nécessite une description préalable, explicite et non ambiguë, des hypothèses faites concernant les mécanismes de régulation. En outre, la simulation à partir de ces modèles génère des prédictions sur le comportement de la cellule, qui peuvent être vérifiées expérimentalement.

Une démarche pour analyser la dynamique d'un réseau de régulation génique – basée sur l'application coordonnée des outils expérimentaux, mathématiques et informatiques – est résumée dans la figure 1.1. Dans un premier temps, un ou plusieurs modèles initiaux sont construits à partir des connaissances du système et des observations de son comportement. Ces modèles sont utilisés pour faire des simulations, sous des conditions expérimentales choisies. Les prédictions ainsi obtenues sont ensuite comparées avec les observations faites sous les mêmes conditions expérimentales, ce qui donne une indication de l'adéquation des modèles et peut conduire à la révision de ces derniers. Ceci inaugure un nouveau cycle de simulation du comportement du système, de vérification expérimentale des prédictions et de révision des modèles.

La modélisation et la simulation des réseaux de régulation génique remontent aux débuts de la biologie moléculaire, à la suite des travaux pionniers de F. Jacob et de J. Monod

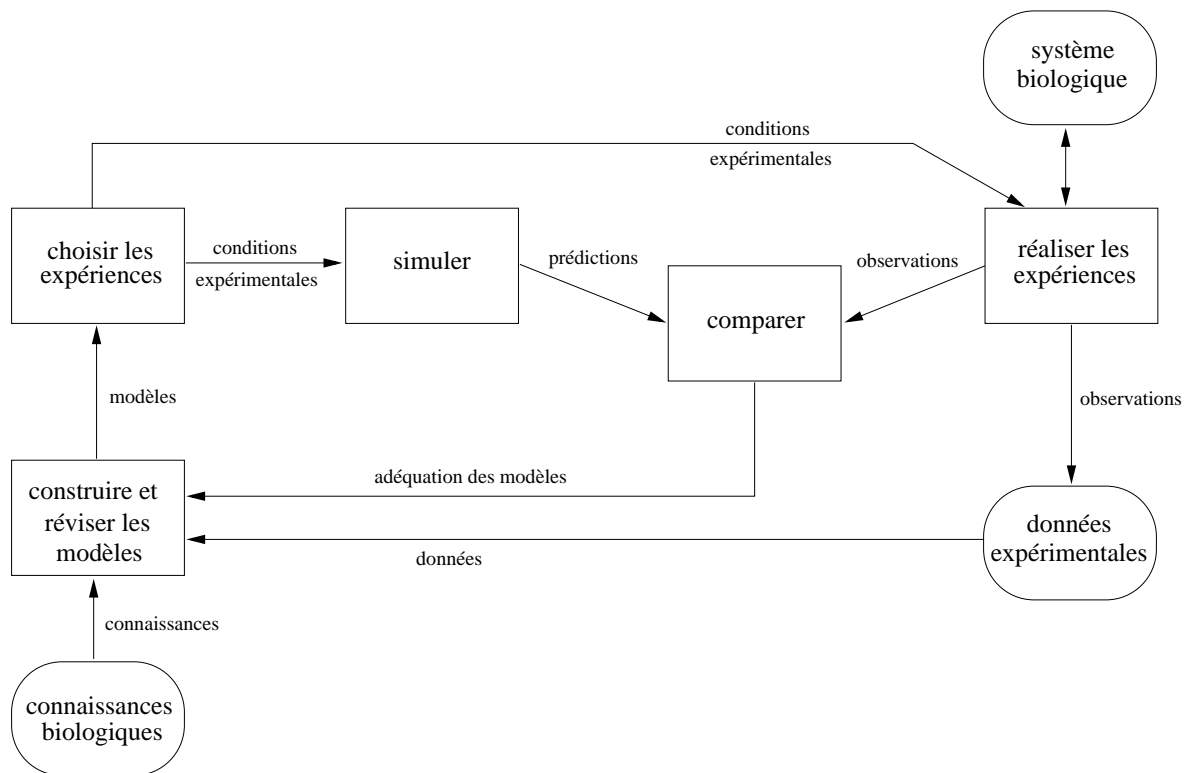


FIG. 1.1 – Démarche pour analyser la dynamique de réseaux de régulation génique : application coordonnée des outils expérimentaux, mathématiques et informatiques.

sur la régulation de l'opéron *lac* [33, 50]. L'un des résultats marquants des années soixante, le modèle d'autorégulation génique de Goodwin [27], est essentiellement une formalisation des travaux expérimentaux cités ci-dessus et leur généralisation à d'autres systèmes de régulation bactériens [45]. Un autre développement majeur était l'idée, proposée par M. Sugita [62], d'interpréter les gènes et leurs interactions en termes logiques, ce qui a ouvert la voie aux réseaux Booléens et leurs généralisations [35, 65]. Les travaux d'A. Turing sur la modélisation de la morphogénèse [70] doivent aussi être mentionnés ici. Réalisés avant même la publication de la structure de double hélice de l'ADN [76], ils ont gardé leur intérêt avec l'élucidation des bases moléculaires du développement. Si la modélisation et la simulation de réseaux de régulation génique sont donc dotées d'une longue histoire, il est toutefois clair que depuis quelques années, avec la fin annoncée de grands projets de séquençage, elles sont entrées dans une nouvelle phase. La modélisation et la simulation suscitent aujourd'hui un intérêt sans précédent, dont témoignent la croissance exponentielle du nombre de publications dans le domaine et la multiplication des conférences et des ateliers spécialisés.

Même si une grande diversité de modèles de réseaux de régulation génique et, plus généralement, de réseaux d'interactions moléculaires a été proposée, deux types de modèles de réseaux sont le plus répandus aujourd'hui : les équations différentielles et les équations maîtresses stochastiques (voir [9, 29, 41, 60] pour des revues). Ces modèles permettent des descriptions détaillées de réseaux de régulation génique, descendant jusqu'au niveau moléculaire. En outre, les modèles différentiels et stochastiques peuvent être utilisés afin de faire des prédictions précises et numériques du comportement d'un système de régulation.

Un grand nombre d'excellents exemples de l'application des modèles différentiels et stochastiques aux réseaux procaryotes et eucaryotes est disponible dans la littérature. H. McAdams et L. Shapiro [42] ont simulé le choix entre lyse et lysogénie de la bactériophage λ par un modèle différentiel, tandis que A. Arkin et collègues [3] ont étudié le même système à l'aide d'un modèle stochastique. Dans une série de publications, les groupes de B. Novak et de J. Tyson ont développé des modèles différentiels des mécanismes cinétiques impliqués dans la régulation du cycle cellulaire chez le crapaud *Xenopus laevis* [47] et la levure *Schizosaccharomyces pombe* [46] (voir [71] pour une revue). Des modèles différentiels de la segmentation chez *Drosophila melanogaster* ont contribué à mieux comprendre la formation des profils d'expression spatio-temporels des gènes *gap*, *pair-rule* et *segment polarity* le long de l'axe antéro-postérieur de l'embryon de la mouche [53, 54, 75].

L'application des modèles différentiels et stochastiques est toutefois soumise à deux contraintes importantes [29]. Premièrement, les mécanismes des réactions biochimiques impliquées dans les interactions du réseau ne sont souvent pas bien connus. Ceci a pour conséquence que des modèles cinétiques détaillés ne peuvent pas être construits. Une deuxième contrainte vient de l'absence générale d'informations quantitatives sur les paramètres cinétiques et les concentrations moléculaires, même dans le cas des systèmes biologiques bien étudiés. Par conséquent, les méthodes traditionnelles de l'analyse numérique sont difficiles à appliquer.

Les deux contraintes mentionnées ci-dessus génèrent une demande de méthodes basées sur des modèles approchés qui, tout en faisant abstraction des mécanismes moléculaires

précis des interactions, prennent en compte des aspects essentiels de la régulation génique. En outre, ces méthodes devraient permettre l'analyse qualitative de la dynamique des réseaux de régulation génique. Mes recherches depuis mon arrivée à l'INRIA, fin 1998, se sont portées sur le développement d'une telle méthode, son implémentation dans un outil informatique convivial et son application à des systèmes biologiques réels en collaboration avec des biologistes expérimentaux.

1.3 Modélisation et simulation qualitative

Le développement d'une méthode de modélisation et de simulation de réseaux de régulation génique, satisfaisant les contraintes relevées dans la section précédente, ne reposait pas sur rien. Je pouvais me baser sur l'expérience et les ressources accumulées en biologie mathématique et en bioinformatique, ainsi qu'en mathématiques et en informatique plus généralement. Deux approches m'ont particulièrement guidé dans mes recherches : l'analyse des réseaux de régulation génique à l'aide de modèles d'équations différentielles linéaires par morceaux et la simulation qualitative de systèmes dynamiques.

Au début des années soixante-dix, L. Glass et S. Kauffman ont proposé de modéliser les réseaux de régulation génique par une classe d'*équations différentielles linéaires par morceaux* [26]. Les variables d'état dans ces modèles correspondent aux concentrations des protéines codées par les gènes du réseau, tandis que les équations différentielles représentent les interactions par lesquelles certaines protéines influencent la synthèse et la dégradation d'autres protéines. La forme des interactions est approchée par des fonctions en escalier, ce qui donne la structure linéaire par morceaux aux équations différentielles. Ces modèles possèdent des propriétés mathématiques qui facilitent l'analyse qualitative du comportement des systèmes de régulation [26, 28, 44, 61]. Sur le plan formel, les modèles linéaires par morceaux sont apparentés aux modèles logiques asynchrones proposés par R. Thomas et collègues [66, 67]. Les modèles linéaires par morceaux et logiques ont donné des résultats pertinents lors de l'étude de nombreux réseaux de régulation procaryotes et eucaryotes (*e.g.*, [1, 24, 43, 48, 51, 56, 64]).

Le manque d'informations précises sur les relations fonctionnelles entre les composantes du système et sur la valeur des paramètres est un problème que l'on retrouve dans de nombreux domaines scientifiques et techniques. Ceci a suscité un intérêt pour les méthodes de *simulation qualitative*, développées pour des situations dans lesquelles l'application des méthodes de simulation numérique n'est pas possible ou pas appropriée [10]. Ainsi, la méthode QSIM, développée par B. Kuipers [37], permet de définir des modèles sous forme d'*équations différentielles qualitatives*. Ces dernières sont des abstractions d'équations différentielles ordinaires, dans lesquelles les valeurs des variables et des paramètres sont spécifiées par des valeurs qualitatives, notamment des signes, tandis que les relations fonctionnelles entre les variables et les paramètres sont transformées en contraintes entre valeurs qualitatives. La simulation qualitative prédit des comportements qualitatifs possibles du système à partir de l'équation différentielle qualitative et des conditions initiales, elles-mêmes spécifiées également de façon qualitative. Il existe quelques exemples de l'application des méthodes

de simulation qualitative aux réseaux de régulation en biologie [30, 34, 69]. Des méthodes ayant des idées similaires à la base ont été proposées pour l'analyse de systèmes hybrides [2, 24].

Dans mes travaux, j'ai fait une synthèse des approches résumées dans les deux paragraphes précédents, en développant une méthode de modélisation et de simulation qualitative basée sur les modèles d'équations différentielles linéaires par morceaux. D'une part, les modèles linéaires par morceaux permettent une description approchée, bien adaptée aux données expérimentales disponibles aujourd'hui. D'autre part, la simulation qualitative apporte un cadre informatique au développement d'une méthode capable de prédire, à partir des modèles linéaires par morceaux, les comportements qualitatifs possibles du réseau. Aussi, les modèles linéaires par morceaux posent des contraintes mathématiques fortes sur la dynamique locale du système. Ceci porte un remède au talon d'Achille des méthodes de simulation qualitative : l'explosion combinatoire du nombre de comportements qualitatifs possibles dans le cas de l'analyse des systèmes complexes rencontrés en pratique.

Dès le début, il me paraissait clair que, si la méthode de modélisation et de simulation qualitative devait avoir un impact dans la communauté des biologistes, il fallait l'implémenter dans un outil informatique convivial. Avec mes collègues, nous avons donc fait de grands efforts pour développer un logiciel appelé *Genetic Network Analyzer (GNA)*. Les utilisateurs visés de GNA sont des modélisateurs de systèmes biologiques, qui sont supposés comprendre les concepts mathématiques à la base de la méthode, sans être pour autant des experts en systèmes dynamiques. GNA a été programmé en Java, parce que ce langage est devenu le standard dans la programmation par objets et il permet de développer de façon relativement facile une interface graphique, composant critique de GNA. En outre, la portabilité de Java fait que GNA peut être utilisé, sans adaptations spécifiques, sur différentes plates-formes.

L'utilité de la méthode est finalement déterminée par son application à des réseaux de régulation génique réellement étudiés. Qu'est-ce que la modélisation et la simulation qualitative pourraient apporter aux biologistes? Afin de répondre à cette question, j'ai procédé en deux étapes. Dans un premier temps, la méthode a été validée au travers de son application à des processus dont le fonctionnement est déjà bien connu, comme l'infection d'une cellule d'*Escherichia coli* par le bactériophage λ et l'initiation de la sporulation chez *Bacillus subtilis*. Dans un deuxième temps, la méthode est appliquée (en ce moment même) à des systèmes qui sont pour l'instant moins bien, voire mal compris, et qui sont activement étudiés aujourd'hui, comme la régulation globale de la transcription chez *E. coli* et *Synechocystis* PCC 6803, ainsi que l'induction de la voie lytique par une cellule d'*E. coli* infectée par le bactériophage Mu.

Les applications choisies concernent uniquement des processus de régulation bactériens. Même si la méthode de modélisation et de simulation n'est pas restreinte aux organismes procaryotes, il me semblait préférable de commencer avec des systèmes *a priori* moins compliqués que ceux trouvés chez les organismes eucaryotes. Déjà, la modélisation mathématique de réseaux de régulation bactériens – en particulier ceux dont le fonctionnement est peu compris par les expérimentalistes – représente un grand défi pour la biologie des systèmes. L'analyse de réseaux eucaryotes ajouterait des niveaux de complexité supplé-

mentaires, en particulier en ce qui concerne la complexité de la régulation génique et les interactions entre différentes cellules de l'organisme. Une autre raison pour se focaliser, au moins dans un premier temps, sur les procaryotes est d'ordre plus pratique : les biologistes avec qui j'entretiens des collaborations, travaillent sur des systèmes procaryotes (chapitre 5).

1.4 Plan

Le mémoire est structuré en six chapitres, ce chapitre d'introduction inclus. À l'exception des premiers et dernier, chaque chapitre est composé d'une reproduction de quelques publications, précédée par un court résumé.

Le deuxième chapitre est entièrement dédié aux états de l'art de la modélisation et de la simulation de réseaux de régulation génique, ainsi que de la simulation qualitative. Ces études de la littérature m'ont donné des orientations pour le développement de la méthode de modélisation et de simulation qualitative, décrite en détail dans le troisième chapitre du mémoire. L'implémentation de la méthode dans l'outil informatique GNA est le sujet du quatrième chapitre, tandis que son application à l'analyse de l'initiation de la sporulation chez *B. subtilis* est présentée dans le cinquième chapitre. Dans le sixième et dernier chapitre, je résumerai les résultats obtenus et je les placerai dans un contexte plus large, esquissant ainsi le programme de mes recherches à court et à moyen terme. Quelques parties de ce programme sont en cours de réalisation depuis un ou deux ans.

Les appendices de ce mémoire contiennent mon curriculum vitae et une liste de toutes mes publications depuis l'obtention de mon doctorat en 1998.

Chapitre 2

États de l'art

Ce chapitre contient trois revues de la littérature résumant l'état de l'art dans deux domaines directement concernés par le sujet de ce mémoire : d'une part, la modélisation et la simulation de réseaux de régulation génique, d'autre part, la simulation qualitative.

La première revue, parue dans le *Journal of Computational Biology* en 2002 [9], résume une dizaine d'approches pour la modélisation et la simulation de réseaux de régulation génique. Ces approches vont des modèles statiques et déterministes (graphes dirigés) aux modèles dynamiques et stochastiques (équations stochastiques maîtresses). J'avais commencé la revue avec l'idée que sa rédaction ne devait pas prendre trop de temps, vu la nouveauté supposée du sujet. Cependant, au fil et à mesure que je découvrais la vaste littérature sur l'analyse mathématique de réseaux de régulation génique, accumulée depuis les années soixante, l'étendue de la tâche m'est apparue clairement. J'ai décidé malgré tout de continuer à m'investir dans ce travail bibliographique car, si les publications anciennes sont souvent ignorées dans la littérature récente, beaucoup d'entre elles sont restées pertinentes. La conclusion principale de la revue est que la plupart des méthodes existantes de modélisation et de simulation sont difficiles à appliquer, parce qu'elles ont du mal à traiter les données sur la régulation génique actuellement disponibles, qui sont incomplètes et essentiellement qualitatives.

La deuxième revue, un chapitre écrit avec J. Geiselmann (Université Joseph Fourier, Grenoble) pour le livre *Informatique pour l'analyse du transcriptome*, qui paraîtra en 2004 chez Hermès, est d'une portée moins large [13]. Elle se focalise sur la classe de modèles les plus utilisés pour modéliser et simuler les réseaux de régulation génique : les équations différentielles ordinaires. Plus précisément, la revue compare les modèles linéaires par morceaux avec, d'une part, les modèles non linéaires et, d'autre part, les modèles linéaires. Elle a pour conclusion que les modèles linéaires par morceaux forment un bon compromis entre l'adéquation empirique des modèles non linéaires et les propriétés mathématiques favorables des modèles linéaires. Ils sont donc un point de départ prometteur pour le développement d'une méthode de modélisation et de simulation qualitative. Une traduction anglaise du chapitre paraîtra dans le livre *Genomic Signal Processing and Statistics*, publié par l'EURASIP [12].

La simulation qualitative est le sujet de la troisième revue de ce chapitre. Cette revue

figure dans la nouvelle édition du livre *Modèles et raisonnements qualitatifs*, paru chez Hermès en 2003 [10]. La revue présente les méthodes classiques de simulation qualitative et met en évidence les problèmes qu'elles ont rencontrées, suivi par une discussion de deux approches alternatives : l'analyse de l'espace de phase et la simulation semi-quantitative. Je me suis inspiré de la première alternative, qui interprète la simulation qualitative dans le cadre de la théorie des systèmes dynamiques. Les méthodes d'analyse de l'espace de phase se focalisent sur des classes de systèmes pour lesquelles existent des contraintes fortes sur les comportements possibles. En adaptant à ces contraintes la représentation de la dynamique du système ainsi que les algorithmes de prédiction, ces méthodes essaient de limiter le nombre de comportements qualitatifs possibles prédits. Une traduction anglaise de la revue de simulation qualitative a été acceptée par la *Knowledge Engineering Review* [11].

J'ai rédigé ou co-rédigé d'autres revues de la littérature, notamment en collaboration avec J. Geiselmann et D. Thieffry (Université de la Méditerranée, Marseille) ([16, 63]; voir l'appendice B pour une liste complète).

H. de Jong (2002), Modeling and simulation of genetic regulatory systems : A literature review, *Journal of Computational Biology*, 9(1) :69-105

H. de Jong, J. Geiselman (2004), Modélisation et simulation de réseaux de régulation génique par des équations différentielles ordinaires, J.-F. Boulicaut, O. Gandrillon (dir.), *Informatique pour l'analyse du transcriptome*, Hermès, Paris, sous presse

H. de Jong (2003), Simulation qualitative, L. Travé-Massuyès, P. Dague (dir.), *Modèles et raisonnements qualitatifs*, Hermès, Paris, 269-329

Chapitre 3

Méthode de modélisation et de simulation qualitative

Les revues de la littérature du chapitre 2 donnent quelques directions à explorer pour le développement d'une méthode de modélisation et de simulation de réseaux de régulation génique adaptée aux données biologiques disponibles aujourd'hui. Notamment, elles suggèrent une classe de modèles appropriée – les modèles d'équations différentielles linéaires par morceaux –, ainsi qu'un cadre informatique pour la prédiction de la dynamique qualitative des systèmes décrits par cette classe de modèles – la simulation qualitative.

Ce chapitre contient une description détaillée de la méthode de modélisation et de simulation qualitative que j'ai développée en collaboration avec M. Page (INRIA Rhône-Alpes et Université Pierre Mendès-France, Grenoble). Il s'agit d'un article qui vient de paraître dans le *Bulletin of Mathematical Biology* [19]. L'article a été précédé par quelques versions préliminaires, publiées dans les actes des conférences en informatique, comme l'*International Joint Conference on Artificial Intelligence* [21], l'*European Conference on Artificial Intelligence* [17, 20] et l'*International Workshop on Hybrid Systems* [18]. Avec le recul, ces versions étaient limitées par des problèmes mathématiques, liés aux discontinuités dans le membre droit des équations différentielles linéaires par morceaux. Aussi, elles étaient difficiles à comprendre à cause des problèmes conceptuels liés à la nature des abstractions qualitatives. Ces problèmes ont été résolus, en grande partie, dans l'article repris dans ce chapitre, grâce à une collaboration avec J.-L. Gouzé (INRIA Sophia-Antipolis) et T. Sari (Université de Haute Alsace, Mulhouse). La collaboration est soutenue par une Action de Recherche Coopérative (ARC) INRIA (2002-2004) et une Action Concertée Incitative (ACI) IMPBio (2003-2006).

L'article paru dans le *Bulletin of Mathematical Biology* montre que, pour la classe d'équations différentielles linéaires par morceaux considérée, l'espace de phase peut être partitionné en régions dans lesquelles le système se comporte de façon qualitativement homogène. À chacune de ces régions, nous associons un *état qualitatif*. Deux états qualitatifs peuvent être reliés par une *transition* s'il existe une solution, débutant dans la région correspondant au premier état, qui atteint la région correspondant au deuxième état, sans passer par une troisième région. L'ensemble des états qualitatifs et des transitions

entre états qualitatifs forme un *graphe de transition*, dont les chemins représentent les comportements qualitatifs possibles du système.

De façon générale, le graphe de transition associé à un modèle linéaire par morceaux diffère en fonction des valeurs des paramètres. Cependant, nous montrons dans l'article qu'il est possible de définir une classe de modèles, déterminée par des contraintes d'inégalité entre paramètres. Sous certaines conditions, pas trop restrictives, chaque modèle dans cette classe produit le même graphe de transition. Nous proposons alors une méthode de simulation qualitative qui permet de calculer, de façon symbolique, le graphe de transition à partir d'un *modèle qualitatif*, c'est-à-dire, un modèle d'équations différentielles linéaires par morceaux assorti des contraintes d'inégalité sur les paramètres.¹ La méthode est correcte, dans le sens où le graphe de transition est garanti de couvrir tous les comportements qualitatifs compatibles avec la structure du système. La méthode de simulation qualitative a été implémentée dans l'outil informatique *Genetic Network Analyzer (GNA)* (chapitre 4).

¹Plus précisément, la méthode calcule la partie du graphe de transition *atteignable* à partir d'un ensemble d'états qualitatifs initiaux.

H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, J. Geiselman (2004), Qualitative simulation of genetic regulatory networks using piecewise-linear models, *Bulletin of Mathematical Biology*, **66(2)** :301-340

Chapitre 4

Outil informatique

La méthode de modélisation et de simulation qualitative du chapitre 3 a été implémentée en Java dans un outil appelé *Genetic Network Analyzer (GNA)*. Le logiciel a été réalisé en collaboration avec M. Page et C. Hernandez (INRIA Rhône-Alpes, maintenant à l'Institut Suisse de Bioinformatique, Genève). GNA a été déposé auprès de l'APP et est disponible gratuitement à des fins de recherches pour des utilisateurs académiques (<http://www-helix.inrialpes.fr/gna>). La version actuelle et les version précédentes ont été distribuées à plus de 100 laboratoires en France et à l'étranger. Le premier article reproduit dans ce chapitre, paru dans la revue *Bioinformatics* en 2003, décrit la version 5.0 de GNA et son utilisation [15].¹ La version actuelle de GNA, la version 5.5, diffère de la version précédente par le renouvellement de l'interface graphique, appelée VisualGNA.

Afin de pouvoir effectuer une simulation qualitative, GNA a besoin d'un modèle qualitatif du réseau – équations différentielles linéaires par morceaux et contraintes d'inégalité entre paramètres – ainsi qu'un ou plusieurs états qualitatifs initiaux. Dans les versions précédentes, ces entrées prenaient la forme de fichiers texte. La version actuelle du programme, développée par M. Page, permet à l'utilisateur de spécifier le modèle et les états initiaux au travers de VisualGNA. Le résultat d'une simulation qualitative est le graphe de transition resumant la dynamique qualitative du réseau. VisualGNA offre plusieurs possibilités d'analyser les comportements qualitatifs contenus dans ce graphe. Les fonctions du programme sont détaillées dans la documentation HTML incluse dans la distribution.

¹Au moment de la rédaction de ce mémoire, l'article sur GNA figure parmi les sept articles les plus consultés sur le site web de *Bioinformatics* dans les cinq dernières années (<http://www3.oup.co.uk/jnls/list/cabios/special/15/default.html>).

H. de Jong, J. Geiselman, C. Hernandez, M. Page (2003),
Genetic Network Analyzer : Qualitative simulation of ge-
netic regulatory networks, *Bioinformatics*, 19(3) :336-344

Chapitre 5

Applications

La méthode de modélisation et de simulation qualitative, ainsi que son implémentation, ont été utilisées pour l'analyse de plusieurs réseaux de régulation génique chez les procaryotes. Ces applications ont été réalisées en étroite collaboration avec le laboratoire de J. Geiselmann (Université Joseph Fourier, Grenoble). Dans ce chapitre, j'ai inclus un article décrivant les résultats d'une étude de l'initiation de la sporulation chez *Bacillus subtilis*. Cet article vient de paraître dans le *Bulletin of Mathematical Biology* [14].

Dans des conditions de privation de nutriments, la bactérie du sol *Bacillus subtilis* peut abandonner la croissance végétative et former une spore dormante et résistante aux facteurs environnementaux. La décision entre la croissance végétative et la sporulation est contrôlée par un réseau de régulation génétique complexe, intégrant divers signaux environnementaux, signaux du cycle cellulaire et signaux métaboliques. Malgré la quantité énorme de travaux réalisés pour élucider ce réseau, très peu de données quantitatives sur les paramètres cinétiques et les concentrations moléculaires sont disponibles. Nous avons construit un modèle qualitatif de ce réseau, composé de neuf équations différentielles linéaires par morceaux et une soixantaine de contraintes de paramètres, à une exception près déterminées par des données biologiques. À partir de ce modèle, nous avons été capables de reproduire les caractéristiques essentielles du choix entre la croissance végétative et la sporulation, en particulier le rôle joué par des boucles de rétroaction positives et négatives. Cette application à un système dont le fonctionnement est bien compris, au moins qualitativement, a permis la validation de la méthode et de l'outil de simulation.

J'ai également construit un modèle de l'un des réseaux de régulation génique les plus connus en biologie moléculaire : la décision entre lyse et lysogénie suivant l'infection d'une cellule d'*E. coli* par le bactériophage λ . Ce modèle, une adaptation du modèle précédemment développé par D. Thieffry [64], n'a pas été publié, mais est utilisé dans les travaux pratiques sur GNA qui accompagnent mes cours. À l'heure actuelle, la méthode et son implémentation sont utilisées dans notre équipe et dans d'autres équipes – par exemple, l'ETH à Zürich [72] – pour analyser des systèmes dont le fonctionnement n'est que partiellement compris (section 6.1).

H. de Jong, J. Geiselman, G. Batt, C. Hernandez, M. Page (2004), Qualitative simulation of the initiation of sporulation in *Bacillus subtilis*, *Bulletin of Mathematical Biology*, 66(2) :261-300

Chapitre 6

Discussion

6.1 Récapitulatif des résultats

Dans ce mémoire, j'ai présenté une approche de modélisation et de simulation de réseaux de régulation génique sur laquelle j'ai travaillé depuis mon arrivée à l'INRIA, fin 1998. La méthode est basée sur deux idées principales : la modélisation de réseaux de régulation génique par des équations différentielles linéaires par morceaux et la simulation qualitative des comportements de ces réseaux. Le résultat est une méthode bien adaptée aux données expérimentales actuellement disponibles, caractérisées par l'absence d'informations sur les mécanismes biochimiques des interactions et les valeurs numériques des paramètres (section 1.2). La méthode a une base mathématique solide, développée depuis le début des années soixante-dix par L. Glass et d'autres [26]. Elle permet le développement d'algorithmes de simulation qualitative performants, capables de traiter les réseaux de régulation génique complexes rencontrés en pratique.

Cependant, l'approche que j'ai présentée dans ce mémoire a un prix. La nature approchée des modèles, ainsi que le caractère qualitatif des prédictions de la méthode, ont pour conséquence que certains types de questions restent sans réponse. L'article sur la modélisation de l'initiation de la sporulation chez *B. subtilis*, inséré dans le chapitre 5, en donne des exemples. Il est possible de prédire que la concentration d'une certaine protéine augmente suite à une perturbation, mais la simulation qualitative ne donne pas de valeurs de concentration numériques. Aussi, la graphe de transition spécifie quelles séquences d'états qualitatifs le système peut suivre à partir d'un état initial donné, mais elle ne permet pas de déterminer laquelle d'entre elles est la plus probable. Néanmoins, pour des questions concernant le comportement global d'un réseau de quelques dizaines de gènes, la méthode de modélisation et de simulation qualitative semble un bon compromis entre validité biologique et applicabilité pratique.

La méthode a été complètement implémentée dans l'outil Genetic Network Analyzer (GNA). La disponibilité d'une implémentation est une condition nécessaire pour l'utilisation de la méthode lors de l'étude de réseaux complexes. Le choix de Java comme langage de programmation a permis une conception transparente et modulaire de GNA, et a facilité

le développement des interfaces graphiques. En outre, la portabilité de Java fait que GNA tourne sur toutes les plates-formes majeures utilisées, notamment Windows, Linux, Unix et Macintosh. Le programme est disponible gratuitement pour des utilisateurs académiques à des fins de recherches non commerciales. Aujourd'hui GNA a été distribué à plus de 100 équipes à travers le monde.

L'implémentation facilite l'application de la méthode, mais elle présuppose néanmoins que l'utilisateur comprend, au moins globalement, les bases mathématiques de la méthode. Aussi, l'interface graphique de GNA aide à spécifier les modèles et à interpréter les résultats, mais elle n'a pas été conçue pour dissimuler le fait que l'utilisateur s'est engagé dans un processus de modélisation mathématique : c'est-à-dire, un processus dans lequel il doit articuler un modèle mathématique du réseau à partir des données biologiques et déterminer la signification biologique des prédictions du simulateur. Ce choix de conception est délibéré, mais restreint l'utilisation de GNA aux personnes ayant des connaissances de la modélisation mathématique en biologie.

L'application de GNA à l'analyse des processus de régulation génique bactériens comme l'infection d'*E. coli* par le bactériophage λ , ou bien l'initiation de sporulation chez *B. subtilis*, a montré que la méthode est capable de reproduire le comportement qualitatif de réseaux de régulation génique complexes, mais déjà bien compris. Même si ces résultats sont confortants, les vrais apports de la modélisation et de simulation à la biologie devraient se faire ailleurs, dans l'analyse de systèmes dont le fonctionnement reste obscur. Cet objectif n'est que très rarement atteint dans les projets de modélisation aujourd'hui, à cause de l'absence des informations précises et quantitatives relevée ci-dessus. L'adaptation de notre méthode à ces conditions difficiles nourrit l'espoir qu'elle pourra réussir là où les méthodes traditionnelles ont du mal à s'imposer.

Même s'il est encore trop tôt pour évaluer le bien-fondé de cet espoir, les premiers résultats des projets de modélisation récemment lancés sont encourageants. Ils concernent justement des systèmes de régulation génique peu compris, comme la régulation globale de la transcription chez *E. coli* et *Synechocystis* PCC 6803. Ma priorité à court terme est d'avancer dans ces projets, par l'application coordonnée de la modélisation qualitative et des mesures de la cinétique. Ces activités s'inscrivent dans le cadre d'une collaboration – commencée en 2000 par un projet inter-EPST Bioinformatique et depuis poursuivie dans d'autres projets – avec les laboratoires de J. Geiselmann (Université Joseph Fourier, Grenoble) et de J. Houmard (École Normale Supérieure, Paris). Les activités ont pris d'ampleur avec l'arrivée de D. Ropers en tant que post-doctorante INRIA fin 2003. GNA est également utilisé par d'autres personnes, à l'extérieur de notre équipe, pour modéliser des réseaux de régulation génique [72].

6.2 Extensions de la méthode

Les fonctionnalités de GNA ne couvrent qu'une partie de la démarche de l'analyse de réseaux de régulation génique résumée dans le chapitre 1 et reproduite ici dans la figure 6.1. En fait, GNA est essentiellement un simulateur muni d'une interface graphique

aidant l'utilisateur à entrer les modèles et à interpréter les résultats. L'approche visualisée dans la figure est plus large et comprend d'autres activités, comme la construction et la révision des modèles à partir des données expérimentales, ainsi que des connaissances du domaine, la sélection des expériences discriminatoires et la comparaison des prédictions et des observations. Dans le cas où les réseaux à analyser sont complexes, la mise en oeuvre de la démarche de la figure demande des outils informatiques pour toutes les activités du cycle, des outils informatiques qui sont intégrés entre eux et avec les outils expérimentaux. L'évolution de GNA en un environnement informatique intégré pour l'analyse de réseaux de régulation génique constitue mon programme de recherche à moyen terme. Ici, je tracerai les contours de ce programme, en mettant en avant les premières explorations commencées ces dernières années.

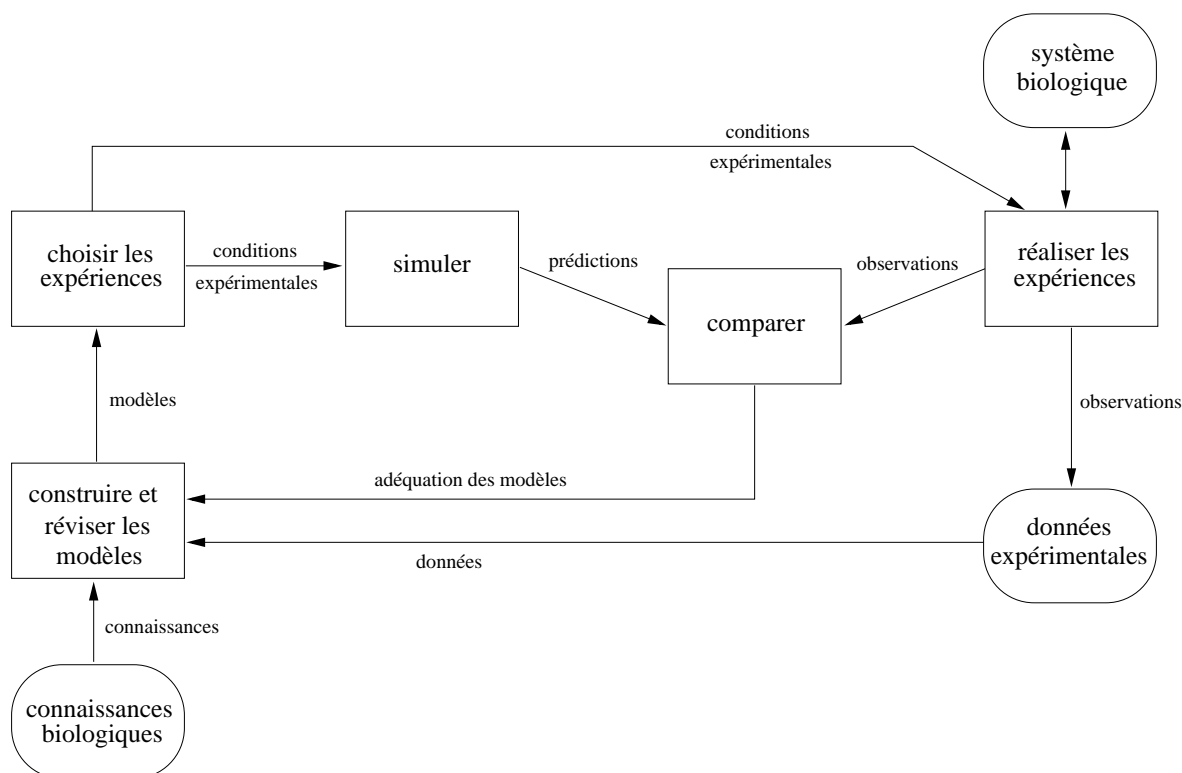


FIG. 6.1 – Démarche pour analyser la dynamique de réseaux de régulation génique : application coordonnée des outils expérimentaux, mathématiques et informatiques.

Une extension directe des travaux présentés dans ce mémoire serait de mieux caractériser les attracteurs (points d'équilibre et cycles limites) des modèles linéaires par morceaux. Outre des prédictions des comportements transitoires possibles du système (*i.e.*, les séquences d'états qualitatifs dans le graphe de transition), *l'identification des attracteurs* donnerait des informations de grande valeur pour la validation expérimentale d'un modèle. Dans le cadre d'une ARC INRIA (2002-2004) et d'une ACI IMPBio (2003-2006), j'étudie en collaboration avec R. Casey (INRIA Sophia-Antipolis), J.-L. Gouzé (INRIA Sophia-

Antipolis) et T. Sari (Université de Haute Alsace, Mulhouse), les attracteurs de modèles d'équations différentielles linéaires par morceaux, ainsi que la stabilité et le bassin d'attraction de ces attracteurs. Les résultats de ces études mathématiques seront utilisés pour le développement d'algorithmes efficaces, capables de trouver les attracteurs d'un modèle de réseau donné. Ces algorithmes seront ensuite implémentés dans un nouveau module de GNA et mis à l'épreuve dans les applications mentionnées à la fin de la section précédente.

La *validation* d'un modèle du réseau consiste à évaluer l'adéquation du modèle avec la réalité biologique. Dans le cas de la méthode de simulation qualitative, la validation passe par la comparaison du graphe de transition prédit avec les comportements qualitatifs observés. Plus précisément, il s'agit de vérifier que le graphe de transition généré par GNA possède certaines propriétés, notamment qu'il contient un chemin correspondant à une séquence d'états qualitatifs observés. Dans le cadre de la thèse de G. Batt (INRIA Rhône-Alpes et Université Joseph Fourier, Grenoble), nous développons une méthode de validation qui est basée sur l'utilisation des techniques de « model checking », développées pour la vérification automatique de systèmes de transition [6, 7]. Ce travail est soutenu financièrement par un projet inter-EPST Bioinformatique (2002-2004). Une exploration initiale de l'approche de « model checking », en partie effectuée en collaboration avec le projet VASY de l'INRIA Rhône-Alpes (D. Bergamini, H. Garavel et R. Mateescu) [4, 5], a donné des résultats tout-à-fait satisfaisants et est actuellement poursuivie de façon plus approfondie. À terme, ce travail devrait déboucher sur un nouveau module de GNA, capable de traiter des modèles complexes de façon efficace, tout en fournissant des garanties sur l'adéquation des résultats, en particulier la garantie qu'un modèle n'est jamais rejeté à tort.

La *construction* et la *révision* de modèles sont des activités clés dans le schéma de la figure 6.1. D'habitude, on suppose que la construction et la révision d'un modèle sont faites à la main, en utilisant des données de la littérature complétées par des hypothèses plausibles. Cependant, avec l'accumulation des connaissances sur les réseaux d'interactions moléculaires et la production en masse des données expérimentales, ces activités ont commencé à faire l'objet d'efforts particuliers de développement de méthodes systématiques, assistées par ordinateur. Globalement, il existe deux approches pour la construction de modèles [9, 59]. D'une part, ils peuvent être composés des connaissances sur les composants moléculaires d'un système de régulation et leurs interactions réciproques. D'autre part, ils peuvent être induits à partir des données d'expression et d'autres mesures de la cinétique du système. En pratique, les deux approches doivent être combinées afin de pouvoir exploiter efficacement la richesse d'informations disponibles.

J'ai commencé à explorer des idées sur l'intégration de la méthode de simulation qualitative avec des méthodes pour la construction et la révision de modèles. Dans l'équipe HELIX à l'INRIA-Rhône-Alpes, A. Morgat et collègues ont développé le système GenoExpertBacteria (GEB), pour l'analyse de données génomiques bactériennes, structuré autour des bases de connaissances sur les composantes de processus métaboliques et géniques et leurs interactions réciproques (<http://www-geb.inrialpes.fr/>). Nous utilisons GEB dans les projets de modélisation chez *E. coli* et *Synechocystis* pour rassembler et expertiser les données de la littérature, et nous étudions la possibilité de connecter GNA et

GEB plus étroitement. Un des moyens d'obtenir des hypothèses plausibles des interactions entre gènes et protéines est d'utiliser des méthodes d'analyse de séquences. M.-F. Sagot et collègues ont développé l'outil SMILE [40], qu'ils appliquent pour identifier des sites de fixation des régulateurs de transcription dans le génome d'*E. coli* et de *Synechocystis*. L'induction de modèles à partir des données d'expression fait l'objet d'une étude exploratoire avec G. Ferrari-Treccate (INRIA Rocquencourt), dans le cadre de l'ARC INRIA mentionné ci-dessus. Nous regardons si et comment la méthode pour l'identification de modèles d'équations différentielles linéaires par morceaux, développée par G. Ferrari-Treccate [23], pourrait être appliquée dans le contexte de GNA. Ces explorations seront poursuivies dans le cadre de la thèse de Samuel Drulhe (INRIA Rhône-Alpes et Université Joseph Fourier) et du projet européen Hygeia (2004-2007).

En règle générale, les données disponibles ne permettent pas d'arriver à un seul modèle du réseau étudié. L'étape de modélisation résulte alors en plusieurs modèles candidats, reflétant les hypothèses alternatives qui peuvent être faites. Afin de choisir parmi ces candidats, un retour à l'expérimentation est nécessaire, donnant pour résultat de nouvelles observations et mesures qui permettent de discriminer les modèles et donc les hypothèses sous-jacentes. Le problème central est évidemment de minimiser le nombre d'expériences discriminatoires, étant donné les coûts de leur exécution. I. Vatcheva, dont j'ai co-encadré (avec N. Mars) la thèse à l'Université de Twente (Pays-Bas), dans le cadre d'un projet PAI Van Gogh (1999-2001), a développé une méthode statistique pour la discrimination de modèles d'équations différentielles, applicable quand des données numériques précises sont absentes [73, 74]. Il semble possible d'appliquer cette méthode à l'analyse de réseaux de régulation génique, mais les travaux dans cette direction n'ont pas encore commencé.

L'intégration des outils informatiques pour toutes ces activités au sein d'un seul système, sous forme de nouveaux modules de GNA ou de connections entre GNA et d'autres logiciels, forme un programme de recherche ambitieux, dont la réalisation sera étalée sur plusieurs années. Le résultat pourrait ressembler aux systèmes d'aide à la découverte que j'ai envisagés lors de mon M.Sc. de philosophie des sciences et des technologies chez A. Rip à l'Université de Twente [22]. Je garde un petit espoir que j'aurai le temps de reprendre ces réflexions à l'interface de l'informatique et de la philosophie des sciences. Même si j'ai dû les négliger un peu ces dernières années, sous la pression d'avancer mon programme de recherche dans un domaine très compétitif, elles me sont restées chères.

6.3 Analyse qualitative de réseaux d'interactions : vers des cellules virtuelles

Comme cela a été mentionné dans le chapitre 1, le comportement observé d'un organisme vivant émerge des interactions entre ses composants moléculaires. Même si dans ce mémoire, je me suis focalisé sur les réseaux de régulation génique – composés surtout des interactions entre protéines et acides nucléiques, contrôlant la transcription et la traduction des gènes – il est clair que la compréhension du fonctionnement d'une cellule entière passe

par des modèles combinant la régulation génique avec le métabolisme, la transduction des signaux et d'autres processus. Plusieurs initiatives visant le développement des modèles de cellules entières, des *cellules virtuelles*, ont été lancées ces dernières années, comme E-CELL [68], Virtual Cell [39], l'« International *E. coli* Alliance » [31], l'« Alliance for Cellular Signalling » [55] et les travaux du « Systems Biology Research Group » à l'UCSD [8, 57, 58].

Toutefois, le bien-fondé des efforts de construire des modèles intégrés des différents processus cellulaires ne devrait pas nous amener à une vision naïve de la cellule virtuelle, une vision que F. Rechenmann et moi avons dénoncée dans un article écrit pour *La Recherche* [52]. En fait, il est parfois avancé que, étant donné les développements de la technologie informatique, le jour arrivera où des cellules entières – voire des organismes multi-cellulaires entiers – pourraient être simulés sur le niveau moléculaire, en suivant en parallèle des centaines de milliers de réactions. Cependant, il semble loin d'être acquis que cette stratégie de « brute force » – si elle devenait possible un jour – va nous aider à mieux comprendre le fonctionnement des systèmes vivants. En fait, il n'est pas du tout évident de trouver des patrons significatifs dans les masses de données générées par la simulation d'une cellule complète sur le niveau moléculaire.

À mon avis, il serait plus fructueux de ne pas développer l'idée de cellule virtuelle à partir des capacités actuelles ou prévues de la technologie informatique, mais de commencer par les questions biologiques. En particulier, le point de départ devrait être que différentes questions biologiques demandent différents types de modèles, sur différents niveaux d'abstraction. Par conséquent, une cellule virtuelle ne consisterait pas en un seul modèle détaillé de tous les processus cellulaires, mais plutôt en une hiérarchie de modèles couvrant différents aspects de la dynamique cellulaire sur différents niveaux d'abstraction. En passant d'un niveau à un autre, une analyse globale de la structure de régulation du système pourrait alors être poursuivie par une étude détaillée des mécanismes biochimiques dans un module de régulation particulier. Si les modèles approchés utilisés dans ce mémoire se prêtent bien au premier type de questions, le deuxième type demandera probablement des modèles plus fins, par exemple des modèles discrets des réactions biochimiques que l'on retrouve dans la simulation stochastique [3, 25].

Les approches qualitatives pour l'analyse de la dynamique cellulaire ont un rôle important à jouer dans la réalisation de cette vision de cellules virtuelles. Une première motivation est bien évidemment que les méthodes numériques classiques sont difficiles à appliquer aujourd'hui, étant donné l'absence générale des informations quantitatives sur les valeurs des paramètres cinétiques et des concentrations moléculaires. Cependant, même si les techniques expérimentales continuent à se développer à un point où il sera possible de mesurer – facilement et avec haute précision – les variables et les paramètres des processus cellulaires, les méthodes qualitatives ne perdront pas leur intérêt. En fait, s'il s'agit de bien comprendre la relation entre la structure et la dynamique d'un réseau d'interactions, elles donnent souvent des réponses plus appropriées que les méthodes quantitatives. Les méthodes qualitatives permettent, par exemple, de vérifier qu'une propriété dynamique d'un système n'est pas dépendante des valeurs précises des paramètres, mais émerge de la structure d'interactions du réseau sous-jacent.

Le développement de méthodes qualitatives pour étudier la dynamique d'une cellule sur différents niveaux d'abstraction, ainsi que l'intégration de ces méthodes dans un environnement pour l'analyse, voire la synthèse, de cellules virtuelles, pose de grands défis scientifiques à l'interface des mathématiques, de l'informatique et de la biologie. Il s'agit d'identifier des classes de modèles ayant des propriétés mathématiques favorables à l'analyse qualitative, de développer des algorithmes de simulation adaptés à ces contraintes, d'implémenter ces algorithmes dans des outils informatiques conviviaux et d'appliquer ces outils à des systèmes biologiques afin de comprendre comment le développement, le fonctionnement et l'évolution des organismes vivants émergent des interactions entre gènes, protéines, métabolites et autres molécules. Il n'en faudrait pas plus pour faire battre plus fort le cœur d'un scientifique...

Bibliography

- [1] R. Alur, C. Belta, F. Ivančić, V. Kumar, M. Mintz, G.J. Pappas, H. Rubin, and J. Schlug. Hybrid modeling and simulation of biomolecular networks. In M.D. Di Benedetto and A. Sangiovanni-Vincentelli, editors, *Hybrid Systems: Computation and Control (HSCC 2001)*, volume 2034 of *Lecture Notes in Computer Science*, pages 19–32. Springer-Verlag, Berlin, 2001.
- [2] R. Alur, T.A. Henzinger, G. Lafferriere, and G.J. Pappas. Discrete abstractions of hybrid systems. *Proceedings of the IEEE*, 88(7):971–984, 2000.
- [3] A. Arkin, J. Ross, and H.A. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected *Escherichia coli* cells. *Genetics*, 149(4):1633–1648, 1998.
- [4] G. Batt, D. Bergamini, H. de Jong, H. Gavarel, and R. Mateescu. Model checking genetic regulatory networks using GNA and CADP. In S. Graf and L. Mounier, editors, *Eleventh International SPIN Workshop on Model Checking of Software, SPIN 2004*, volume 2989 of *Lecture Notes in Computer Science*, pages 158–163, Berlin, 2004. Springer-Verlag.
- [5] G. Batt, H. de Jong, J. Geiselman, and M. Page. Qualitative analysis of genetic regulatory networks: A model-checking approach. In B. Bredeweg and P. Salles, editors, *Proceedings of the Fifteenth International Workshop on Qualitative Reasoning, QR-03*, pages 31–38, Brasilia, Brazil, 2003.
- [6] B. Bérard, M. Bidoit, A. Finkel, F. Laroussinie, A. Petit, L. Petrucci, Ph. Schnoebelen, and P. McKenzie. *Systems and Software Verification: Model-Checking Techniques and Tools*. Springer-Verlag, Berlin, 2001.
- [7] E.M. Clarke, O. Grumberg, and D.A. Peled. *Model Checking*. MIT Press, Boston, MA, 1999.
- [8] M.W. Covert, C.H. Schilling, and B. Palsson. Regulation of gene expression in flux balance models of metabolism. *Journal of Theoretical Biology*, 213(1):73–88, 2001.
- [9] H. de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9(1):69–105, 2002.

- [10] H. de Jong. Simulation qualitative. In L. Travé-Massuyès and P. Dague, editors, *Modèles et raisonnements qualitatifs*, pages 269–329. Hermès, Paris, 2003.
- [11] H. de Jong. Qualitative simulation and related approaches for the analysis of dynamical systems. *Knowledge Engineering Review*, 2004. Accepted for publication.
- [12] H. de Jong and J. Geiselmann. Modeling and simulation of genetic regulatory networks by ordinary differential equations. In J. Chen, E.R. Dougherty, I. Shmulevich, and Z.J. Wang, editors, *Genomic Signal Processing and Statistics*. EURASIP, 2004. To appear.
- [13] H. de Jong and J. Geiselmann. Modélisation et simulation de réseaux de régulation génique par des équations différentielles ordinaires. In J.-F. Boulicaut and O. Gandrillon, editors, *Informatique pour l'analyse du transcriptome*. Hermès, Paris, 2004. In press.
- [14] H. de Jong, J. Geiselmann, G. Batt, C. Hernandez, and M. Page. Qualitative simulation of the initiation of sporulation in *B. subtilis*. *Bulletin of Mathematical Biology*, 66(2):261–300, 2004.
- [15] H. de Jong, J. Geiselmann, C. Hernandez, and M. Page. Genetic Network Analyzer: Qualitative simulation of genetic regulatory networks. *Bioinformatics*, 19(3):336–344, 2003.
- [16] H. de Jong, J. Geiselmann, and D. Thieffry. Qualitative modelling and simulation of developmental regulatory networks. In S. Kumar and P.J. Bentley, editors, *On Growth, Form, and Computers*, pages 109–134. Academic Press, London, 2003.
- [17] H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Dealing with discontinuities in the qualitative simulation of genetic regulatory networks. In F. van Harmelen, editor, *Proceedings of Fifteenth European Conference on Artificial Intelligence, ECAI-02*, pages 412–416, Amsterdam, 2002. IOS Press.
- [18] H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Hybrid modeling and simulation of genetic regulatory networks: A qualitative approach. In A. Pnueli and O. Maler, editors, *Hybrid Systems: Computation and Control (HSCC 2003)*, volume 2623 of *Lecture Notes in Computer Science*, pages 267–282. Springer-Verlag, Berlin, 2003.
- [19] H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bulletin of Mathematical Biology*, 66(2):301–340, 2004.
- [20] H. de Jong and M. Page. Qualitative simulation of large and complex genetic regulatory systems. In W. Horn, editor, *Proceedings of the Fourteenth European Conference on Artificial Intelligence, ECAI-00*, pages 141–145, Amsterdam, 2000. IOS Press.

- [21] H. de Jong, M. Page, C. Hernandez, and J. Geiselman. Qualitative simulation of genetic regulatory networks: Method and application. In B. Nebel, editor, *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI-01*, pages 67–73, San Mateo, CA, 2001. Morgan Kaufmann.
- [22] H. de Jong and A. Rip. The computer revolution in science: Steps towards the realization of computer-supported discovery environments. *Artificial Intelligence*, 91(2):225–256, 1997.
- [23] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari. A clustering technique for the identification of piecewise affine systems. *Automatica*, 39(1):205–218, 2003.
- [24] R. Ghosh and C.J. Tomlin. Lateral inhibition through Delta-Notch signaling: A piecewise affine hybrid model. In M.D. Di Benedetto and A. Sangiovanni-Vincentelli, editors, *Hybrid Systems: Computation and Control (HSCC 2001)*, volume 2034 of *Lecture Notes in Computer Science*, pages 232–246. Springer-Verlag, Berlin, 2001.
- [25] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
- [26] L. Glass and S.A. Kauffman. The logical analysis of continuous non-linear biochemical control networks. *Journal of Theoretical Biology*, 39(1):103–129, 1973.
- [27] B.C. Goodwin. *Temporal Organization in Cells*. Academic Press, New York, N.Y., 1963.
- [28] J.-L. Gouzé and T. Sari. A class of piecewise linear differential equations arising in biological models. *Dynamical Systems*, 17(4):299–316, 2003.
- [29] J. Hasty, D. McMillen, F. Isaacs, and J.J. Collins. Computational studies of gene regulatory networks: *In numero* molecular biology. *Nature Review Genetics*, 2(4):268–279, 2001.
- [30] K.R. Heidtke and S. Schulze-Kremer. Design and implementation of a qualitative simulation model of λ phage infection. *Bioinformatics*, 14(1):81–91, 1998.
- [31] C. Holden. Alliance launched to model *E. coli*. *Science*, 297(5586):1459–1460, 2002.
- [32] T. Ideker, T. Galitski, and L. Hood. A new approach to decoding life: Systems biology. *Annual Review of Genomics and Human Genetics*, 2:343–372, 2001.
- [33] F. Jacob and J. Monod. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3:318–356, 1961.
- [34] P.D. Karp. Design methods for scientific hypothesis formation and their application to molecular biology. *Machine Learning*, 12(12):89–116, 1993.

- [35] S.A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22:437–467, 1969.
- [36] H. Kitano. Systems biology: A brief overview. *Science*, 295(5560):1662–1664, 2002.
- [37] B.J. Kuipers. *Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge*. MIT Press, Cambridge, MA, 1994.
- [38] D.J. Lockhart and E.A. Winzeler. Genomics, gene expression and DNA arrays. *Nature*, 405(6788):827–836, 2000.
- [39] L.M. Loew and J.C. Schaff. The Virtual Cell: A software environment for computational cell biology. *Trends in Biotechnology*, 19(10):401–405, 2001.
- [40] L. Marsan and M.-F. Sagot. Algorithms for extracting structured motifs using a suffix tree with an application to promoter and regulatory site consensus identification. *Journal of Computational Biology*, 7(3-4):345–62, 2000.
- [41] H.H. McAdams and A. Arkin. Simulation of prokaryotic genetic circuits. *Annual Review of Biophysics and Biomolecular Structure*, 27:199–224, 1998.
- [42] H.H. McAdams and L. Shapiro. Circuit simulation of genetic networks. *Science*, 269(5224):650–656, 1995.
- [43] L. Mendoza, D. Thieffry, and E.R. Alvarez-Buylla. Genetic control of flower morphogenesis in *Arabidopsis thaliana*: A logical analysis. *Bioinformatics*, 15(7-8):593–606, 1999.
- [44] T. Mestl, E. Plahte, and S.W. Omholt. A mathematical framework for describing and analysing gene regulatory networks. *Journal of Theoretical Biology*, 176(2):291–300, 1995.
- [45] J. Monod and F. Jacob. General conclusions: Teleonomic mechanisms in cellular metabolism, growth, and differentiation. In *Cold Spring Harbor Symposium on Quantitative Biology*, volume 26, pages 389–401, Cold Spring Harbor, NY, 1961.
- [46] B. Novák, A. Csikász-Nagy, B. Györfy, K.C. Chen, and J.J. Tyson. Mathematical model of the fission yeast cell cycle with checkpoint controls at the G1/S, G2/M and metaphase/anaphase transitions. *Biophysical Chemistry*, 72:185–200, 1998.
- [47] B. Novak and J.J. Tyson. Numerical analysis of a comprehensive model of M-phase control in *Xenopus* oocyte extracts and intact embryos. *Journal of Cell Science*, 106:1153–1168, 1993.
- [48] S.W. Omholt, X. Kefang, Ø. Andersen, and E. Plahte. Description and analysis of switchlike regulatory networks exemplified by a model of cellular iron homeostasis. *Journal of Theoretical Biology*, 195(3):339–350, 1998.

- [49] A. Pandey and M. Mann. Proteomics to study genes and genomes. *Nature*, 405(6788):837–846, 2000.
- [50] A.B. Pardee, F. Jacob, and J. Monod. The genetic control and cytoplasmic expression of ‘inducibility’ in the synthesis of β -galactosidase by *E. coli*. *Journal of Molecular Biology*, 1:165–178, 1959.
- [51] E.I. Prokudina, R.Y. Valeev, and R.N. Tchuraev. A new method for the analysis of the dynamics of the molecular genetic control systems. II. Application of the method of generalized threshold models in the investigation of concrete genetic systems. *Journal of Theoretical Biology*, 151(1):89–110, 1991.
- [52] F. Rechenmann and H. de Jong. La cellule virtuelle ? In preparation, 2004.
- [53] J. Reinitz, D. Kosman, C.E. Vanario-Alonso, and D.H. Sharp. Stripe forming architecture of the gap gene system. *Developmental Genetics*, 23:11–27, 1998.
- [54] J. Reinitz, E. Mjolsness, and D.H. Sharp. Model for cooperative control of positional information in *Drosophila* by bicoid and maternal hunchback. *Journal of Experimental Zoology*, 271:47–56, 1995.
- [55] N. Ruiz, C.N. Peterson, and T.J. Silhavy. RpoS-dependent transcriptional control of *sprE*: Regulatory feedback loop. *Journal of Bacteriology*, 183(20):5974–5981, 2001.
- [56] L. Sánchez and D. Thieffry. A logical analysis of the *Drosophila* gap genes. *Journal of Theoretical Biology*, 211(2):115–141, 2001.
- [57] C.H. Schilling, D. Letscher, and B.O. Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology*, 203:229–248, 2000.
- [58] C.H. Schilling and B.O. Palsson. Assessment of the metabolic capabilities of *Haemophilus influenzae* rd through a genome-scale pathway analysis. *Journal of Theoretical Biology*, 203:249–283, 2000.
- [59] C. Schut and B. Bredeweg. An overview of approaches to qualitative model construction. *Knowledge Engineering Review*, 11(1):1–25, 1996.
- [60] P. Smolen, D.A. Baxter, and J.H. Byrne. Modeling transcriptional control in gene networks: Methods, recent results, and future directions. *Bulletin of Mathematical Biology*, 62(2):247–292, 2000.
- [61] E.H. Snoussi. Qualitative dynamics of piecewise-linear differential equations: A discrete mapping approach. *Dynamics and Stability of Systems*, 4(3-4):189–207, 1989.
- [62] M. Sugita. Functional analysis of chemical systems *in vivo* using a logical circuit equivalent: II. The idea of a molecular automaton. *Journal of Theoretical Biology*, 4:179–192, 1963.

- [63] D. Thieffry and H. de Jong. Modélisation, analyse et simulation des réseaux génétiques. *Médecine/Sciences*, 18(4):492–502, 2002.
- [64] D. Thieffry and R. Thomas. Dynamical behaviour of biological networks: II. Immunity control in bacteriophage lambda. *Bulletin of Mathematical Biology*, 57(2):277–297, 1995.
- [65] R. Thomas. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, 42:563–585, 1973.
- [66] R. Thomas and R. d’Ari. *Biological Feedback*. CRC Press, Boca Raton, FL, 1990.
- [67] R. Thomas, D. Thieffry, and M. Kaufman. Dynamical behaviour of biological regulatory networks: I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bulletin of Mathematical Biology*, 57(2):247–276, 1995.
- [68] M. Tomita, K. Hashimoto, K. Takahashi, T. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J.C. Venter, and C.A. Hutchison. E-CELL: Software environment for whole-cell simulation. *Bioinformatics*, 15(1):72–84, 1999.
- [69] R.B. Trelease, R.A. Henderson, and J.B. Park. A qualitative process system for modeling NF- κ B and AP-1 gene regulation in immune cell biology research. *Artificial Intelligence in Medicine*, 17(3):303–321, 1999.
- [70] A.M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London, Series B*, 237:37–72, 1951.
- [71] J.J. Tyson, K. Chen, and B. Novak. Network dynamics and cell physiology. *Nature Reviews Molecular Cell Biology*, 2(12):908–916, 2001.
- [72] A. Usseglio Viretta and M. Fussenegger. Modeling the quorum sensing regulatory network of human-pathogenic *Pseudomonas aeruginosa*. *Biotechnology Progress*, 2004. In press.
- [73] I. Vatcheva, O. Bernard, H. de Jong, J.-L. Gouzé, and N.J.I. Mars. Discrimination of semi-quantitative models by experiment selection: Method and application in population biology. In B. Nebel, editor, *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI-01*, pages 74–79, San Mateo, CA, 2001. Morgan Kaufmann.
- [74] I. Vatcheva, H. de Jong, and N.J.I. Mars. Selection of perturbation experiments for model discrimination. In W. Horn, editor, *Proceedings of Fourteenth European Conference on Artificial Intelligence, ECAI-00*, pages 191–195, Amsterdam, 2000. IOS Press.
- [75] G. von Dassow, E. Meir, E.M. Munro, and G.M. Odell. The segment polarity network is a robust developmental module. *Nature*, 406(6792):188–192, 2000.

- [76] J.D. Watson and F.H. Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953.

Annexe A

Curriculum vitae

État civil

Nom : De Jong
Prénom : Hidde (H.)
Sexe : masculin
Date et lieu de naissance : le 7 juillet 1968, Delft (Pays-Bas)
Nationalité : néerlandaise
Adresse : 12 bis, rue des Clercs, 38000 Grenoble

Formation

1986 : Baccalauréat au Carmel Lyceum, Oldenzaal (Pays-Bas).

1992 : Diplôme de M.Sc. en Informatique (*cum laude*) à l'Université de Twente, Enschede (Pays-Bas).

Sujet de spécialisation : intelligence artificielle.

1993 : Diplôme de M.Sc. en Gestion et Technologie (*cum laude*) à l'Université de Twente.

Sujet de spécialisation : theorie des organisations et technologies d'information.

1994 : Diplôme de M.Sc. en Philosophie des Sciences, des Technologies et de la Société (*cum laude*) à l' Université de Twente.

Sujet de spécialisation : philosophie des sciences et des technologies.

1998 : Doctorat en Informatique (*cum laude*) à l'Université de Twente.

Sujet : développement et application des techniques d'intelligence artificielle pour un système d'aide à l'analyse des expériences scientifiques.

Titre de la thèse : « Computer-supported analysis of scientific measurements ».

Recherche

- 1991 : Stagiaire au Research Institute for Knowledge Systems (RIKS), Maastricht (Pays-Bas).
- 1993-1998 : Doctorant en Informatique à l'Université de Twente.
- 1998- : Chargé de recherche dans l'équipe SHERPA, puis HELIX à l'Institut National de Recherche en Informatique et en Automatique (INRIA), Unité de Recherche Rhône-Alpes, Montbonnot Saint-Martin.
Sujet de recherche : modélisation et simulation de réseaux de régulation génique.

Encadrement

- 1993-1998 : Encadrement d'étudiants en M.Sc. à l'Université de Twente.
- 1999-2001 : Encadrement d'Ivayla Vatcheva, doctorante en Informatique à l'Université de Twente.
- 2002- : Encadrement de Grégory Batt, doctorant en Informatique à l'Université Joseph Fourier, Grenoble.
- 2004- : Encadrement de Samuel Drulhe, doctorant en Informatique à l'Université Joseph Fourier, Grenoble.
- 2000- : Encadrement d'étudiants en stage de DESS, en stage de magistère et en stage de DEA à l'INRIA Rhône-Alpes.

Enseignement

- 1993-1998 : Cours d'Informatique à l'Université de Twente.
Sujets : conception de systèmes d'information, applications industrielles de techniques d'intelligence artificielle.
- 2000-2003 : Cours de Bioinformatique à l'Université Joseph Fourier.
Sujet : modélisation et simulation de réseaux de régulation génique.
- 2002- : Cours de Bioinformatique à l'Institut National des Sciences Appliquées (INSA), Lyon.
Sujet : modélisation et simulation de réseaux de régulation génique.
- 2002 : Cours à l'atelier de formation « From functional genomic data to the dynamical analysis of molecular networks » (INSERM), La Londe-Les-Maures, France.
- 2003 : Cours à la troisième école thématique de Biologie Végétale (CNRS-INRA), Roscoff, France.

- 2004 : Cours à l'école d'été « Mathematical and Computational Methods in Biology » (CIMPA-UNESCO), Valdivia, Chile.
- 2004 : Cours à l'école de printemps « BioData Mining » (EUNITE/JCB), Jena, Allemagne.
- 2004 : Cours à l'école « Coupled Map Lattices », Institut H. Poincare, Paris, France.
- 2004 : Cours à l'école d'été « Multiple Aspects of DNA and RNA : from biophysics to bioinformatics » (Marie Curie Training Course), Les Houches, France.
- 2004 : Cours à l'école « European School on Nanosciences & Nanotechnologies », Grenoble, France.
- 2004 : Tutoriel à l'« International Conference on Systems Biology », Heidelberg, Allemagne.

Projets de recherche

- 1999-2001 : Co-responsable du projet « Analyse de données scientifiques : une approche par modèles », Programme d'Actions Intégrées franco-hollandais Van Gogh.
- 2000-2002 : Co-responsable du projet « Modélisation et simulation de réseaux de régulation génique : transduction des signaux chez la cyanobactérie *Synechocystis* PCC 6803 », Programme Bioinformatique inter-EPST.
- 2002- : Co-responsable du projet « Validation de modèles de réseaux de régulation génique : régulation globale de la transcription chez *Escherichia coli* et *Synechocystis* PCC 6803 », Programme Bioinformatique inter-EPST.
- 2002- : Co-responsable du projet « GDyn : analyse dynamique de réseaux de régulation génique », Action de Recherche Coopérative INRIA.
- 2003- : Responsable du projet « BacAttract : analyse théorique et expérimentale des attracteurs de réseaux de régulation génique : régulation globale de la transcription chez *Escherichia coli* et *Synechocystis* PCC 6803 », Action Concertée Incitative IMPBio.
- 2004- : Responsable d'un workpackage dans le projet européen « Hygeia : Hybrid systems for biochemical network modeling and analysis », NEST-2003-1 ADVENTURE.

Animation

- 2004- : Rédacteur adjoint de l'« IEEE Transactions on Computational Biology and Bioinformatics ».
- 2002- : Membre du comité de programme de l'« International Workshop on Qualitative Reasoning ».

- 2004- : Membre du comité de programme de l'« IEEE Workshop on Genomic Signal Processing and Statistics ».
- 2004- : Membre du comité de programme des « Journées Ouvertes de Biologie, Informatique et Mathématiques ».
- 2004- : Membre du comité de programme de la « Conference on Algorithms and Computational Methods for Biochemical and Evolutionary Networks ».
- 2003 : Co-organisateur de la réunion satellite ECCB « Modeling and simulation of biological regulatory processes », ENS, Paris.
- 2004 : Co-organisateur de la réunion satellite JOBIM « Modélisation et la simulation de processus de régulation biologique », Montréal, Canada.
- 2004- : Membre du comité scientifique du groupe de travail VICANNE « Modélisation dynamique de réseaux de régulation biologiques », financé dans le cadre de l'Action Concertée Incitative IMPBio.
- 2004 : Co-organisateur de la réunion « Réseaux géniques et métaboliques : vers des modèles intégrés » du groupe de travail « Modélisation dynamique de réseaux de régulation biologiques », Grenoble.
- 2001- : Membre de la commission de spécialistes (section 27) de l'Université de Provence, Marseille.
- 2004- : Membre de la section d'audition du concours CR2 de l'INRIA Rocquencourt.

Séminaires invités

- 1997 : Filosofisch Instituut, Rijksuniversiteit Groningen, Groningue, Pays-Bas.
- 1999 : INRIA Sophia-Antipolis, Sophia-Antipolis, France.
- 2000 : Centre for Biometrics, CPRO- DLO, Wageningen University & Research Centre, Pays-Bas.
- 2000 : Séminaire « Algorithmique et biologie », Institut Pasteur, Paris, France.
- 2000 : INRIA Sophia-Antipolis, Sophia-Antipolis, France.
- 2001 : Laboratoire Génétique et Cancer (CNRS UMR5641), Lyon, France.
- 2001 : Atelier CONSENSUS, Génopole d'Evry, Evry, France.
- 2001 : École Supérieure d'Ingénieurs de Luminy (ESIL), Université de la Méditerranée, Marseille, France.

- 2001 : Séminaire « Algorithmique et Biologie », Université Claude Bernard, Lyon, France.
- 2001 : Réunion du groupe de travail « Bioinformatique fonctionnelle des systèmes de régulations génétiques » de l'action « Informatique, Mathématique, Physique pour la Génomique (IMPG) », Marseille, France.
- 2002 : Institut de Biologie de Lille, Lille, France.
- 2002 : Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA), Nancy, France.
- 2002 : Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Allemagne.
- 2002 : LION Bioscience, Heidelberg, Allemagne.
- 2003 : Séminaire « Dynamical Stochastic Modeling in Biology », University of Copenhagen, Copenhague, Danmark.
- 2003 : Séminaire MAGMA, Institut de Mathématiques de Luminy, Marseille, France.
- 2003 : Vlaams Interuniversitair Instituut voor Biotechnologie, University of Ghent, Belgique.
- 2003 : Laboratoire Leibniz, Institut d'Informatique et de Mathématiques Appliquées de Grenoble (IMAG), Grenoble, France.
- 2003 : Laboratoire des Interactions Plantes-Microorganismes, INRA/CNRS, Toulouse, France.
- 2003 : Séminaire thématique de Bioinformatique de la Génopole Ouest, IRISA, Rennes, France.
- 2003 : AC Graphes et Algorithmique pour la Bioinformatique, LIRMM, Montpellier, France.
- 2003 : Journées thématiques en Bioinformatique, Bordeaux, France.
- 2004 : Centrum voor Wiskunde en Informatica (CWI), Amsterdam, Pays-Bas.
- 2004 : Session « Biological Networks », 6th World Congress of the Bernoulli Society, Barcelone, Espagne.
- 2004 : Second Bertinoro Computational Biology Meeting, Bertinoro, Italie.
- 2004 : Séminaire « Integrative Bioinformatics – Aspects of the Virtual Cell », Dagstuhl, Allemagne.
- 2004 : Rencontre du groupe de travail « Analyse dynamique de réseaux de régulation biologique », Dijon, France.

- 2004 : Laboratoire TIMC, Institut d'Informatique et de Mathématiques Appliquées de Grenoble (IMAG), Grenoble, France.
- 2004 : Department of Physiology, McGill University, Montréal, Canada.
- 2004 : Averøy symposium en honneur de René Thomas, CIGENE (Centre for Integrative Genetics), and Department of Chemistry, Biotechnology and Food Science, Agricultural University of Norway, Norvège.
- 2004 : Forum Franco-indien « Environmental and Therapeutical Applications of Nanobiotechnology and Genomics », Sophia-Antipolis, France.

Annexe B

Publications depuis 1998

Articles dans des journaux

H. de Jong, F. van Raalte (1999), « Comparative envisionment construction: A technique for the comparative analysis of dynamical systems », *Artificial Intelligence*, 115(2), 145-214.

H. de Jong, N.J.I. Mars, P.E. van der Vet (1999), « Computer-supported resolution of measurement conflicts: A case-study in materials science », *Foundations of Science*, 4, 427-461.

H. de Jong, M. Page, C. Hernandez, J. Geiselman, S. Maza (2000), « Modeling and simulation of genetic regulatory networks », *ERCIM News*, 43, 18-19.

H. de Jong (2002), « Modeling and simulation of genetic regulatory systems: A literature review », *Journal of Computational Biology*, 9(1), 69-105.

D. Thieffry, H. de Jong (2002), « Modélisation, analyse et simulation des réseaux génétiques », *Médecine/Sciences*, 18(4), 492-502.

H. de Jong, J. Geiselman, C. Hernandez, M. Page (2003), « Genetic Network Analyzer: Qualitative simulation of genetic regulatory networks », *Bioinformatics*, 19(3), 336-344.

I. Vatcheva, O. Bernard, H. de Jong, N.J.I. Mars (2003), « Experiment selection for the discrimination of semi-quantitative models of dynamical systems », soumis pour publication.

H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, J. Geiselman (2004), « Qualitative simulation of genetic regulatory networks using piecewise-linear models », *Bulletin of Mathematical Biology*, 66(2):301-340.

H. de Jong, J. Geiselman, G. Batt, C. Hernandez, M. Page (2004), « Qualitative simulation of the initiation of sporulation in *B. subtilis* », *Bulletin of Mathematical Biology*, 66(2):361-300.

D. Eveillard, D. Ropers, H. de Jong, C. Branlant, A. Bockmayr (2004), « A multi-site constraint programming model of alternative splicing regulation », *Theoretical Computer Science*, sous presse.

H. de Jong (2004), « Qualitative simulation and related approaches for the analysis of dynamical systems », *Knowledge Engineering Reviewe*, à paraître.

F. Rechenmann, H. de Jong (2004), « La cellule virtuelle ? », *La Recherche*, à paraître.

Chapitres dans des livres

H. de Jong, J. Geiselman, D. Thieffry (2003), « Qualitative modeling and simulation of developmental regulatory networks », S. Kumar, P.J. Bentley (dir.), *On Growth, Form, and Computers*, Academic Press, London, 109-134.

H. de Jong (2003), « Simulation qualitative », L. Travé-Massuyès, P. Dague (dir.), *Le raisonnement qualitatif pour les sciences de l'ingénieur*, Hermès, Paris, 269-329.

D. Schneider, A. Latifi, H. de Jong, J. Geiselman (2003), « Dynamics and simulation of the genetic regulatory networks in bacteria », *Recent Research Developments in Genetics*, 3:55-83.

H. de Jong, J. Geiselman (2004), « Modeling and simulation of genetic regulatory networks by ordinary differential equations », J. Chen, E.R. Dougherty, I. Shmulevich, Z.J. Wang (dir.), *Genomic Signal Processing and Statistics*, EURASIP, à paraître.

H. de Jong, J. Geiselman (2004), « Modélisation et simulation de réseaux de régulation génique par des équations différentielles ordinaires », J.-F. Boulicaut, O. Gandrillon (dir.), *Informatique pour l'analyse du transcriptome*, Hermès, Paris, sous presse.

Publications dans des actes

H. de Jong (1998), « Computer-supported model construction and its application », *Working Notes of the International Conference on Model-Based Reasoning in Scientific Discovery, MBR'98*, Pavía, 19-20.

I. Vatcheva, H. de Jong (1999), « Semi-quantitative comparative analysis », T.D. Dean (ed.), *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, IJCAI-99*, Morgan Kaufmann, San Francisco, CA, 1034-1040.

I. Vatcheva, H. de Jong (1999), « Semi-quantitative comparative analysis and its application », C. Price (ed.), *Working Notes of the Thirteenth International Workshop on Qualitative Reasoning, QR-99*, AAAI Press, Menlo Park, CA, 231-239.

H. de Jong, M. Page (2000), « Qualitative simulation of large and complex genetic regulatory systems », J. Flores, K. Forbus (dir.), *Working Notes of the 14th International Workshop on Qualitative Reasoning, QR 2000*, Morelia, Mexico, 32-39.

H. de Jong, M. Page (2000), « Qualitative simulation of large and complex genetic regulatory systems », W. Horn (ed.), *Proceedings of the 14th European Conference on Artificial Intelligence, ECAI 2000*, IOS Press, Amsterdam, 141-145.

I. Vatcheva, H. de Jong, N.J.I. Mars (2000), « Selection of perturbation experiments for model discrimination », J. Flores, K. Forbus (dir.), *Working Notes of the 14th International Workshop on Qualitative Reasoning, QR 2000*, Morelia, Mexico, 24-31.

I. Vatcheva, H. de Jong, N.J.I. Mars (2000), « Selection of perturbation experiments for model discrimination », W. Horn (ed.), *Proceedings of the 14th European Conference on Artificial Intelligence, ECAI 2000*, IOS Press, Amsterdam, 191-195.

I. Vatcheva, H. de Jong, N.J.I. Mars (2000), « Selection of perturbation experiments for model discrimination », A. van den Bosch, H. Weigand (dir.), *Proceedings of the Twelfth Belgium-Netherlands Artificial Intelligence Conference, BNAIC 2000*, De Efteling, Kaatsheuvel, 353-354.

H. de Jong, M. Page, C. Hernandez, J. Geiselmann (2001), « Qualitative simulation of genetic regulatory networks: Method and application », B. Nebel (ed.), *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI-01*, Morgan Kaufmann, San Francisco, CA, 67-73.

H. de Jong, M. Page, C. Hernandez, J. Geiselmann (2001), « Qualitative simulation of genetic regulatory networks: Method and application », G. Biswas, T. Bylander (dir.), *Working Notes of the 15th International Workshop on Qualitative Reasoning, QR-01*, San Antonio, TX, 134-141.

H. de Jong, J. Geiselmann, C. Hernandez, M. Page (2001), « Qualitative simulation of the initiation of sporulation in *B. subtilis* », *Recueil des Actes des Journées Ouvertes Biologie Informatique Mathématiques, JOBIM 2001*, Toulouse, 187-194.

I. Vatcheva, O. Bernard, H. de Jong, J.-L. Gouzé, N.J.I. Mars (2001), « Discrimination of semi-quantitative models by experiment selection: Method and application in population biology », B. Nebel (ed.), *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI-01*, Morgan Kaufmann, San Francisco, CA, 74-79.

I. Vatcheva, O. Bernard, H. de Jong, J.-L. Gouzé, N.J.I. Mars (2001), « Discrimination of

semi-quantitative models by experiment selection: Method and application in population biology », G. Biswas, T. Bylander (dir.), *Working Notes of the 15th International Workshop on Qualitative Reasoning, QR-01*, San Antonio, TX, 121-127.

H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, J. Geiselman (2002), « Dealing with discontinuities in the qualitative simulation of genetic regulatory networks », *Proceedings of Fifteenth European Conference on Artificial Intelligence, ECAI-02*, F. van Harmelen (ed.), IOS Press, Amsterdam, 412-416.

H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, J. Geiselman (2002), « Dealing with discontinuities in the qualitative simulation of genetic regulatory networks », N. Agell, J.A. Ortega (dir.), *Working Notes of Sixteenth International Workshop on Qualitative Reasoning, QR 2002*, Sitges, Barcelona, Spain, 67-74.

C. Hernandez, H. de Jong, J. Geiselman, M. Page (2002), « VisualGNA: A graphical interface for the qualitative simulation of genetic regulatory networks », J. Nicolas, C. Thermes (dir.), *Recueil des Actes des Journées Ouvertes Biologie Informatique Mathématiques, JOBIM 2002*, Saint-Malo, 335-336.

G. Batt, H. de Jong, J. Geiselman, M. Page (2003), « Qualitative analysis of genetic regulatory networks: A model-checking approach », *IJCAI Workshop on Model Checking and Artificial Intelligence (MoChArt-03)*, Acapulco, Mexico, 51-58.

G. Batt, H. de Jong, J. Geiselman, M. Page (2003), « Qualitative analysis of genetic regulatory networks: A model-checking approach », B. Bredeweg, P. Salles (dir.), *Working Notes of Seventeenth International Workshop on Qualitative Reasoning, QR 2003*, Brasília, Brazil, 31-38.

H. de Jong (2003), « Modeling and simulation of genetic regulatory networks », L. Benvenuti, A. De Santis, L. Farina (dir.), *First Multidisciplinary International Symposium on Positive Systems: Theory and Applications, POSTA 2003*, Lectures Notes in Control and Information Sciences 294, Springer-Verlag, Berlin, 111-118.

H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, J. Geiselman (2003), « Hybrid modeling and simulation of genetic regulatory networks: A qualitative approach », A. Pnueli, O. Maler (dir.), *Hybrid Systems: Computation and Control (HSCC 2003)*, Lecture Notes in Computer Science 2623, Springer-Verlag, Berlin, 267-282.

H. de Jong, J. Geiselman, C. Hernandez, M. Page (2003), « Qualitative simulation of the initiation of sporulation in *Bacillus subtilis* », V. Capasso (ed.), *Mathematical Modeling, Computing in Biology and Medicine. Proceedings of the 5th ESMTB Conference*, Milan Research Centre for Industrial and Applied Mathematics (MIRIAM), Milan, 23-28.

D. Eveillard, D. Ropers, H. de Jong, C. Branlant, A. Bockmayr (2003), « Multiscale modeling

of alternative splicing regulation », C.Priami (ed.), *Computational Methods in Systems Biology, CMSB-03*, Lecture Notes in Computer Science, 2602, Springer-Verlag, Berlin, 75-87.

G. Batt, D. Bergamini, H. de Jong, H. Gavarel, R. Maateescu (2004), « Model checking genetic regulatory networks using GNA and CADP », *Eleventh International SPIN Workshop on Model Checking of Software, SPIN 2004*, Lecture Notes in Computer Science 2989, Springer-Verlag, Berlin, 158-163.

Rapports de recherche

I. Vatcheva, H. de Jong [1999], « Semi-quantitative comparative analysis », CTIT Technical Report 99-10, University of Twente, Enschede, the Netherlands.

H. de Jong (2000), « Modeling and simulation of genetic regulatory systems: A literature review », RR-4032, INRIA, France.

H. de Jong, J. Geiselman, C. Hernandez, M. Page (2001), « Genetic Network Analyzer: A tool for the qualitative simulation of genetic regulatory networks », RR-4262, INRIA, France.

H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, J. Geiselman (2001), « Qualitative simulation of genetic regulatory networks using piecewise-linear models », RR-4407, INRIA, France.

H. de Jong, J. Geiselman, G. Batt, C. Hernandez, M. Page (2002), « Qualitative simulation of the initiation of sporulation in *B. subtilis*, RR-4527, INRIA, France.

D. Eveillard, D. Ropers, H. de Jong, C. Branlant, A. Bockmayr (2003), « A multi-site constraint programming model of alternative splicing regulation », RR-4830, INRIA, France.

I. Vatcheva, O. Bernard, H. de Jong, N.J.I. Mars (2003), « Experiment selection for the discrimination of semi-quantitative models of dynamical systems », RR-4940, INRIA, France.

H. de Jong (2004), « Qualitative simulation and related approaches for the analysis of dynamical systems », RR-5218, INRIA, France.

