GDyn : Analyse dynamique de réseaux de régulation génique

Proposition d'une Action de Recherche Coopérative (ARC)

Coordinateurs : Jean-Luc Gouzé et Hidde de Jong

5 février 2003

Résumé

L'étude de réseaux de régulation génique est fortement stimulée par le développement des nouvelles techniques génomiques permettant de mesurer simultanément le niveau d'expression de tous les gènes d'un organisme. Outre des méthodes expérimentales à haut débit, des approches mathématiques et bioinformatiques sont indispensables pour analyser la dynamique des réseaux de régulation génique. Un formalisme basé sur des équations différentielles linéaires par morceaux (LPM) s'est montré particulièrement adapté pour la modélisation de ces réseaux.

Le premier but de cette ARC est d'étudier le comportement dynamique des systèmes LPM (solutions, points d'équilibre, cycles limites, stabilité, ...). Ensuite, dans un deuxième volet, nous tenterons d'appliquer des concepts issus de l'automatique et de la théorie de systèmes hybrides à cette classe de systèmes pour mieux décrire, simplifier et identifier un réseau. Les résultats de ces études mathématiques seront utilisés afin de développer de nouvelles méthodes informatiques pour l'analyse de réseaux de régulation génique. Ces méthodes seront implémentées et mises à l'épreuve dans l'étude des réseaux impliqués dans la régulation globale de la transcription chez les bactéries *Escherichia coli* et *Synechocystis*. Plus généralement, nous cherchons à rassembler des compétences multiples et à créer des collaborations enrichissantes autour de la thématique de la dynamique de réseaux de régulation génique, de sa description et son analyse, sur le plan mathématique, informatique et biologique.

Équipes partenaires

Nom	Responsable	Laboratoire
CEG	Johannes Geiselmann	Laboratoire Plasticité et Expression des Génomes Microbiens, université Joseph Fourier, Grenoble
COMORE	Jean-Luc Gouzé	INRIA Sophia Antipolis
HELIX	François Rechenmann	INRIA Rhône-Alpes
LMA	Michel Goze	Laboratoire de Mathématiques et Applications, université de Haute Alsace, Mulhouse
SOSSO	Michel Sorine	INRIA Rocquencourt
TOP	Jean Houmard	Laboratoire Organismes Photosynthétiques et Environnement, École Normale Supérieure, Paris

Chercheurs participants

Nom	Grade	Équipe	Courriel
Olivier Bernard	CR	COMORE	`Olivier.Bernard@sophia.inria.fr`
Frédérique Clément	CR	SOSSO	`Frederique.Clement@inria.fr`
Giancarlo Ferrari-Trecate	CR	SOSSO	`Giancarlo.Ferrari-Trecate@inria.fr`
Johannes Geiselmann	Pr	CEG	`Hans.Geiselmann@ujf-grenoble.fr`
Jean-Luc Gouzé	DR	COMORE	`Gouze@sophia.inria.fr`
Chantal Guidi-Rontani	CR	TOP	`rontani@biologie.ens.fr`
Jean Houmard	DR	TOP	`jhoumard@biologie.ens.fr`
Hidde de Jong	CR	HELIX	`Hidde.de-Jong@inrialpes.fr`
Michel Page	MdC	HELIX	`Michel.Page@inrialpes.fr`
Tewfik Sari	Pr	LMA	`T.Sari@uha.fr`
Dominique Schneider	MdC	CEG	`Dominique.Schneider@ujf-grenoble.fr`
Michel Sorine	DR	SOSSO	`Michel.Sorine@inria.fr`

1 Introduction

L'étude de *réseaux de régulation génique* est fortement stimulée par le développement des nouvelles techniques génomiques permettant, entre autre, de mesurer simultanément le niveau d'expression de tous les gènes d'un organisme. Outre ces méthodes expérimentales à haut débit, des approches mathématiques et bioinformatiques sont indispensables pour analyser les réseaux géniques. À cause du grand nombre de composantes des réseaux, connectées par des boucles de retroaction, une compréhension intuitive de la dynamique du système est souvent difficile, voire impossible. La modélisation et la simulation par ordinateur peuvent contribuer à l'élucidation du fonctionnement d'un réseau d'interactions en permettant au biologiste de tester un nombre plus restreint d'hypothèses plausibles.

Une grande variété d'approches pour la *modélisation* et la *simulation* de réseaux de régulation génique a été proposée dans les trente dernières années [4, 14, 23]. 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, comme celui presenté pour l'infection d'une bactérie par le phage λ [18], 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 numerique sont difficiles à appliquer.

Il n'est en général pas possible de prendre en compte de telles contraintes avec les méthodes de modélisation et de simulation existantes. Toutefois les approches utilisant une classe de modèles définis par des équations différentielles *linéaires par morceaux (LPM)*, proposés par Glass et Kauffman [11], sont une exception notable. Les modèles LPM permettent une description approchée des réseaux de régulation génique, particulièrement 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 [11, 21, 24]. Sur le plan formel, les modèles LPM sont proches des modèles logiques asynchrones proposés par Thomas *et al.* [26]. Les modèles LPM et logiques ont déjà donné des résultats intéressants lors de l'étude de nombreux réseaux de régulation procaryotes et eucaryotes.

La nature discontinue des fonctions en escalier utilisées dans les modèles LPM pose des problèmes mathématiques non trivaux. Nous avons montré récemment qu'une approche capable de prendre en compte des équations différentielles discontinues, largement utilisée en théorie du contrôle, pouvait être utilisée afin de résoudre ces problèmes de façon à la fois mathématiquement rigoureuse et facilement utilisable [13]. Cette approche, proposée initialement par Fillipov [9], est basée sur une généralisation des équations différentielles aux inclusions différentielles. À partir de ces résultats, nous avons développé une *méthode de simulation qualitative* de réseaux de régulation génique. La méthode a été implémentée en Java dans l'outil GNA (*Genetic Network Analyser*) [6].

Nous proposons dans cette ARC d'approfondir et d'étendre ces travaux. Dans un premier volet, nous poursuivrons l'application de l'approche de Filippov et nous caractériserons les attracteurs (points d'équilibre et cycles limites) des systèmes LPM, ainsi que leur stabilité et leur bassin d'attraction. Ensuite, dans un deuxième volet, nous tenterons d'appliquer des concepts issus de l'automatique et de la théorie de systèmes hybrides à cette classe de systèmes pour mieux décrire, simplifier et identifier un réseau. Les résultats de ces études mathématiques seront utilisés afin de développer de nouvelles méthodes informatiques pour l'analyse de réseaux de régulation génique. Ces méthodes seront implémentées et mises à l'épreuve dans l'étude des réseaux responsables pour la *régulation globale de la transcription* chez les bactéries *Escherichia coli* et *Synechocystis*.

Plusieurs participants à l'ARC proposée ici sont également impliqués dans un projet soumis pour financement dans le cadre du Programme Bioinformatique inter-EPST. Dans ce projet, intitulé « Validation de modèles de réseaux de régulation génique », nous développerons une méthode capable de tester la validité d'un modèle en comparant les prédictions obtenues par simulation qualitative avec des données expérimentales. La méthode sera utilisée pour la validation des modèles LPM des réseaux intervenant dans la régulation globale de la transcription chez *E. coli* et *Synechocystis*. L'ARC est complémentaire au projet inter-EPST, dans le sens où nous nous focaliserons non pas sur la modélisation de systèmes biologiques et leur validation, mais sur l'étude des propriétes dynamiques des modèles LPM.

Dans la deuxième section de la proposition nous introduisons de façon intuitive la modélisation de réseaux de régulation génique par les modèles LPM, ainsi que l'analyse mathématique des propriétes dynamiques de ces réseaux. Ensuite, nous décrivons les deux volets de l'ARC plus en détail : l'analyse des attracteurs d'un système de régulation génique (section 3) et l'application des concepts issus de l'automatique et des systèmes hybrides (section 4). Le calendrier du projet est presenté dans la quatrième section, suivis par un récapitulatif des principales contributions dans la section finale.

2 Modélisation et analyse de réseaux de régulation génique

La dynamique des réseaux de régulation génique peut être modélisée par une classe d'équations différentielles LPM de la forme:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{x})\,\boldsymbol{x},\ \boldsymbol{x} \geq \boldsymbol{0}, \tag{1}$$

où $\boldsymbol{x} = (x_1, \ldots, x_n)'$ est un vecteur de concentrations de protéines, $\boldsymbol{f} = (f_1, \ldots, f_n)'$ et $\boldsymbol{g} = \mathrm{diag}(g_1, \ldots, g_n)$. Le taux de variation de chaque concentration x_i est défini comme la différence entre le taux de synthèse $f_i(\boldsymbol{x})$ et le taux de dégradation $g_i(\boldsymbol{x})\,x_i$ de chaque protéine. La fonction $f_i : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$ est définie comme une somme de fonctions en escalier ponderées par un paramètre de taux et exprimant la logique de la régulation du gène i. La fonction $g_i : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{>0}$ est définie de manière analogue.

La Fig. 1 représente un réseau de régulation génique très simple. Les gènes *a* et *b* codent pour les protéines A et B, qui chacune modulent l'expression des deux gènes. Plus précisement, les protéines A et B répriment l'expression des gènes *a* et *b*, pour des concentrations différentes. Le réseau de régulation génique représenté dans la Fig. 1 peut être décrit par les deux équations suivantes :

$$\dot{x}_a = \kappa_a\, s^-(x_a,\theta_a^2)\, s^-(x_b,\theta_b^1) - \gamma_a\, x_a \tag{2}$$
$$\dot{x}_b = \kappa_b\, s^-(x_a,\theta_a^1)\, s^-(x_b,\theta_b^2) - \gamma_b\, x_b. \tag{3}$$

Le gène *a* est exprimé à un taux $\kappa_a > 0$, si la concentration de la protéine A est au-dessous du seuil θ_a^2 et si la concentration de la protéine B est au-dessous du seuil θ_b^1, c'est à dire, si $s^-(x_a,\theta_a^2)\, s^-(x_b,\theta_b^1) = 1$, avec $s^-(x,\theta)$ la fonction en escalier valant 1 si $x < \theta$ et 0 si $x > \theta$. La protéine A est dégradée spontanément à un taux proportionnel à sa concentration ($\gamma_a > 0$ est une constante). L'équation d'état du gène *b* est interprétée de façon analogue.

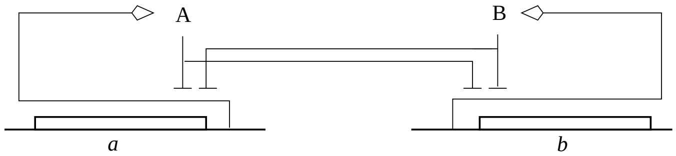

Fig. 1 – *Exemple d'un réseau de régulation génique composé de deux gènes (a et b) codant chacun pour une protéine régulatrice (A et B).*

Les propriétés dynamiques des modèles LPM peuvent être analysées en partitionnant l'espace de phase dans des régions (hyper)rectangulaires appellées *domaines*. La Fig. 2(a) montre la partition pour le réseau de deux gènes donné en exemple. Nous appelons *domaines de seuil* les domaines qui sont localisés dans des (intersections de) plans de seuil, comme D^4 et D^7, et *domaines de régulation* qui ne le sont pas, comme D^1.

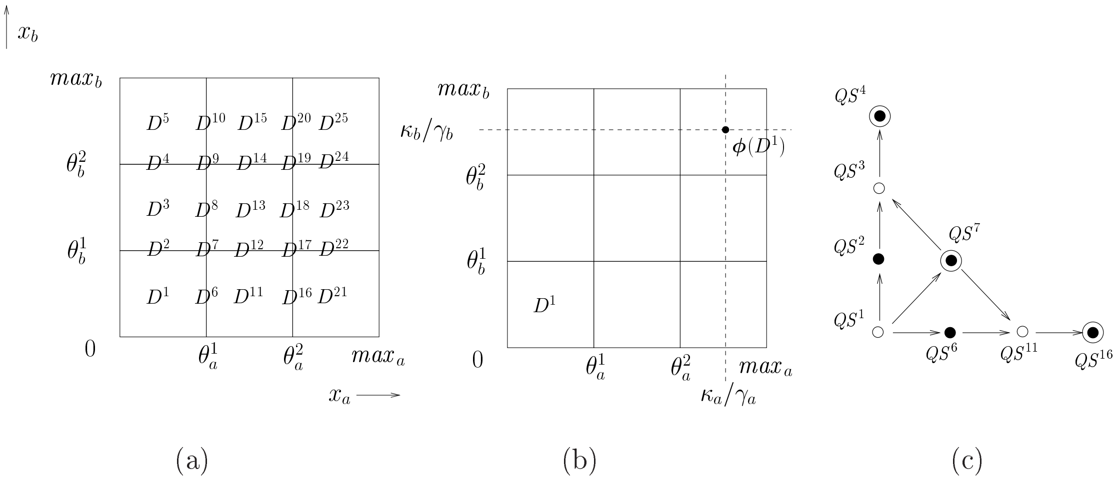

Fig. 2 – *(a) Partition de l'espace de phase dans des domaines de régulation et des domaines de seuil. (b) Analyse du comportement du système dans le domaine D^1, sous les hypothèses $\theta_a^2 < \kappa_a/\gamma_a < max_a$ et $\theta_b^2 < \kappa_b/\gamma_b < max_b$. (c) Graphe de transition produit dans une simulation qualitative à partir de l'état qualitatif QS^1. Les états qualitatifs associés avec des domaines de régulation et de seuil sont indiqués par des ronds pleins et vides, respectivement. Les états qualitatifs associés avec des domaines dans lesquels se trouve un point d'équilibre sont circlés [7].*

En évaluant les fonctions en escalier dans un domaine de régulation D, nous obtenons des équations différentielles d'une forme particulièrement simple : elles sont linéaires et découplées. Il est facile de démontrer que toutes les trajectoires de solutions dans D tendent de façon monotone vers un équilibre stable $\Phi(D)$, appellé *équilibre cible* [11, 21, 24]. Par exemple, dans le domaine D^1, le système (2)-(3) est décrit par les équations différentielles

$$\dot{x}_a = \kappa_a - \gamma_a\, x_a,$$
$$\dot{x}_b = \kappa_b - \gamma_b\, x_b$$

et tend vers l'équilibre cible $\Phi(D^1) = \{(\kappa_a/\gamma_a, \kappa_b/\gamma_b)'\}$. La situation est plus compliquée dans le cas de domaines de seuil, où des discontinuités peuvent se produire. J.-L. Gouzé

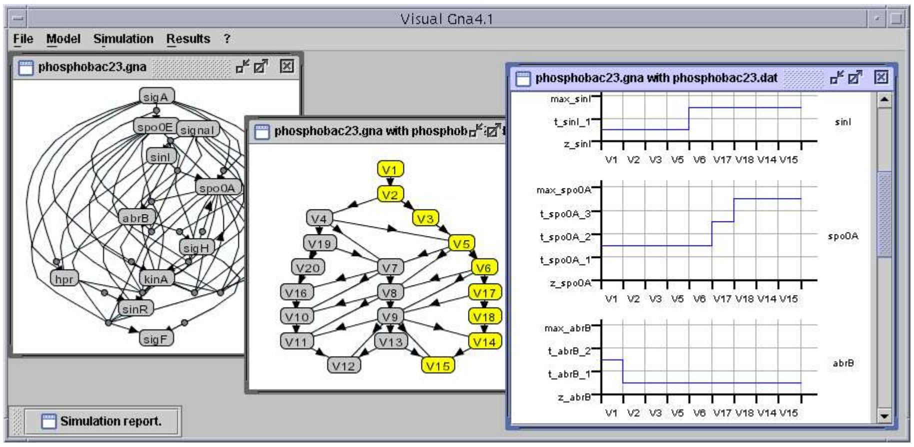

FIG. 3 – *Simulation d'un réseau de régulation génique en utilisant GNA. La fenêtre de gauche montre les gènes et les interactions du réseau contrôlant l'initiation de la sporulation chez B. subtilis. La fenêtre du milieu contient le graphe des transitions entre états résultant de la simulation du réseau pour des conditions initiales induisant la sporulation, tandis que la fenêtre de droite montre une séquence d'états qualitatifs sélectionnée dans le graphe des transitions.*

(COMORE) et T. Sari (LMA) ont montré que, en généralisant les équations différentielles aux inclusions différentielles suivant l'approche de Filippov [13], le comportement du système dans les domaines de seuil peut être décrit de manière analogue au comportement dans les domaines de régulation. L'un des premiers objectifs de l'ARC sera de continuer et d'approfondir ces travaux.

L'analyse mathématique des modèles LPM suggère une description qualitative de la dynamique du système. En fait, nous pouvons définir un *état qualitatif* comme un domaine D et la position relative de son équilibre cible $\Phi(D)$. Le système peut faire une *transition* entre deux états qualitatifs, associés avec des domaines D et D', s'il existe des solutions qui à partir de D' entrent D ou inversement. La *simulation qualitative* du système consiste alors à trouver tous les états qualitatifs atteignables à partir d'un état initial par une ou plusieurs transitions. La Fig. 2(c) montre le graphe de transition obtenu pour une simulation du réseau de deux gènes, à partir de l'état qualitatif associé à D^1. La description des comportements qualitatifs possibles du système à l'aide du graphe de transition fait abstraction du temps pendant lequel le système peut rester dans un état qualitatif avant de faire une transition à un autre état. H. de Jong et M. Page (HELIX), en collaboration avec J.-L. Gouzé (COMORE) et T. Sari (LMA), ont formalisé ce qui est présenté ici succinctement dans une méthode de simulation qualitative [7], assisté par l'outil informatique *GNA (Genetic Network Analyzer)* [6] (Fig. 3).[1] GNA a été utilisé par J. Geiselmann (CEG) et H. de Jong (HELIX) pour l'analyse qualitative du réseau de gènes contrôlant l'initiation de la sporulation chez la bactérie *Bacillus subtilis* [5].

1. GNA est librement disponible pour des fins de recherche académique à travers `http://www-helix.inrialpes.fr/gna`.

3 Analyse d'attracteurs

Les domaines dans l'espace de phase peuvent contenir des points d'équilibre, appellés *équilibres reguliers* dans le cas de domaines de régulation et *équilibres singuliers* dans le cas de domaines de seuil [25]. Tandis que les équilibres réguliers sont toujours stables, les équilibres singuliers peuvent être stables ou instables. Le réseau de régulation analysé dans la Fig. 2 admet trois équilibres singuliers, dont deux sont stables (localisés dans D^4 et D^{16}) et un est instable (localisé dans D^7). Des ensembles de domaines peuvent également contenir des cycles limites [12] et, à partir de $n = 4$, des attracteurs de nature chaotique [20].

Une extension directe des travaux résumés dans la section précédente serait de mieux caractériser les attracteurs des systèmes décrits par les modèles LPM. Cette problématique a fait l'objet de plusieurs études dans la biologie mathématique au cours des trente dernières années. Cependant, les résultats obtenus ne concernent que des cas particuliers de (1) ou sont limités par une description inadéquate du comportement du système dans les domaines de seuil. Dans le cadre de l'ARC, nous préciserons les notions d'équilibre et de cycle limite à l'aide de l'approche de Filippov. Partant de ces définitions, nous formulerons ensuite des critères pour localiser des attracteurs dans l'espace de phase et pour déterminer leur stabilité et leur bassin d'attraction.

Sur la base de ces travaux mathématiques, nous développerons une méthode d'analyse qualitative d'attracteurs qui complétera la méthode de simulation présentée dans la section précédente. Dans un premier temps, des algorithmes pour rechercher les attracteurs d'un modèle LPM seront conçus. L'efficacité de ces algorithmes est un critère de tout premier ordre, étant donné que le nombre de domaines dans l'espace de phase croit de façon exponentielle avec le nombre de variables. Dans un deuxième temps, ces algorithmes seront implémentés dans un nouveau module de GNA et intégrés dans l'interface utilisateur.

La méthode d'analyse qualitative d'attracteurs et son implémentation informatique seront utilisées pour étudier des problématiques biologiques concrètes. En particulier, nous appliquerons ces outils à l'analyse des réseaux intervenant dans la régulation globale de la transcription chez *E. coli* et *Synechocystis*. L'étude de ces réseaux est une activité centrale des deux laboratoires biologiques impliqués (CEG et TOP). Dans le cas d'*E. coli*, dix a vingt protéines, appelés *régulateurs globaux*, assurent la coordination globale de l'organisme [3, 15, 19]. Les interactions entre ces régulateurs constituent donc la charpente des réseaux de régulation génique, qui permettent à un organisme non seulement de s'adapter aux changements de son environnement, mais également une adaptation évolutive à plus long terme. Les modèles LPM de ces réseaux, ainsi que les données d'expression pour tester leur prédictions, seront obtenus à travers le projet inter-EPST « Validation de modèles de réseaux de régulation génique ».

4 Exploration de méthodes de l'automatique et des systèmes hybrides

L'analyse des modèles LPM de réseaux de régulation génique touche à plusieurs sujets de recherche innovants à l'interface des mathématiques appliquées et de l'informatique. Dans le deuxieme volet de l'ARC, nettement plus exploratif que le premier, nous tenterons de creuser ces sujets en appliquant des méthodes de l'automatique et des systèmes hybrides à l'analyse de réseaux de régulation génique. Le but est avant tout de lier les compétences des divers participants en vue d'obtenir des avancées nouvelles qui pourront être poursuivie audelà de l'ARC. Même si la traduction des résultats en outils informatiques ou applications biologiques réelles sera probablement moins immédiate, nous garderons toujours à l'esprit que nos techniques doivent avoir un sens dans le contexte biologique et s'implémenter de façon raisonnablement efficace.

Les systèmes LPM étudiés dans cette ARC ont été décrit comme présentant un caractère hybride [1, 8, 10]. D'une part, l'espace de phase du système est divisé de façon discrète en domaines, mais d'autre part, le système peut évoluer de façon continue à l'intérieur des domaines. Nous profiterons de nos compétences sur l'approche des *systèmes hybrides* (G. Ferrari, SOSSO) pour étudier si et comment les techniques existantes dans ce domaine permettent d'approfondir ou faciliter l'analyse de réseaux de régulation génique.

La méthode de simulation qualitative mentionnée dans la section 2 est liée aux travaux dans le domaine de *raisonnement qualitatif* [17, 22, 27], tout comme la méthode pour l'analyse de systèmes en boucles dévéloppée par O. Bernard et J.-L. Gouzé (COMORE) [2]. Ces méthodes partagent le principe d'utiliser des techniques originales de nature qualitative pour tenir compte du fait que les paramètres des modèles sont souvent mal connus. Dans cette ARC nous comparons les deux méthodes et étudierons la possibilité d'interpreter des réseaux de régulation génique comme des systèmes en boucles.

Nous voulons aussi explorer une voie que nous pensons prometteuse et que l'on pourrait désigner par *automatique des grands réseaux*. Les réseaux de régulation génique que nous avons étudiés jusqu'à maintenant sont composés d'une dizaine de gènes [5]. Les réseaux contrôlant la régulation globale de la transcription chez *E. coli* et *Synechocystis* sont certainement plus grands : ils peuvent impliquer plusieurs dizaines de gènes. Afin de pousser les limites de taille et de complexité des réseaux qu'on peut traiter avec notre approche qualitative, nous explorerons l'applicabilité de certaines notions de l'automatique classique aux modèles LPM. On peut déjà dégager quelques problématiques qui feront appèl aux compétences des projets COMORE et SOSSO :

- Est-il possible de voir un grand réseau de gènes comme composé de plusieurs sous-systèmes comportant un nombre réduit d'entrées et de sorties, interconnectés entre eux ?
- Peut-on simplifier le sous-système en le représentant par son comportement entrée-sortie ? Peut-on ainsi simplifier le système entier ?
- Les techniques classiques d'observation (reconstruire des variables non mesurées) et d'identification (estimer des paramètres, par exemple les interactions dans le réseau...) vont-elles s'appliquer ? Quels aménagements nécessitent-elles ?
- À partir d'une approche fine de description du réseau, peut-on dégager des variables plus générales (par un passage micro-macro) ?

Il est à noter que ces idées commencent à poindre en analyse de réseaux biologiques [16].

5 Déroulement

Le projet est planifié sur deux ans (Fig. 4). Le premier volet du projet, rappelons-nous, concerne l'application de la méthode de Filippov aux systèmes LPM et l'analyse d'attracteurs. Dans les premiers neuf mois, le fondement mathématique de l'approche sera établié, suivi par le développement des algorithmes (six mois), leur implémentation dans GNA (trois mois) et l'application de la méthode aux réseaux étudiés par les biologistes impliqués dans l'ARC (six mois). La nature du deuxième volet étant plus exploratoire, il est plus difficile de distinguer les activités avec la même précision que pour le premier volet. En gros, les neuf premier mois seront dédiés à des études préliminaires de l'application des méthodes de l'automatique et de la théorie des systèmes hybrides. Dans les neuf mois suivants nous préciserons nos idées sur les voies jugées les plus prometteuses, autant sur le plan mathématique que sur le plan informatique. Les derniers six mois seront utilisés pour une première confrontation de nos méthodes avec des exemples biologiques.

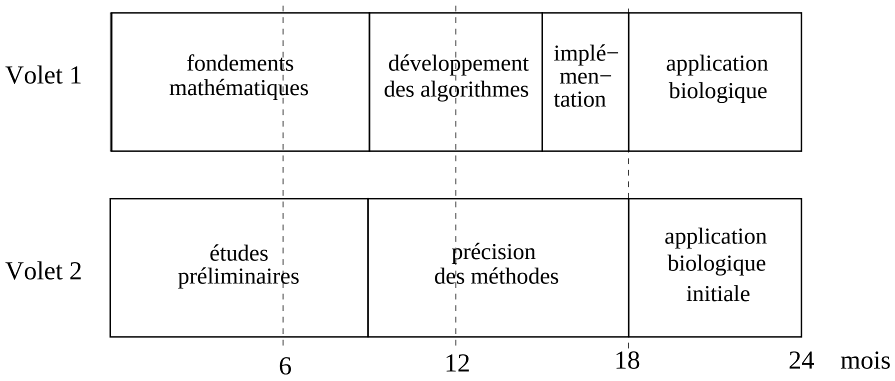

FIG. 4 – *Calendrier de l'ARC.*

6 Conclusions

L'étude de réseaux de régulation génique est fortement stimulée par le développement des nouvelles techniques génomiques permettant de mesurer simultanément le niveau d'expression de tous les gènes d'un organisme. Outre des méthodes expérimentales à haut débit, des approches mathématiques et bioinformatiques sont indispensables pour analyser la dynamique des réseaux de régulation génique. Au sein de l'INRIA, plusieurs projets s'intéressent à cette problématique. Outre les projets participant à cette proposition, nous citons par exemple les projets CONTRAINTES (F. Fages, INRIA Rocquencourt) et MODBIO (A. Bockmayr, INRIA Lorraine), impliqués dans l'ARC CPBIO, ansi que le projet SYMBIOSE (J. Nicolas, INRIA Rennes).

Notre approche se différentie sur deux aspects. D'une part, nous nous concentrons sur une classe de modèles linéaires par morceaux (LPM), bien adaptée aux connaissances biologiques disponibles actuellement. D'autre part, nous utilisons des techniques qualitatives pour l'analyse du comportement des réseaux de régulation génique décrits par ces modèles. Le premier but de l'ARC est d'étudier les propriétés dynamiques des systèmes LPM (solutions, point d'équilibre, cycles limites, stabilité, ...). Ensuite, dans un deuxième volet plus exploratoire, nous tenterons d'appliquer des concepts issus de l'automatique et de la théorie de systèmes hybrides à cette classe de systèmes. Les résultats de ces études mathématiques seront utilisés afin de développer de nouvelles méthodes informatiques pour l'analyse de réseaux de régulation génique. Ces méthodes seront implémentées et mises à l'épreuve dans l'étude des réseaux impliqués dans la régulation globale de la transcription chez les bactéries *Escherichia coli* et *Synechocystis*.

Plus généralement, nous cherchons dans cette ARC à rassembler des compétences multiples et à créer des collaborations enrichissantes autour de la thématique de la dynamique de réseaux de régulation génique, de sa description et son analyse, sur le plan mathématique, informatique et biologique.

Références

[1] R. Alur, C. Belta, F. Ivančíc, 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] O. Bernard and J.-L. Gouzé. Global qualitative description of a class of nonlinear dynamical systems. *Artificial Intelligence*, 136:29–59, 2002.

[3] D. Chatterji and A.K. Ojha. Revisiting the stringent response, ppGpp and starvation signaling. *Current Opinion in Microbiology*, 4(2):160–165, 2001.

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

[5] H. de Jong, J. Geiselmann, G. Batt, C. Hernandez, and M. Page. Qualitative simulation of the initiation of sporulation in *B. subtilis.* Technical Report RR-4527, INRIA, 2002. Soumis pour publication.

[6] H. de Jong, J. Geiselmann, C. Hernandez, and M. Page. Genetic Network Analyzer: Qualitative simulation of genetic regulatory networks. *Bioinformatics*, 2002. Sous presse.

[7] H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and H. Geiselmann. Qualitative simulation of genetic regulatory networks using piecewise-linear models. Technical Report RR-4407, INRIA, 2002. Soumis pour publication.

[8] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari. A clustering technique for the identification of piecewise affine systems. *Automatica*, 39(2), 2002. À paraître.

[9] A.F. Filippov. *Differential Equations with Discontinuous Righthand Sides.* Kluwer Academic Publishers, Dordrecht, 1988.

[10] 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.

[11] L. Glass and S.A. Kauffman. The logical analysis of continuous non-linear biochemical control networks. *Journal of Theoretical Biology*, 39:103–129, 1973.

[12] L. Glass and J.S. Pasternack. Prediction of limit cycles in mathematical models of biological oscillations. *Bulletin of Mathematical Biology*, 40:27–44, 1978.

[13] J.-L. Gouzé and T. Sari. A class of piecewise linear differential equations arising in biological models. *Dynamical Systems*, 2002. À paraître.

[14] 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.

[15] R. Hengge-Aronis. The general stress response in *Escherichia coli.* In *Bacterial Stress Responses*, pages 161–178. ASM Press, Washington, D.C., 2000.

[16] B.N. Kholodenko, A. Kiyatkin, F.J. Bruggeman, E. Sontag, H.V. Westerhoff, and J.B. Hoek. Untangling the wires: A strategy to trace functional interactions in signaling

and gene networks. *Proceedings of the National Academy of Sciences of the USA*, 99(20):12841–12846, 2002.

[17] B. Kuipers. *Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge*. MIT Press, Cambridge, MA, 1994.

[18] H.H. McAdams and L. Shapiro. Circuit simulation of genetic networks. *Science*, 269:650–656, 1995.

[19] S.M. McLeod and R.C. Johnson. Control of transcription by nucleoid proteins. *Current Opinion in Microbiology*, 4(2):152–159, 2001.

[20] T. Mestl, C. Lemay, and L. Glass. Chaos in high-dimensional neural and gene networks. *Physica D*, 98:33–52, 1996.

[21] T. Mestl, E. Plahte, and S.W. Omholt. A mathematical framework for describing and analysing gene regulatory networks. *Journal of Theoretical Biology*, 176:291–300, 1995.

[22] E. Sacks. Automatic qualitative analysis of dynamic systems using piecewise linear approximations. *Artificial Intelligence*, 41:313–364, 1990.

[23] 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*, 62:247–292, 2000.

[24] 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.

[25] E.H. Snoussi and R. Thomas. Logical identification of all steady states: The concept of feedback loop characteristic states. *Bulletin of Mathematical Biology*, 55(5):973–991, 1993.

[26] 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.

[27] L. Travé-Massuyès, P. Dague, and F. Guerrin, editors. *Le raisonnement qualitatif pour les sciences de l'ingénieur*. Hermes, Paris, 1997.

A Équipes et rôles des participants

A.1 CEG (Laboratoire Plasticité et Expression des Génomes Microbiens, FRE2383, université Joseph Fourier, Grenoble)

Johannes Geiselmann et Dominique Schneider font partie de l'équipe CEG dans le laboratoire Plasticité et Expression des Génomes Microbiens, placé sous la double tutelle de l'université Joseph Fourier (UJF) et du CNRS (FRE2383). Ils s'intéressent entre autres aux mécanismes moléculaires de l'adaptation des micro-organismes à leur environnement, ainsi qu'aux régulation de l'expression génique et rythmes circadiens chez les bactéries. Les organismes modèles surtout étudiés dans CEG sont *Escherichia coli* et *Synechocystis* PCC6803.

Participants à l'ARC : Johannes Geiselmann (Pr université Joseph Fourier) et Dominique Schneider (MdC université Joseph Fourier).

Responsable de l'équipe et adresse : Johannes Geiselmann, CERMO, 460 rue de la Piscine, BP53, 38041 Grenoble Cedex 9.

Téléphone : 04 76 63 56 62
Courriel : `Hans.Geiselmann@ujf-grenoble.fr`
Page web : `http://www.ujf-grenoble.fr/ujf/fr/recherche/labujf/pegm.phtml`

A.2 COMORE (INRIA Sophia Antipolis)

Le projet COMORE s'intéresse aux systèmes biologiques, et plus particulièrement aux écosystèmes et aux bioréacteurs. L'objectif global de COMORE est d'appliquer et de développer des méthodes de l'automatique (modélisation, identification, estimation, régulation, contrôle optimal, théorie des jeux) et de la théorie des systèmes dynamiques aux ressources vivantes exploitées (ressources renouvelables), afin d'en améliorer la gestion. Notre cadre de pensée est celui de l'automatique : un système, décrit par des variables d'état, est soumis à des entrées (des actions sur ce système que l'on maîtrise ou non), et est décrit par des sorties (les mesures possibles sur le système).

Plus spécifiquement, Olivier Bernard et Jean-Luc Gouzé s'intéressent à l'analyse qualitative des systèmes dynamiques issus de la biologie. Ils ont mis au point des techniques pour étudier des systèmes connus de manière qualitative seulement (par les signes de la matrice jacobienne par exemple).

Participants à l'ARC : Olivier Bernard (CR INRIA) et Jean-Luc Gouzé (DR INRIA).

Chef de projet et adresse : Jean-Luc Gouzé, INRIA Sophia Antipolis, 2004 route des Lucioles, B.P. 93, 06902 Sophia-Antipolis Cedex.
Téléphone : 04 92 38 78 75
Courriel : `gouze@sophia.inria.fr`
Page web : `http://www.inria.fr/comore`

A.3 HELIX (INRIA Rhône-Alpes)

Le projet HELIX, bi-localisé à l'INRIA Rhône-Alpes (Grenoble) et l'université Claude Bernard (Lyon), s'intéresse à la bioinformatique. Au sein du projet, la bioinformatique est vue comme l'ensemble des méthodes et des outils informatiques destinés à modéliser, analyser et visualiser les diverses entités impliquées dans les processus d'expression et de transmission de l'information génétique, ainsi que les relations que ces entités entretiennent entre elles, en particulier au sein des réseaux géniques et métaboliques. Parmi les axes de recherche de HELIX, on compte la modélisation des gènes et l'inférence de motifs, l'organisation des génomes et cartographie comparée, la phylogénie et évolution, la modélisation dynamique des réseaux de régulation génique, la protéomique et la modélisation de données génomiques et post-génomiques.

Hidde de Jong et Michel Page travaillent sur la modélisation et la simulation de réseaux de régulation génique. Ils ont développé une méthode de simulation qualitative qui a été implémentée dans l'outil GNA (Genetic Network Analyzer) et appliquée à des réseaux de régulation réels.

Participants à l'ARC : Hidde de Jong (CR INRIA) et Michel Page (MdC université Pierre

Mendès France, Grenoble).

Chef de projet et adresse : François Rechenmann, INRIA Rhône-Alpes, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier CEDEX.
Téléphone : 04 76 61 53 65
Courriel : `francois.rechenmann@inrialpes.fr`
Page web : `http://www.inrialpes.fr/helix` (en reconstruction)

A.4 LMA (Laboratoire Mathématiques et Applications, EA 1108, université de Haute Alsace, Mulhouse)

Tewfik Sari est membre du Laboratoire Mathématiques et Applications de l'université de Haute Alsace à Mulhouse (EA 1108). Il s'intéresse aux méthodes géométriques et asymptotiques pour les équations différentielles ordinaires, ainsi qu'à leurs applications, principalement en biologie. Pendant 2001-2002, Tewfik Sari a été accueilli en délégation dans le projet COMORE.

Participants à l'ARC : Tewfik Sari (Pr université de Haute Alsace).

Responsable du laboratoire et adresse : Michel Goze, Laboratoire de Mathématiques et Applications, université de Haute Alsace, 4 rue des Frères Lumière, 68093 Mulhouse.
Téléphone : 03 89 33 64 28
Courriel : M.Goze@uha.fr
Page web : `http://www.math.univ-mulhouse.fr/`

A.5 SOSSO (INRIA Rocquencourt)

Le projet SOSSO mène des recherches en modélisation et commande motivées par des problèmes dans deux domaines d'application privilégiés, l'automobile (modélisation, commande et diagnostic des systèmes de dépollution) et la santé (modélisation des systèmes cardio-respiratoire et de reproduction et de leur contrôle en vue de l'aide au diagnostic). Cela conduit à des problèmes de modélisation mathématique, d'identification et de commande de divers types de systèmes, principalement des systèmes à retards, à hystérésis ou hybrides. Récemment, les travaux de modélisation du coeur (ARC ICEMA2) l'ont conduit à s'intéresser à la modélisation et à la commande de systèmes multi-échelles.

Dans le domaine des systèmes hybrides, les compétences de Giancarlo Ferrari-Trecate sont l'analyse de l'observabilité et de la stabilité, l'estimation d'état et l'identification des systèmes affines par morceaux. Dans le projet, Frédérique Clément et Michel Sorine s'intéressent plus particulièrement aux problèmes de régulation dans des systèmes biologiques (régulation ovarienne par exemple).

Participants à l'ARC : Frédérique Clément (CR INRIA), Giancarlo Ferrari-Trecate (CR INRIA) et Michel Sorine (DR INRIA).

Chef de projet et adresse : Michel Sorine, INRIA Rocquencourt, Domaine de Voluceau,

Roquencourt, B.P. 105, 78153 Le Chesnay Cedex.
Téléphone : 01 39 63 56 48
Courriel : `Michel.Sorine@inria.fr`
Page web : `http://www-rocq.inria.fr/sosso`

A.6 TOP (Laboratoire Organismes Photosynthétiques et Environnement, UMR 8543, École Normale Supérieure, Paris)

L'équipe TOP, dirigée par Jean Houmard, fait partie du Laboratoire Organismes Photosynthétiques et Environnement (UMR 8543), placé sous la double tutelle de l'École Normale Supérieure de Paris et du CNRS. Les membres de TOP ont pour objectif d'élucider les mécanismes moléculaires mis en jeu par les cyanobactéries – procaryotes photosynthétiques qui produisent environ 20% de l'oxygène de la Planète – lors de leur adaptation aux changements dans les paramètres environnementaux. Le rôle que pourraient jouer les nucléotides cycliques (AMPc et GMPc) dans ces phénomènes est plus particulièrement étudié. Chez les organismes eucaryotes, ces seconds messagers sont en effet impliqués dans nombre de régulations globales et les cyanobactéries sont, à ce jour, les seuls procaryotes à posséder les deux types de nucléotides cycliques.

Pour ces études, des approches aussi globales que possible sont menées, transcriptome (J.C. Cadoret) et protéome (J. Wang), sur des cyanobactéries modèles, pour lesquelless les séquences complètes de génome sont disponibles. Parallèlement, une caractérisation biochimique des enzymes et protéines impliquées dans les voies de transduction du signal est menée par Chantal Guidi-Rontani.

Participants à l'ARC : Chantal Guidi-Rontani (CR CNRS) et Jean Houmard (DR CNRS).

Responsable de l'équipe et adresse : Jean Houmard, département de biologie, École Normale Supérieure, 46 rue d'Ulm, 75230 Paris Cedex 05.
Téléphone : 01 44 32 35 19
Courriel : `jhoumard@biologie.ens.fr`
Page web : `http://www.biologie.ens.fr/fr/photoreg/umr8543_99.html`

B Publications récentes des participants

A. Bemporad, **G. Ferrari-Trecate** et M. Morari. Observability and controllability of piecewise affine and hybrid systems. *IEEE Transactions on Automatic Control*, 45(10):1864–1876, 2000.

O. Bernard et **J.-L. Gouzé**. Estimation d'état. In D. Dochain, editeur, *Automatique des bioprocédés*, 87–120. Hermes Science, Paris, 2001.

O. Bernard et **I. Queinnec**. Modèles dynamiques de procédés biochimiques. propriétés des modèles. In D. Dochain, editeur, *Automatique des bioprocédés*, 23–52. Hermes Science, Paris, 2001.

O. Bernard et **J.-L. Gouzé**. Global qualitative behavior of a class of nonlinear biological systems: application to the qualitative validation of phytoplankton growth models. *Artificial Intelligence*, 136:29–59, 2002.

J.A.G. Ochoa de Alda, G. Ajlani et **J. Houmard**. *Synechocystis* PCC 6803 *cya2*, a prokaryotic gene that encodes a guanylyl cyclase. *Journal of Bacteriology*, 182(13):3839–3842, 2000.

J.A.G. Ochoa de Alda et **J. Houmard**. Genomic survey of cAMP and cGMP signalling components in the cyanobacterium *Synechocystis* PCC 6803. *Microbiology*, 146:3183–3194, 2000.

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

H. de Jong, **J. Geiselmann**, C. Hernandez et **M. Page**. Genetic Network Analyzer: Qualitative simulation of genetic regulatory networks. *Bioinformatics*, 2002. Sous presse.

H. de Jong, **J.-L. Gouzé**, C. Hernandez, **M. Page**, **T. Sari** et **J. Geiselmann**. Dealing with discontinuities in the qualitative simulation of genetic regulatory networks. In F. van Harmelen, editeur, *Proceedings of 15th European Conference on Artifical Intelligence, ECAI 2002*, pages 412–416, Amsterdam, 2002. IOS Press.

H. de Jong, **M. Page**, C. Hernandez et **J. Geiselmann**. Qualitative simulation of genetic regulatory networks: Method and application. In B. Nebel, editeur, *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI-01*, 67–73, San Mateo, CA, 2001. Morgan Kaufmann.

G. Ferrari-Trecate, D. Mignone et M. Morari. Moving horizon estimation for hybrid systems. *IEEE Transactions on Automatic Control*, 47(10):1663–1676, 2002.

G. Ferrari-Trecate, M. Muselli, D. Liberati et M. Morari. A clustering technique for the identification of piecewise affine systems. *Automatica*, 39(2), 2002. À paraître.

J.-L. Gouzé and **T. Sari**. A class of piecewise linear differential equations arising in biological models. *Dynamical Systems*, 2002. À paraître.

I. Luque, A. Contreras, G. Zabulon, A. Herrero et **J. Houmard**. Expression of the glutamyl-tRNA synthetase gene from the cyanobacterium *Synechococcus* sp. PCC 7942 depends on nitrogen availability and the global regulator NtcA. *Molecular Microbiology*, 46(4):1157–1167, 2002.

I. Luque, G. Zabulon, A. Contreras et **J. Houmard**. Convergence of two global transcriptional regulators on nitrogen induction of the stress-acclimation gene *nblA* in the cyanobacterium *Synechococcus* sp. PCC 7942. *Molecular Microbiology*, 41(4):937–947, 2001.

D. Papadopoulos, **D. Schneider**, J. Meier-Eiss, W. Arber, R.E. Lenski et M. Blot. Genomic evolution during a 10,000-generation experiment with bacteria. *Proceedings of the National*

Academy of Sciences of the USA, 96(7):3807–3812, 1999.

C. Ranquet, **J. Geiselmann** et A. Toussaint. The tRNA function of SsrA contributes to controlling repression of bacteriophage Mu prophage. *Proceedings of the National Academy of Sciences of the USA*, 98(18):10220–10225, 2001.

D. Schneider, E. Duperchy, J. Depeyrot, E. Coursange, R. Lenski et M. Blot. Genomic comparisons among *Escherichia coli* strains B, K-12, and O157:H7 using IS elements as molecular markers. *BMC Microbiology*, 2(1):18, 2002.