
Analyse qualitative de la dynamique de réseaux de régulation génique par des modèles linéaires par morceaux

Grégory Batt^{1,2} — Hidde de Jong¹ — Johannes Geiselman³ — Jean-Luc Gouzé⁴ — Michel Page^{1,5} — Delphine Ropers¹ — Tewfik Sari⁶ — Dominique Schneider³

¹*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
{Gregory.Batt, Hidde.de-Jong, Michel.Page, Delphine.Ropers}@inrialpes.fr*

²*Université Joseph Fourier, Grenoble*

³*Institut National de Recherche en Informatique et en Automatique (INRIA)
Unité de recherche Sophia Antipolis
2004 route des Lucioles, BP 93, 06902 Sophia Antipolis
Jean-Luc.Gouze@sophia.inria.fr*

⁴*Laboratoire Adaptation et Pathogénie des Microorganismes (CNRS UMR 5163),
Université Joseph Fourier, Bâtiment Jean Roget, Faculté de Médecine-Pharmacie
Domaine de la Merci, 38700 La Tronche
{Hans.Geiselman, Dominique.Schneider}@ujf-grenoble.fr*

⁵*École Supérieure des Affaires, Université Pierre Mendès France, Grenoble*

⁶*Laboratoire de Mathématiques, Université de Haute-Alsace
4 rue des Frères Lumière, 68093 Mulhouse
T.Sari@uha.fr*

RÉSUMÉ. Le fonctionnement et le développement des organismes vivants sont contrôlés par des réseaux complexes de gènes, protéines, petites molécules et leurs interactions, appelés réseaux de régulation génique. Nous discutons dans cette revue une méthode de modélisation et de simulation qualitative de réseaux de régulation génique, bien adaptée au manque actuel de données biologiques quantitatives et précises. Cette méthode est basée sur une classe d'équations différentielles linéaires par morceaux (LPM) possédant des propriétés mathématiques favorables.

Technique et sciences informatiques – /2006. Modélisation et simulation pour la post-génomique

Nous décrivons quelques résultats de base concernant la dynamique de systèmes LPM, introduisons une abstraction discrète sur laquelle repose la méthode de simulation qualitative et illustrons notre approche avec une application concernant la réponse au stress nutritionnel de la bactérie Escherichia coli.

ABSTRACT. The functioning and development of living organisms is controlled by large and complex networks of genes, proteins, small molecules, and their interactions, so-called genetic regulatory networks. In this review we discuss a method for the qualitative modeling and simulation of genetic regulatory networks which is well-adapted to the current lack of precise and quantitative biological data. The method is based on a class of piecewise-linear (PL) differential equation models with favorable mathematical properties. We review some basic results concerning the dynamics of PL systems, describe a discrete abstraction underlying the qualitative simulation method, and illustrate our approach by means of an application in the context of the nutritional stress response in the bacterium Escherichia coli.

MOTS-CLÉS : Réseaux de régulation génique, modélisation et simulation qualitative, équations différentielles linéaires par morceaux, abstractions discrètes, Genetic Network Analyzer (GNA), réponse au stress nutritionnel, Escherichia coli

KEYWORDS: Genetic regulatory networks, qualitative modeling and simulation, piecewise-linear differential equations, discrete abstractions, Genetic Network Analyzer (GNA), nutritional stress response, Escherichia coli

1. Introduction

Le fonctionnement et le développement des organismes vivants sont contrôlés par des réseaux complexes de gènes, protéines, petites molécules et leurs interactions, appelés *réseaux de régulation génique*. Grâce aux efforts concertés de la génétique, de la biologie moléculaire, de la biochimie et de la physiologie, des quantités énormes de données ont été accumulées sur les composants de ces réseaux. Mais de façon surprenante, malgré les nombreuses avancées dans la caractérisation de la structure des réseaux, il n'est actuellement pas bien compris comment le comportement des systèmes contrôlés par ces réseaux émerge des interactions entre leurs composants. Ceci a incité un nombre croissant de chercheurs à passer de l'étude de la structure à celle du comportement des réseaux de régulation génique, dans le cadre d'un mouvement plus large, aussi bien nommé génomique fonctionnelle, biologie des systèmes ou encore biologie intégrative (Bower *et al.*, 2001; Collado-Vides *et al.*, 1996; Kitano, 2001; Szallasi *et al.*, 2006; Pécou *et al.*, 2006).

Ces dix dernières années, la boîte à outils du biologiste moléculaire a été étendue à des techniques expérimentales à haut débit de transcriptomique, de protéomique et de métabolomique. Ces techniques permettent d'étudier à l'échelle du génome la structure et le comportement des réseaux de régulation génique (Goodacre *et al.*, 2004; Lockhart *et al.*, 2000; Mockler *et al.*, 2005; Pandey *et al.*, 2000; Sauer, 2004; Zhu *et al.*, 2003). Il est cependant clair qu'en plus d'outils expérimentaux puissants, l'étude du comportement des réseaux de régulation génique doit faire appel à des outils mathématiques et informatiques. En effet, étant donné que la plupart des réseaux d'intérêt biologique sont constitués d'un grand nombre de composants impliqués dans des boucles de rétroaction complexes, la prédiction du comportement du système par un simple raisonnement intuitif devient rapidement infaisable. L'utilisation combinée de modèles mathématiques et d'outils informatiques permet au contraire de décrire le réseau de façon claire et précise et de prédire son comportement, à la fois systématiquement et efficacement.

Plusieurs méthodes de *modélisation, analyse et simulation de réseaux de régulation génique* ont été proposées ces vingt dernières années (de Jong, 2002; Hasty *et al.*, 2001; McAdams *et al.*, 1998; Smolen *et al.*, 2000). Cependant, ces travaux se sont essentiellement focalisés sur deux approches, basées respectivement sur l'utilisation d'*équations différentielles* et d'*équations maîtresses stochastiques*. L'application de ces dernières repose sur des théories bien établies décrivant la cinétique de systèmes de réactions biochimiques, aussi bien d'un point de vue déterministe que stochastique (Cornish-Bowden, 1995; Gillespie, 1992; Heinrich *et al.*, 1996; van Kampen, 1997). Plusieurs méthodes mathématiques et outils informatiques permettent de transformer les équations d'un modèle en prédictions testables expérimentalement. De nombreux exemples témoignent de la capacité des approches classiques à permettre d'appréhender le fonctionnement de réseaux de régulation génique d'importance biologique, tels que le réseau contrôlant la réponse d'*Escherichia coli* à l'infection par le bactériophage λ (Arkin *et al.*, 1998; McAdams *et al.*, 1995), la régulation du cycle cellulaire

chez la levure et les eucaryotes supérieurs (Tyson *et al.*, 2001) ou encore, le contrôle du rythme circadien chez la drosophile (Leloup *et al.*, 2003).

Les approches classiques permettent de faire des prédictions numériques précises sur les propriétés dynamiques déterministes ou stochastiques des réseaux de régulation génique. Cependant, dans la plupart des cas, l'application d'équations différentielles et d'équations maîtresses stochastiques est loin d'être évidente. Premièrement, les interactions mettent en jeu des réactions biochimiques dont le mécanisme n'est souvent pas connu ou alors de façon incomplète, ce qui complique la formulation des modèles. Deuxièmement, les données quantitatives sur les paramètres cinétiques et les concentrations moléculaires sont généralement absentes, même dans le cas de systèmes bien connus, ce qui rend les méthodes numériques standards difficiles à appliquer.

En pratique, le modélisateur dispose d'informations bien plus faibles sur les composants du réseaux et leurs interactions. Typiquement, au lieu de détails sur le mécanisme par lequel une protéine régule l'expression d'un gène, il sait seulement si la protéine active ou inhibe ce gène. Et quand bien même il a été montré, par exemple, que la protéine se fixe à un ou plusieurs sites dans la région promotrice du gène, les valeurs numériques des constantes de dissociation et des autres paramètres ne sont souvent pas disponibles. Il est possible au mieux d'inférer que la protéine régulatrice se fixe fortement ou faiblement à l'ADN, avec une affinité plus forte pour un site qu'un autre. Même si le développement de techniques permettant une caractérisation plus fine et plus quantitative des processus cellulaires est un domaine en pleine expansion (Levin *et al.*, 2004; Mihalcescu *et al.*, 2004; Phair *et al.*, 2001; Shav-Tal *et al.*, 2004), l'essentiel des progrès des techniques expérimentales va plus dans le sens d'une augmentation du nombre d'observations simultanées que dans la précision des mesures.

Au-delà de ces considérations pratiques, on peut se demander si des informations quantitatives précises sont vraiment essentielles pour comprendre le fonctionnement des réseaux de régulation génique. En fait, on peut raisonnablement faire l'hypothèse que de nombreuses propriétés dynamiques importantes des systèmes vivants, telles que la réponse d'une cellule bactérienne à la perturbation de son environnement, ne dépendent pas de valeurs numériques précises, ou même d'un mécanisme réactionnel spécifique. Alors qu'en général différents individus d'une même espèce se développent ou répondent à une perturbation de la même façon, il est fortement probable que les valeurs des paramètres décrivant les interactions varient d'un individu à l'autre, du fait du polymorphisme génétique ou de fluctuations physiologiques. De même, du fait de la diversification des espèces au cours de l'évolution, des espèces différentes peuvent créer la même fonction régulatrice par des mécanismes différents (Rao *et al.*, 2004). En d'autres termes, les *propriétés dynamiques qualitatives*, c'est-à-dire, les propriétés qui sont invariantes pour une certaine gamme de valeurs de paramètres et de mécanismes réactionnels, pourraient être plus importantes que les propriétés dynamiques quantitatives. Les propriétés qualitatives décrivent la connection étroite existant entre le comportement du système et la structure du réseau d'interactions, indépendamment des détails quantitatifs de ce dernier.

Pour l'ensemble des raisons décrites ci-dessus, il y a aujourd'hui un intérêt croissant pour les *approches qualitatives* de modélisation, d'analyse et de simulation des réseaux de régulation génique et des autres réseaux d'interactions biologiques. Ces approches ont toutes en commun le fait d'être capables d'inférer des propriétés qualitatives de la dynamique du système à partir des données incomplètes ou non quantitatives. Ceci n'exclut évidemment pas que des variantes de ces approches peuvent également être utilisés pour faire des prédictions quantitatives à partir de modèles très détaillés. Parmi les approches qualitatives nous citons : réseaux booléens (Kauffman, 1993; Thomas *et al.*, 1990), réseaux de Pétri (Koch *et al.*, 2005; Simão *et al.*, 2005), graphes (dirigés) (Barabasi *et al.*, 2004; Shen-Orr *et al.*, 2002), algèbres de processus (Regev *et al.*, 2001), équations différentielles qualitatives (Heidtke *et al.*, 1998), automates hybrides (Ghosh *et al.*, 2004). Plusieurs revues ont justifié l'utilisation des approches qualitatives, et les ont comparées avec des formalismes existants (de Jong *et al.*, 2003b; Gagneur *et al.*, 2005).

L'objectif de cet article est de donner une revue d'une méthode particulière de modélisation et de simulation des réseaux de régulation génique (de Jong *et al.*, 2004b), basée sur une classe d'*équations différentielles linéaires par morceaux (LPM)* proposée il y a une trentaine d'années par Glass et Kauffman (Glass *et al.*, 1973). Tout en faisant abstraction des mécanismes réactionnels précis, les modèles LPM décrivent des aspects essentiels de la régulation génique. Leur forme mathématique simple permet en outre l'analyse qualitative du comportement du réseau de régulation génique (Coutinho *et al.*, 2006; Edwards, 2000; Edwards *et al.*, 2001; Farcot, 2006; Gouzé *et al.*, 2002; Glass *et al.*, 1978; Mestl *et al.*, 1995; Plahte *et al.*, 2005). Plus précisément, la méthode présentée dans cet article décrit la dynamique qualitative d'un système LPM à l'aide d'un graphe dit de transition, constitué d'états qualitatifs et de transitions entre états qualitatifs. Les états qualitatifs correspondent à des régions de l'espace de phase dans lesquelles le système se comporte de façon qualitativement homogène, alors que les transitions correspondent à des trajectoires de solutions connectant des régions adjacentes. Nous avons montré que le graphe de transition est invariant pour de larges gammes de valeurs de paramètres, définies par des contraintes d'inégalités pouvant être inférées de la littérature expérimentale. De plus, les états qualitatifs et les transitions entre états qualitatifs peuvent être calculés à partir de ces contraintes à l'aide de règles symboliques. La méthode de modélisation et de simulation qualitative a été implémentée dans le logiciel Genetic Network Analyzer (GNA) (de Jong *et al.*, 2003a), et appliquée à l'analyse de divers réseaux de régulation génique bactériens (de Jong *et al.*, 2004a; Ropers *et al.*, 2006; Usseglio Viretta *et al.*, 2004).

Dans les deux sections suivantes, nous introduisons les modèles LPM de réseaux de régulation génique et décrivons certains résultats mathématiques de base concernant leur dynamique. C'est sur ces résultats que repose l'abstraction discrète utilisée dans la section 4 pour décrire la dynamique qualitative d'un réseau au moyen d'un graphe de transition. Dans la section 5, la méthode de modélisation et simulation qualitative est appliquée à l'analyse d'un réseau génique complexe, contrôlant la réponse au stress nutritionnel de la bactérie *E. coli*. L'article se termine par une discussion de

la méthode dans le contexte d'autres travaux et présente brièvement une extension de la méthode qui permet de traiter des réseaux plus grands.

2. Modèles LPM de réseaux de régulation génique

La figure 1(a) montre l'exemple d'un réseau de régulation génique simple. Les gènes a et b , transcrits à partir de promoteurs séparés, codent les protéines A et B, dont chacune contrôle l'expression des deux gènes.¹ Les protéines A et B répriment l'expression des gènes a et b à des concentrations différentes. La répression est réalisée par fixation des protéines sur des sites de régulation chevauchant les promoteurs. Tout en étant simple, ce réseau d'*inhibition réciproque* est un composant élémentaire de réseaux réels plus complexes et nous permet d'analyser quelques phénomènes caractéristiques de la différenciation cellulaire (Monod *et al.*, 1961; Thomas *et al.*, 1990).

La dynamique de réseaux de régulation génique peut être modélisée par un système d'équations différentielles linéaires par morceaux (LPM) ayant la forme suivante (Glass *et al.*, 1973; Mestl *et al.*, 1995; Thomas *et al.*, 1990) :

$$\dot{\mathbf{x}} = \mathbf{h}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) - \mathbf{g}(\mathbf{x}) \mathbf{x}, \quad [1]$$

où $\mathbf{x} = (x_1, \dots, x_n)' \in \Omega$ représente le vecteur de concentrations cellulaires des protéines, $\mathbf{f} = (f_1, \dots, f_n)'$, $\mathbf{g} = \text{diag}(g_1, \dots, g_n)$ et $\Omega \subset \mathbb{R}_{\geq 0}^n$ est un hypercube n -dimensionnel de l'espace des phases. Le taux de variation de chaque concentration protéique x_i , $i \in \{1, \dots, n\}$, est ainsi défini comme la différence entre le taux de synthèse $f_i(\mathbf{x})$ et le taux de dégradation $g_i(\mathbf{x}) x_i$ de la protéine. Les modèles LPM peuvent être facilement étendus afin de prendre en compte les *variables d'entrée* $\mathbf{u} = (u_1, \dots, u_m)'$, représentant la concentration de protéines et de petites molécules dont la synthèse et la dégradation sont régulées à l'extérieur du système (de Jong *et al.*, 2004b).

La fonction $f_i : \Omega \rightarrow \mathbb{R}_{\geq 0}$ décrit comment le taux de synthèse de la protéine codée par le gène i dépend des concentrations \mathbf{x} des protéines dans la cellule. Plus précisément, la fonction f_i est définie par

$$f_i(\mathbf{x}) = \sum_{l \in L_i} \kappa_i^l b_i^l(\mathbf{x}), \quad [2]$$

où $\kappa_i^l > 0$ est un *paramètre de synthèse*, $b_i^l : \Omega \rightarrow \{0, 1\}$ une *fonction de régulation* continue par morceaux et L_i un ensemble, éventuellement vide, d'indices de fonctions de régulation. La fonction g_i décrit la régulation de la dégradation des protéines. Elle

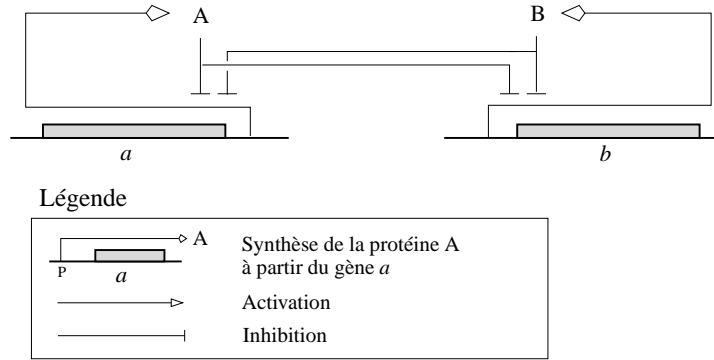
1. Par convention, les gènes sont notés en italique et les noms des protéines commencent par une majuscule.

est définie de façon analogue à f_i , excepté que g_i doit être strictement positive. En outre, nous notons γ les *paramètres de dégradation* afin de les distinguer formellement des paramètres de synthèse κ . Une conséquence directe de ces définitions est que h est une fonction *linéaire par morceaux (LPM)*.

La fonction de régulation b_i^l décrit les conditions sous lesquelles la protéine codée par le gène i est synthétisée (ou dégradée) à un taux κ_i^l (ou $\gamma_i^l x_i$). Elle est définie par une combinaison de fonctions en escalier et constitue l'équivalent arithmétique d'une fonction booléenne exprimant la logique de la régulation génique (Glass *et al.*, 1973; Mestl *et al.*, 1995; Thomas *et al.*, 1990). Plus précisément, les conditions de synthèse ou de dégradation sont exprimées à l'aide des fonctions en escalier s^+ , s^- :

$$s^+(x_j, \theta_j) = \begin{cases} 1, & \text{si } x_j > \theta_j, \\ 0, & \text{si } x_j < \theta_j, \end{cases} \quad \text{et } s^-(x_j, \theta_j) = 1 - s^+(x_j, \theta_j), \quad [3]$$

avec x_j un élément de x et θ_j une constante correspondant à un seuil de concentration. Il est à noter que s^+ , s^- ne sont pas définies au seuil, où une discontinuité se produit. L'utilisation de fonctions en escalier est motivée par l'observation que l'activité d'un gène dépend souvent de façon sigmoïdale de la concentration d'une protéine régulatrice (Ptashne., 2004; Yagil *et al.*, 1971).



(a)

$$\dot{x}_a = \kappa_a s^-(x_a, \theta_a^2) s^-(x_b, \theta_b^1) - \gamma_a x_a,$$

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

(b)

Figure 1. (a) Exemple d'un réseau de régulation génique simple, composé des deux gènes a et b , leurs protéines A et B , ainsi que leurs interactions régulatrices. Voir la figure 8 pour les notations. (b) Equations d'état pour le réseau décrit en (a).

Les équations d'état pour le réseau simple donné en exemple sont montrées dans la figure 1(b). Le gène a est exprimé à un taux κ_a , si la concentration de la protéine

A est en-dessous de son seuil θ_a^2 et la concentration de la protéine B en-dessous de son seuil θ_b^1 , c'est-à-dire, si $s^-(x_a, \theta_a^2) s^-(x_b, \theta_b^1)$ vaut 1. De façon similaire, le gène b est exprimé à un taux κ_b , si la concentration de la protéine A est en-dessous du seuil θ_a^1 et la concentration de la protéine B en-dessous de son seuil θ_b^2 . La dégradation des protéines n'est pas régulée dans l'exemple et est donc simplement proportionnelle à leur concentration (avec des paramètres de dégradation γ_a ou γ_b).

L'utilisation de fonctions en escalier $s^\pm(x_j, \theta_j)$ dans l'équation [1] amène des complications. En effet, les fonctions en escalier sont discontinues pour $x_j = \theta_j$ et le second membre h de l'équation [1] peut donc présenter des discontinuités dans l'union des hyperplans de seuil

$$\Theta = \bigcup_{i \in \{1, \dots, n\}, l_i \in \{1, \dots, p_i\}} \{x \in \Omega \mid x_i = \theta_i^{l_i}\}, \quad [4]$$

où p_i est le nombre de seuils de concentration de la protéine codée par le gène i . Les discontinuités dans Θ sont une conséquence directe de l'approximation des fonctions sigmoïdales par des fonctions en escalier. Comme nous le montrons plus loin, cette approximation facilite l'analyse mathématique, mais nécessite une définition explicite du système dans Θ pour préserver des propriétés dynamiques importantes, *e.g.*, l'occurrence de points d'équilibre.

Afin de compléter la définition de modèles LPM dans les plans de seuil, nous suivons une approche souvent utilisée en théorie du contrôle, proposée à l'origine par A.F. Filippov (Filippov, 1988). Elle consiste à étendre l'équation différentielle $\dot{x} = h(x)$, $x \in \Omega \setminus \Theta$, à une *inclusion* différentielle

$$\dot{x} \in K(x), \text{ avec } K(x) = \overline{\text{co}}(\{\lim_{y \rightarrow x, y \notin \Theta} h(y)\}), \quad x \in \Omega, \quad [5]$$

où $\overline{\text{co}}(P)$ est le plus petit ensemble convexe fermé contenant l'ensemble P et où $\{\lim_{y \rightarrow x, y \notin \Theta} h(y)\}$ représente l'ensemble des valeurs limites de $h(y)$ pour $y \notin \Theta$ et $y \rightarrow x$. Cette approche a été appliquée dans le contexte de la modélisation de réseaux de régulation génique par J.-L. Gouzé et T. Sari (Gouzé *et al.*, 2002). Ce n'est pas la seule approche possible (voir, *e.g.*, (Mestl *et al.*, 1995; Plahte *et al.*, 2005)), mais elle a l'avantage d'être mathématiquement rigoureuse, de se prêter bien à un traitement informatique et, au moins dans les cas simples, de retrouver des résultats intuitifs.

En pratique, $K(x)$ est souvent difficile à calculer parce que le plus petit ensemble convexe fermé peut être un polyèdre complexe dans Ω . Nous utilisons donc une extension alternative de l'équation différentielle :

$$\dot{x} \in H(x), \text{ avec } H(x) = \overline{\text{rect}}(\{\lim_{y \rightarrow x, y \notin \Theta} h(y)\}), \quad x \in \Omega, \quad [6]$$

où $\overline{rect}(P)$ est le plus petit ensemble *hyperrectangulaire* contenant l'ensemble P (de Jong *et al.*, 2004b). L'utilisation de \overline{rect} a l'avantage que l'on peut décomposer $H(\mathbf{x})$ en produit cartésien $H_1(\mathbf{x}) \times \dots \times H_n(\mathbf{x})$, et réécrire [6] comme le système d'inclusions différentielles $\dot{x}_i \in H_i(\mathbf{x})$, $i \in \{1, \dots, n\}$. Il est à noter que $H(\mathbf{x})$ est une surapproximation de $K(\mathbf{x})$ pour tout $\mathbf{x} \in \Omega$.

Formellement, nous définissons le *système LPM* Σ comme le triplet (Ω, Θ, H) , c'est-à-dire, la fonction H , définie sur l'espace des phases Ω de dimension n , l'union des hyperplans de seuil Θ incluse. Une *solution* du système LPM Σ sur un intervalle de temps I est une solution de l'inclusion différentielle [6] sur I , c'est-à-dire, une fonction absolument continue $\xi(t)$ telle que $\dot{\xi}(t) \in H(\xi(t))$ est vrai presque partout sur I (Filippov, 1988). En particulier, il se peut que la dérivée de ξ n'existe pas, et que donc $\dot{\xi}(t) \in H(\xi(t))$ ne soit pas satisfait, si ξ atteint ou quitte Θ au temps t .

Pour tout $\mathbf{x}_0 \in \Omega$ et $\tau \in \mathbb{R}_{>0}$, $\Xi_\Sigma(\mathbf{x}_0, \tau)$ représente l'ensemble des solutions $\xi(t)$ du système LPM Σ , pour la condition initiale $\xi(0) = \mathbf{x}_0$ et $t \in [0, \tau]$. L'existence d'au moins une solution ξ sur un intervalle de temps $[0, \tau]$, $\tau > 0$, est garantie pour chaque $\xi(0) = \mathbf{x}_0$ dans Ω (Filippov, 1988). Cependant, il n'existe en général pas de solution unique. L'ensemble $\Xi_\Sigma = \bigcup_{\mathbf{x}_0 \in \Omega, \tau > 0, \tau \text{ fini}} \Xi_\Sigma(\mathbf{x}_0, \tau)$ est composé de toutes les solutions de Σ sur un intervalle de temps fini. Nous restreignons ici notre analyse à l'ensemble des solutions dans Ξ_Σ qui atteignent et quittent un hyperplan de seuil un nombre fini de fois. Ceci permet d'exclure le comportement asymptotiques de certaines solutions du type Zeno (Johansson *et al.*, 1999), qui n'a pas une grande importance pratique pour les applications visées.

3. Analyse mathématique des systèmes LPM

Les propriétés dynamiques de Σ peuvent être analysées dans un hyperrectangle de dimension n de l'espace des phases $\Omega = \Omega_1 \times \dots \times \Omega_n$, où $\Omega_i = \{x_i \in \mathbb{R} \mid 0 \leq x_i \leq \max_i\}$, $i \in \{1, \dots, n\}$. \max_i représente la concentration maximale de la protéine codée par le gène i . Les hyperplans de seuil de dimension $(n-1)$, $\{\mathbf{x} \in \Omega \mid x_i = \theta_i^i\}$, $l_i \in \{1, \dots, p_i\}$, $i \in \{1, \dots, n\}$, partitionnent Ω en des régions (hyper)rectangulaires appelées *domaines*. Dans chacune de ces régions, la concentration d'une protéine est égale à un seuil ou est bornée par des seuils. Plus précisément, le domaine $D \subseteq \Omega$ est défini par $D = D_1 \times \dots \times D_n$, où chaque D_i , $i \in \{1, \dots, n\}$, est donné par l'une des équations suivantes :

$$\begin{aligned}
 D_i &= \{x_i \in \Omega_i \mid 0 \leq x_i < \theta_i^1\}, \\
 D_i &= \{x_i \in \Omega_i \mid x_i = \theta_i^1\}, \\
 D_i &= \{x_i \in \Omega_i \mid \theta_i^1 < x_i < \theta_i^2\}, \\
 D_i &= \{x_i \in \Omega_i \mid x_i = \theta_i^2\}, \\
 &\dots \\
 D_i &= \{x_i \in \Omega_i \mid x_i = \theta_i^{p_i}\}, \\
 D_i &= \{x_i \in \Omega_i \mid \theta_i^{p_i} < x_i \leq \max x_i\}.
 \end{aligned} \tag{7}$$

\mathcal{D} est l'ensemble des domaines dans Ω . La figure 2(a) montre la partition en domaines de l'espace des phases correspondant au réseau à deux gènes de l'exemple. Nous distinguons les domaines tels que D^2 et D^7 , qui sont localisés sur des (intersections de) hyperplans de seuil, contrairement à des domaines tels que D^1 et D^3 . Les domaines du premier type sont appelés *singuliers* et les domaines du second type, *réguliers*. Nous notons respectivement \mathcal{D}_s et \mathcal{D}_r les ensembles de domaines singuliers et réguliers.

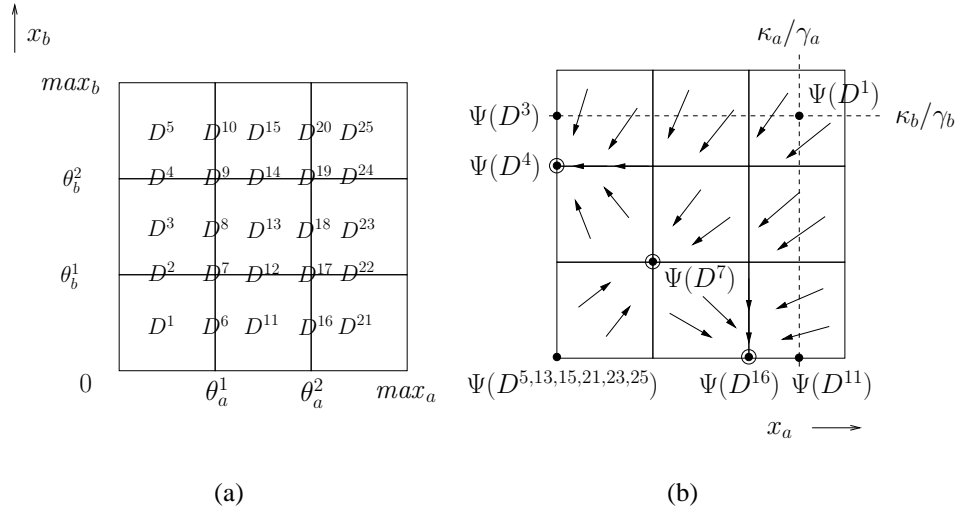


Figure 2. (a) Partition en domaines de l'espace des phases correspondant au modèle de la figure 1(b). (b) Champ de vecteurs et ensembles focaux des domaines (persistants) dans (a). Les ensembles focaux entourés sont des points d'équilibre du système. Les contraintes suivantes ont été utilisées : $\kappa_a/\gamma_a > \theta_a^2$ et $\kappa_b/\gamma_b > \theta_b^2$.

Afin de préparer l'analyse des propriétés dynamiques de [1] dans les domaines réguliers et singuliers, nous définissons quelques concepts topologiques simples. Pour tout domaine, $D \subseteq \Omega$, de dimension $k \in \{0, \dots, n\}$, l'hyperplan support de D ,

$supp(D) \subseteq \Omega$, est l'hyperplan de dimension k contenant D . La *frontière* de D dans $supp(D)$ est notée ∂D . Supposons que D est un domaine singulier, *i.e.* $D \in \mathcal{D}_s$. Alors, $R(D)$ est défini comme l'ensemble des domaines réguliers D' ayant D dans leur frontière, *i.e.* $R(D) = \{D' \in \mathcal{D}_r \mid D \subseteq \partial D'\}$.

Suivant la définition de l'inclusion différentielle [6], dans un domaine régulier D , $H(\mathbf{x})$ se réduit à un singleton $\{\boldsymbol{\mu}^D - \boldsymbol{\nu}^D \mathbf{x}\}$, pour tout $\mathbf{x} \in D$, où $\boldsymbol{\mu}^D$ est un vecteur de (sommes de) paramètres de synthèse et $\boldsymbol{\nu}^D$ une matrice diagonale de (sommes de) paramètres de dégradation. On trouve le résultat selon lequel toutes les solutions $\boldsymbol{\xi}$ dans D convergent de façon monotone vers l'*ensemble focal*

$$\Psi(D) = \{\boldsymbol{\psi}(D)\}, \quad [8]$$

où $\boldsymbol{\psi}(D)$ est appelé le *point focal*, défini par $\boldsymbol{\psi}(D) = (\boldsymbol{\nu}^D)^{-1} \boldsymbol{\mu}^D$ (Glass *et al.*, 1973; Plahte *et al.*, 1994; Snoussi, 1989). Nous faisons l'hypothèse générique que les ensembles focaux $\Psi(D)$, pour tout $D \in \mathcal{D}_r$, ne sont pas localisés dans les hyperplans de seuil Θ . Si $\boldsymbol{\psi}(D) \in D$, alors pour $t \rightarrow \infty$ toutes les solutions dans D approchent l'ensemble focal, qui est alors un point d'équilibre asymptotiquement stable du système. Si $\boldsymbol{\psi}(D) \notin D$, toutes les solutions quitteront D après un certain temps.

La figure 2(b) montre les ensembles focaux des domaines réguliers dans l'espace des phases associé au système de l'exemple. Dans D^1 par exemple, le modèle LPM de la figure 1(b) se réduit à $\dot{x}_a = \kappa_a - \gamma_a x_a$ et $\dot{x}_b = \kappa_b - \gamma_b x_b$, si bien que l'ensemble focal $\Psi(D^1)$ est égal à $\{(\kappa_a/\gamma_a, \kappa_b/\gamma_b)\}$. Parce que $\Psi(D^1)$ est en-dehors de D^1 , les trajectoires dans D^1 quitteront le domaine après un certain temps. Des domaines réguliers différents ont généralement des ensembles focaux différents. Par exemple, l'ensemble focal de D^3 est donné par $\Psi(D^3) = \{(0, \kappa_b/\gamma_b)\}$.

Que se passe-t-il quand une solution quitte un domaine régulier et entre dans un domaine singulier ? L'extension des équations différentielles aux inclusions différentielles permet de répondre à cette question. En effet, dans un domaine singulier, le second membre de l'inclusion différentielle [6] se réduit à $H(\mathbf{x}) = \overline{rect}(\{\boldsymbol{\mu}^{D'} - \boldsymbol{\nu}^{D'} \mathbf{x} \mid D' \in R(D)\})$, pour tout $\mathbf{x} \in D$ (de Jong *et al.*, 2004b). L'ensemble focal associé avec le domaine s'écrit alors

$$\Psi(D) = supp(D) \cap \overline{rect}(\{\boldsymbol{\psi}(D') \mid D' \in R(D)\}), \quad [9]$$

et n'est en général pas constitué d'un seul point dans les domaines de dimension élevée (de Jong *et al.*, 2004b). Deux cas peuvent être distingués. Si $\Psi(D)$ est vide, alors chaque solution traverse D instantanément et D est appelé un domaine *instantané*. Dans le cas contraire, certaines solutions atteignant D (mais pas nécessairement toutes) resteront dans ce domaine pendant un certain temps, glissant le long de l'hyperplan support de D . Le domaine est alors appelé *persistant*. Si $\Psi(D)$ est constitué d'un seul point, alors toutes les solutions dans D convergent de façon monotone vers

ce point. Dans le cas où $\Psi(D)$ n'est pas constitué d'un seul point, une propriété de monotonie plus faible peut être démontrée (de Jong *et al.*, 2004b).

Pour illustrer ces concepts, le domaine singulier D^2 dans l'exemple est instantané ($\Psi(D^2) = \emptyset$), car l'intersection de $\overline{\text{rect}}(\{\psi(D^1), \psi(D^3)\})$ et de $\text{supp}(D^2)$ est vide (fig. 2(b)). Le fait que les solutions atteignant D^2 traversent ce domaine de façon instantanée correspond à ce qu'on attend intuitivement si on prend en compte la direction du champ de vecteur dans D^1 et D^3 (figure 2). D^4 est en revanche un domaine persistant, dans lequel les solutions glissent le long du plan de seuil et convergent vers l'ensemble focal $\Psi(D^4) = \text{supp}(D^4) \cap \overline{\text{rect}}(\{\psi(D^3), \psi(D^5)\}) = \{(0, \kappa_b/\gamma_b)\}$. De nouveau, ce comportement est cohérent avec le champ de vecteurs dans D^3 et D^5 , et un résultat similaire serait généré par une simulation numérique à l'aide de la méthode d'Euler ou un autre schéma d'intégration numérique. En outre, les solutions ressemblent à celles du système d'équations différentielles ordinaires obtenu en remplaçant les fonctions en escalier par des fonctions sigmoïdales raides.

Si $\Psi(D) \cap D = \emptyset$, toutes les solutions quitteront D après un certain temps. Par contre, si $\Psi(D) \cap D \neq \emptyset$, il existe des solutions dans D atteignant ou approchant de façon asymptotique l'ensemble focal $\Psi(D)$. A partir de [6], il peut être démontré que pour chaque $\psi \in \Psi(D) \cap D$ nous avons $\mathbf{0} \in H(\psi)$, et qu'il existe ainsi une solution $\xi(t)$ telle que $\xi(t) = \psi$ pour tout $t \geq 0$. Par conséquent, $\Psi(D) \cap D$ est un ensemble de points d'équilibre du système, appelé *ensemble d'équilibre*. La présence d'ensembles d'équilibre, ainsi que la non-unicité des solutions des inclusions différentielles (section 2), nécessitent des définitions étendues de stabilité (Casey *et al.*, 2005).

Il est facile de vérifier que le domaine D^4 contient un ensemble d'équilibre, ici constitué d'un seul point, $(0, \kappa_b/\gamma_b)$. Ce point d'équilibre est asymptotiquement stable, parce qu'il est atteint ou asymptotiquement approché par toutes les solutions dans le voisinage de $\Psi(D^4)$, propriété qui découle de la direction du champ de vecteurs dans les domaines réguliers D^3 et D^5 . La figure montre également les ensembles focaux des domaines singuliers persistants, D^7 et D^{16} . L'ensemble focal de D^7 coïncide avec le domaine lui-même, réduit à un point dans Ω . $\Psi(D^7)$ est un point d'équilibre instable, parce que des solutions partant d'un voisinage de $\Psi(D^4)$ peuvent s'éloigner du point d'équilibre, comme le suggère le champ de vecteurs dans D^1 , D^3 , D^{11} et D^{13} . Tout comme $\Psi(D^4)$, $\Psi(D^{16})$ est un point d'équilibre asymptotiquement stable.

En résumé, l'utilisation d'inclusions différentielles rend possible l'extension de la définition des modèles LPM [1] à l'ensemble de l'espace des phases d'une manière naturelle et mathématiquement correcte. En particulier, les résultats ci-dessus montrent que la dynamique locale des systèmes LPM Σ est relativement simple : dans chaque domaine de l'espace des phases, le système converge de façon monotone – dans le sens habituel du terme dans le cas des domaines réguliers, dans un sens plus faible dans le cas des domaines singuliers – vers un ensemble focal. Ceci n'implique pas que la dynamique globale des systèmes LPM est également simple. Au contraire, étant donné que l'ensemble focal change généralement en passant d'un domaine à un autre, la dynamique globale des systèmes LPM peut être extrêmement complexe (Edwards *et*

al., 2001; Lewis *et al.*, 1991; Mestl *et al.*, 1997). Cependant, le fait que dans chaque domaine le système se comporte de façon qualitativement homogène suggère une abstraction discrète de la dynamique qui se révèlera très utile pour l'étude des propriétés qualitatives du comportement des réseaux de régulation génique.

4. Simulation qualitative par des modèles LPM

La définition d'une *abstraction discrète* ou *qualitative* de la dynamique de systèmes LPM requiert l'introduction préalable d'un système de transition continu ayant les mêmes propriétés d'atteignabilité que le système LPM Σ original (Batt *et al.*, 2005b). Considérons $x \in D$ et $x' \in D'$, où $D, D' \in \mathcal{D}$ sont des domaines.² S'il existe une solution ξ de Σ passant en x à un instant τ et atteignant x' à un instant τ' , sans quitter $D \cup D'$ dans l'intervalle de temps $[\tau, \tau']$, alors l'absolue continuité de ξ implique que D et D' sont soit égaux, soit contigus. Plus précisément, l'un des trois cas suivants peut se produire : $D = D'$, $D \subseteq \partial D'$ ou $D' \subseteq \partial D$. Par conséquent, nous distinguons trois types de transitions continues, correspondant à ces trois cas : *interne*, notée $x \xrightarrow{int} x'$, *dimension croissante*, notée $x \xrightarrow{dim^+} x'$ et *dimension décroissante*, notée $x \xrightarrow{dim^-} x'$. Ces deux derniers termes expriment que la dimension du domaine augmente ou diminue, en allant de D à D' , respectivement. Le *système de transition LPM continu* d'un système LPM $\Sigma = (\Omega, \Theta, H)$ est alors donné par $\Sigma\text{-TS} = (\Omega, L, \rightarrow)$, où $L = \{int, dim^+, dim^-\}$ est l'ensemble d'étiquettes précisant les types de transition, et $\rightarrow \subseteq \Omega \times L \times \Omega$ est la relation de transition décrivant l'évolution continue du système. Plus précisément, étant donné que $x \in D$ et $x' \in D'$, nous avons $x \xrightarrow{l} x'$ si et seulement si il existe $\xi \in \Xi_\Sigma$ et τ, τ' , tels que $0 \leq \tau < \tau'$, $\xi(\tau) = x$, $\xi(\tau') = x'$ et

- 1) si $l = int$, alors pour tout $t \in [\tau, \tau']$ on a $\xi(t) \in D$ et $D = D'$,
- 2) si $l = dim^+$, alors pour tout $t \in (\tau, \tau']$ on a $\xi(t) \in D'$ et $D' \neq D$,
- 3) si $l = dim^-$, alors pour tout $t \in [\tau, \tau')$ on a $\xi(t) \in D$ et $D \neq D'$.

Nous appelons toute séquence de points (x^0, \dots, x^m) dans Ω , $m \geq 0$, un *chemin* de $\Sigma\text{-TS}$ si et seulement si une séquence de transitions permet d'atteindre x^m à partir de x^0 . C'est-à-dire, (x^0, \dots, x^m) est un chemin si et seulement si pour tout $i \in \{0, \dots, m-1\}$, il existe $l \in L$ tel que $x^i \xrightarrow{l} x^{i+1}$. Il n'est pas difficile de montrer qu'un système LPM Σ et $\Sigma\text{-TS}$, le système de transition LPM continu correspondant, ont des propriétés d'atteignabilité équivalentes. En effet, pour tout $x, x' \in \Omega$, il existe un chemin dans $\Sigma\text{-TS}$ menant de x à x' si et seulement si il existe une solution ξ de Σ sur un intervalle de temps fini $[\tau, \tau']$ passant de x à x' .

Considérons une solution ξ du système LPM Σ pour notre exemple du réseau à deux gènes (figure 3(a)). ξ passe par les points x^0, \dots, x^4 , à des instants τ^0, \dots, τ^4 ,

2. Nous précisons que x' représente ici un point localisé dans le domaine D' et non pas la dérivée de x . Cette dernière est notée \dot{x} , cf. l'équation [1].

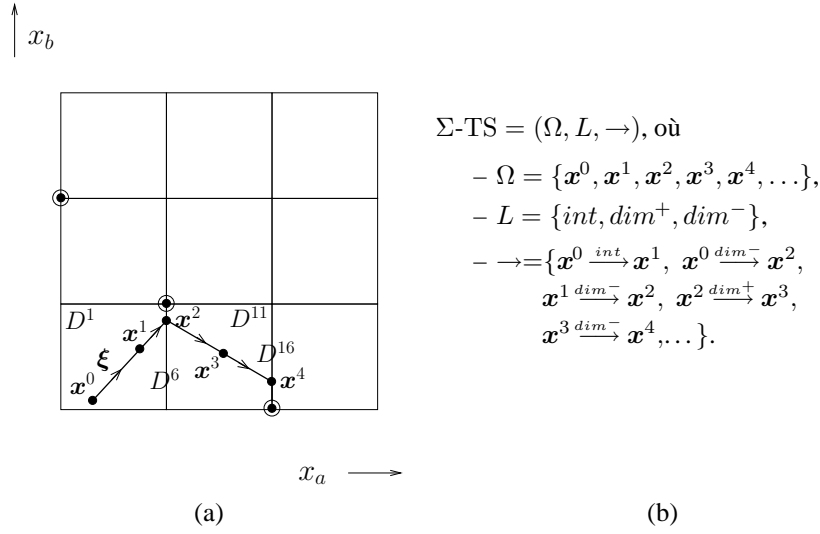


Figure 3. Trajectoire d'une solution ξ du système LPM Σ , débutant dans D^1 et atteignant D^{16} , en passant par les points x^0, \dots, x^4 . (b) Description partielle du système de transition LPM continu Σ -TS, illustrant la relation de ce dernier avec le système LPM Σ .

respectivement. A partir de la définition du système de transition continu, et du fait que x^0 et x^1 font partie du même domaine D^1 , nous inférons l'existence d'une transition *int* de x^0 à x^1 dans Σ -TS. En suivant de nouveau la définition de Σ -TS, parce que $x^1 \in D^1$, $x^2 \in D^6$ et $\xi(t) \in D^1$ pour tout $t \in [\tau^1, \tau^2)$, il existe une transition de type *dim*⁻ de x^1 à x^2 . De la même façon, il existe une transition *dim*⁻ de x^0 à x^2 . En outre, il existe une transition *dim*⁺ de x^2 à x^3 , parce que $x^2 \in D^6$, $x^3 \in D^{11}$ et $\xi(t) \in D^{11}$ pour tout $t \in (\tau^2, \tau^3]$. Il est à noter que les étiquettes de transition *dim*⁺ et *dim*⁻ sont cohérentes avec la dimension des domaines, qui vaut 2 pour D^1 et D^{11} , et 1 pour D^6 . La séquence de points $(x^0, x^1, x^2, x^3, x^4)$ est un chemin de Σ -TS.

A l'aide de la partition \mathcal{D} de l'espace des phases Ω , introduite dans la section 3, une abstraction discrète ou qualitative peut être formulée. L'abstraction est basée sur la relation d'équivalence \sim_Ω , par laquelle deux points sont équivalents quand ils appartiennent au même domaine, *i.e.* $x \sim_\Omega x'$ si et seulement si il existe un $D \in \mathcal{D}$, tel que $x \in D$ et $x' \in D$. La description discrète du système LPM Σ , appelée *système de transition LPM qualitatif*, prend alors la forme du système de transition quotient de Σ -TS par la relation d'équivalence \sim_Ω (Alur *et al.*, 2000; Chutinan *et al.*, 2001). Plus précisément, étant donné que Σ -TS = (Ω, L, \rightarrow) , nous avons Σ -QTS = $(\Omega/\sim_\Omega, L, \rightarrow_{\sim_\Omega})$, où $\Omega/\sim_\Omega = \mathcal{D}$, et $\rightarrow_{\sim_\Omega} \subseteq \mathcal{D} \times L \times \mathcal{D}$, telle que $D \xrightarrow{l}_{\sim_\Omega} D'$ si et seulement si il existe $\xi \in \Xi_\Sigma$ et τ, τ' , $0 \leq \tau < \tau'$, tels que $\xi(\tau) \in D$, $\xi(\tau') \in D'$ et

- 1) si $l = int$, alors pour tout $t \in [\tau, \tau']$ on a $\xi(t) \in D$ et $D = D'$,
- 2) si $l = dim^+$, alors pour tout $t \in (\tau, \tau']$ on a $\xi(t) \in D'$ et $D' \neq D$,
- 3) si $l = dim^-$, alors pour tout $t \in [\tau, \tau')$ on a $\xi(t) \in D$ et $D \neq D'$.

Autrement dit, il existe une transition d'un domaine D à un autre domaine D' si et seulement si il existe une solution de Σ atteignant D' à partir de D en temps fini, sans quitter $D \cup D'$. La différence principale entre Σ -TS et Σ -QTS est que ce dernier est un système de transition *fini* (au lieu d'*infini*), puisque l'espace des états de Σ -QTS est égal à un ensemble fini, \mathcal{D} . Il s'agit d'une propriété qui rend l'analyse de la dynamique du système plus facile à réaliser.

Les domaines d'un système de transition LPM qualitatif sont aussi appelés *états qualitatifs*. Comme pour Σ -TS, nous appelons une séquence de domaines (D^0, \dots, D^m) , $m \geq 0$, un *chemin* de Σ -QTS, si et seulement si, pour tout $i \in \{0, \dots, m-1\}$, il existe un $l \in L$ tel que $D^i \xrightarrow{l, \sim_\Omega} D^{i+1}$. Un chemin dans Σ -QTS décrit de façon qualitative une évolution temporelle possible des concentrations des protéines. Si un domaine D contient un ensemble d'équilibre, c'est-à-dire, si $\Psi(D) \cap D \neq \emptyset$, alors le domaine est appelé un *état d'équilibre qualitatif* du système de transition. Les états qualitatifs et les transitions entre états qualitatifs de Σ -QTS sont souvent représentés sous forme d'un *graphe de transition*, défini comme $G = (\mathcal{D}, \rightarrow_{\sim_\Omega})$.

Considérons à nouveau la situation illustrée dans la figure 3. Selon la définition ci-dessus d'un système de transition qualitatif, l'existence des transitions de x^0 à x^1 , de x^1 à x^2 et de x^2 à x^3 dans Σ -TS, donne respectivement lieu à une transition *int* de D^1 à D^1 , une transition *dim*⁻ de D^1 à D^6 et une transition *dim*⁺ de D^6 à D^{11} . Le système de transition qualitatif et le graphe de transition correspondant à l'exemple du réseau à deux gènes sont représentés dans la figure 4. L'évolution temporelle des concentrations des protéines le long du chemin $(D^1, D^6, D^{11}, D^{16})$ de Σ -QTS est montrée dans la figure 5.

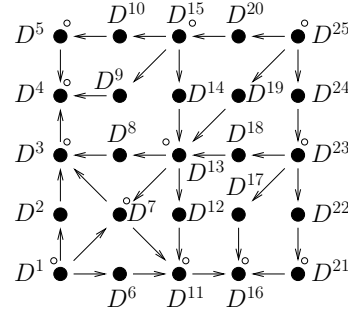
Des résultats standards de la théorie des systèmes hybrides nous permettent directement de conclure de la définition du système de transition LPM qualitatif Σ -QTS que celui-ci est une *approximation conservative* du système de transition LPM continu Σ -TS. C'est-à-dire que s'il existe un chemin (x^0, \dots, x^m) dans Σ -TS, alors il existe également un chemin (D^0, \dots, D^m) dans Σ -QTS tel que $x^i \in D^i$ pour tout $i \in \{0, \dots, m\}$. Il est à noter que l'inverse n'est pas nécessairement vrai,³ ce qui a pour une conséquence qu'une certaine vigilance est de rigueur quand Σ -QTS est utilisé pour vérifier les propriétés d'atteignabilité de Σ -TS (section 6).

La dynamique continue du système LPM dépend évidemment des valeurs des paramètres du modèle, c'est-à-dire, des paramètres de synthèse et de dégradation et des seuils de concentration. Deux systèmes LPM ayant des valeurs de paramètres différentes peuvent donc correspondre aux systèmes de transition LPM qualitatifs diffé-

3. Dans la terminologie de la théorie des systèmes hybrides (Alur *et al.*, 2000; Chutinan *et al.*, 2001), Σ -QTS est une *simulation*, mais pas une *bisimulation* de Σ -TS, par rapport à la relation d'équivalence \sim_Ω .

Σ -QTS = $(\mathcal{D}, L, \rightarrow_{\sim\Omega})$, où

- $\mathcal{D} = \{D^1, D^2, \dots, D^{25}\}$,
- $L = \{int, dim^+, dim^-\}$,
- $\rightarrow_{\sim\Omega} = \{D^1 \xrightarrow{int} \sim_{\Omega} D^1, D^1 \xrightarrow{dim^-} \sim_{\Omega} D^2, D^1 \xrightarrow{dim^-} \sim_{\Omega} D^6, \dots, D^{25} \xrightarrow{dim^-} \sim_{\Omega} D^{24}\}$.



(a)

(b)

$$0 < \theta_a^1 < \theta_a^2 < max_a, \theta_a^2 < \kappa_a/\gamma_a < max_a,$$

$$0 < \theta_b^1 < \theta_b^2 < max_b, \theta_b^2 < \kappa_b/\gamma_b < max_b.$$

(c)

Figure 4. (a) Système de transition qualitatif Σ -QTS et (b) graphe de transition $G = (\mathcal{D}, \rightarrow_{\sim\Omega})$, correspondant à l'exemple du réseau à deux gènes de la figure 1. Les transitions d'un état vers lui-même, qui distinguent les états persistants des états instantanés, sont représentées par de petits cercles vides. (c) Contraintes d'inégalité sur les paramètres, définissant une région dans l'espace des paramètres dans laquelle le graphe G est invariant.

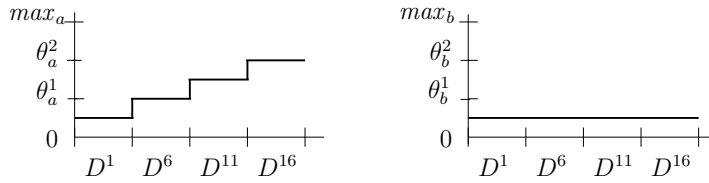


Figure 5. Profil qualitatif de l'évolution temporelle des concentrations des protéines le long du chemin $(D^1, D^6, D^{11}, D^{16})$ du système de transition qualitatif Σ -QTS, défini dans la figure 4. Les profils montrent comment les bornes sur les concentrations, définissant les domaines, varient le long du chemin.

rents, et satisfaire des propriétés dynamiques qualitatives différentes. Cependant, en suivant la logique de la section 1, on peut se demander si l'on ne peut pas caractériser des ensembles de valeurs de paramètres pour lesquelles le système a la même dynamique qualitative.

Dans cette optique, soit $\Gamma \subseteq \mathbb{R}_{>0}^q$ l'espace des paramètres du système LPM Σ , où q est le nombre de paramètres dans [1]. Nous introduisons une deuxième relation

d'équivalence, $\sim_\Gamma \subseteq \Gamma \times \Gamma$, telle que deux vecteurs de paramètres $\mathbf{p}, \mathbf{p}' \in \Gamma$ sont équivalents, si les systèmes de transition LPM qualitatifs correspondants, ou leurs graphes de transition, sont isomorphes. Γ/\sim_Γ dénote l'espace de paramètres quotient, étant donné la relation d'équivalence \sim_Γ . Nous avons montré dans (de Jong *et al.*, 2004b) que des régions $P \subseteq \Gamma$ peuvent être définies, telles que pour tout $\mathbf{p}, \mathbf{p}' \in P$, il est vrai que $\mathbf{p} \sim_\Gamma \mathbf{p}'$. Autrement dit, il existe un $Q \in \Gamma/\sim_\Gamma$, tel que $P \subseteq Q$. Par conséquent, pour tous les vecteurs de paramètres dans P , le système possède les mêmes propriétés dynamiques qualitatives.

Les régions P sont définies par deux types de contraintes d'inégalité qui peuvent généralement être obtenues à partir des données expérimentales disponibles. La spécification de l'ordre des concentrations de seuil $\theta_i^{l_i}$ de chaque protéine constitue le premier type de contraintes, appelées les *inégalités de seuil*. Le second type de contraintes sont les *inégalités focales*. Ces contraintes sont données par la spécification de l'ordre des composantes des points focaux, $\psi_i(D)$, $D \in \mathcal{D}_r$, par rapport aux seuils de concentration. La figure 4(c) montre les contraintes d'inégalité pour notre réseau à deux gènes. Les contraintes sont définies de telle sorte que la protéine A inhibe l'expression du gène b à une concentration plus faible que celle nécessaire à la repression de son propre gène ($\theta_a^1 < \theta_a^2$). En outre, quand le gène a est actif, la concentration de la protéine A tend vers un niveau au-dessus duquel l'auto-inhibition se produit ($\kappa_a/\gamma_a > \theta_a^2$). Les contraintes d'inégalité pour la protéine B sont interprétées de façon analogue. Le graphe de transition dans la figure 4(b) est invariant pour toutes les valeurs de paramètres satisfaisant les contraintes d'inégalité dans la figure 4(c).

Les contraintes d'inégalité jouent aussi un rôle important dans le calcul du système de transition LPM qualitatif (de Jong *et al.*, 2004b). Ce calcul est grandement simplifié par le fait que les domaines D et les ensembles focaux $\Psi(D)$ sont des ensembles hyperrectangulaires. Ceci nous permet d'exprimer les ensembles sous forme de produits, *i.e.* $D = D_1 \times \dots \times D_n$ et $\Psi(D) = \Psi_1(D) \times \dots \times \Psi_n(D)$. Par conséquent, le calcul peut être réalisé séparément pour chaque dimension à l'aide des contraintes d'inégalité. Par exemple, afin de déterminer les transitions à partir du domaine D^1 , il suffit de connaître l'ordre relatif de κ_a/γ_a et θ_a^1 dans la dimension de x_a , ainsi que l'ordre relatif de κ_b/γ_b et θ_b^1 dans la dimension de x_b . Étant donné les contraintes d'inégalité données dans la figure 4(c), et la convergence monotone des solutions dans D^1 vers $\Psi(D^1)$, on peut inférer qu'il existe des solutions débutant dans D^1 et atteignant D^2 , D^6 ou D^7 . Des règles symboliques pour accomplir ces calculs sont données dans (de Jong *et al.*, 2004b).

Les règles symboliques de calcul du système de transition LPM qualitatif à partir d'un modèle LPM complété par des contraintes d'inégalité ont été implémentées en Java. Ceci a donné lieu à l'outil informatique *Genetic Network Analyzer (GNA)* (de Jong *et al.*, 2003a) (figure 6),⁴ qui est équipé d'une interface utilisateur graphique appelée *VisualGNA*. En pratique, étant donné que le nombre de domaines dans l'es-

4. GNA est librement disponible pour les chercheurs académiques à <http://www-helix.inrialpes.fr/gna>.

pace des phases croît de façon exponentielle avec le nombre de gènes du réseau, il n'est pas toujours possible de calculer le système de transition LPM qualitatif entier (*i.e.*, le graphe de transition entier). La plupart du temps, GNA est donc utilisé pour faire ce qui est appelé une *simulation qualitative*, c'est-à-dire, déterminer la partie du graphe de transition atteignable à partir d'un ensemble spécifié de domaines initiaux. GNA a été utilisé pour l'analyse de plusieurs réseaux de régulation génique, comme l'initiation de la sporulation chez la bactérie du sol *Bacillus subtilis* (de Jong *et al.*, 2004a) et le « quorum sensing » chez la bactérie pathogène *Pseudomonas aeruginosa* (Usseglio Viretta *et al.*, 2004).

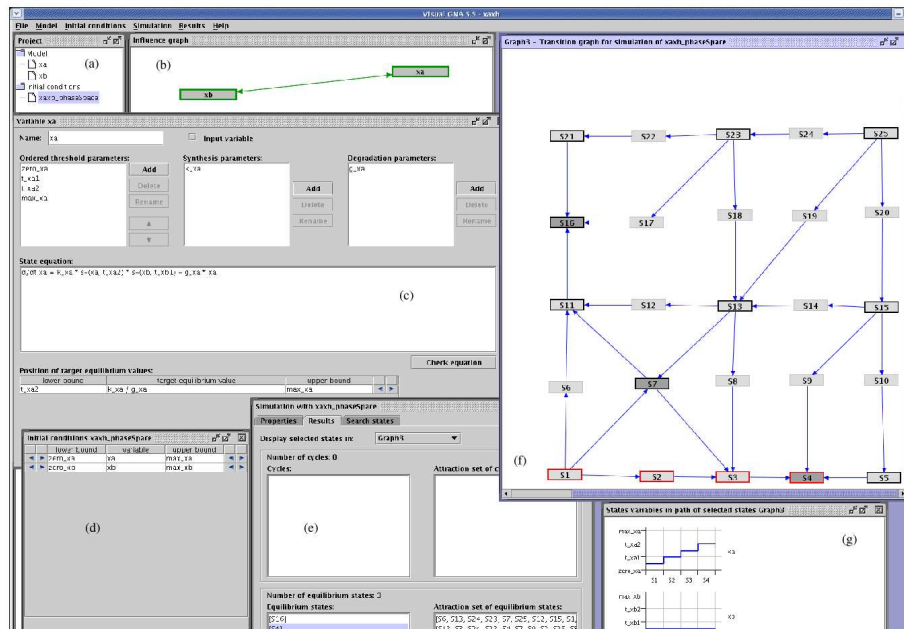


Figure 6. Capture d'écran de GNA, montrant la simulation du réseau à deux gènes de la figure 1 : (a) variables du modèle et conditions initiales, (b) graphe d'interactions entre les gènes du modèle, (c) spécification de l'équation différentielle et des contraintes d'inégalité d'une variable, (d) spécification de domaines initiaux, (e) résumé des résultats de simulation, (f) graphe de transition produit par la simulation et (g) évolution temporelle qualitative des variables de concentration le long d'un chemin.

5. Simulation qualitative de la réponse au stress nutritionnel chez *E. coli*

Dans leur environnement naturel, des bactéries comme *Escherichia coli* rencontrent rarement des conditions de croissance continue. Tant que des nutriments sont

disponibles, les cellules d'*E. coli* croissent et se divisent rapidement, ce qui conduit à une augmentation exponentielle de leur biomasse, un état appelé *phase exponentielle*. Lorsqu'un nutriment essentiel vient à manquer, les bactéries ne peuvent plus maintenir des vitesses de croissance élevées et la population entre dans un état de croissance nulle, appelé *phase stationnaire* (figure 7). Lors de la transition de la phase exponentielle à la phase stationnaire, chaque bactérie individuelle d'*E. coli* subit de profonds changements physiologiques, notamment de la morphologie, du métabolisme cellulaire ou encore de l'expression des gènes (Huisman *et al.*, 1996). Les changements rendent la cellule capable de survivre à des périodes prolongées de privation et de résister à des stress multiples. Cette *réponse au stress nutritionnel* est réversible et la croissance peut donc reprendre, dès que des nutriments sont à nouveau disponibles.

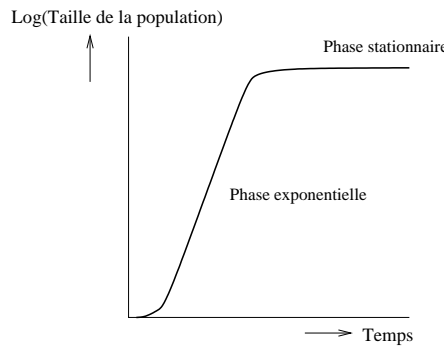


Figure 7. États de croissance possibles d'une population bactérienne : les phases exponentielle et stationnaire.

Au niveau moléculaire, la transition de la phase exponentielle à la phase stationnaire est contrôlée par un réseau de régulation génique complexe intégrant une variété de signaux environnementaux (Hengge-Aronis, 2000; Martinez-Antonio *et al.*, 2003; Wick *et al.*, 2004). La base moléculaire de l'adaptation de la croissance d'*E. coli* à des conditions de stress nutritionnel a fait l'objet de nombreuses études depuis plusieurs décennies (Hengge-Aronis, 2000). Cependant, malgré l'énorme accumulation de données sur les gènes, les protéines et les autres molécules connues pour être impliquées dans le processus d'adaptation au stress, il n'y a pas à l'heure actuelle de compréhension globale de la façon dont la réponse de la cellule émerge du réseau d'interactions moléculaires. De plus, les valeurs numériques des paramètres caractérisant les interactions et les concentrations moléculaires sont, à quelques exceptions près (Bettenbrock *et al.*, 2005; Botsford *et al.*, 1992; Harman, 2001; Wang *et al.*, 2001), absentes. Ceci rend difficile l'application des méthodes traditionnelles de modélisation dynamique des réseaux de régulation génique.

Les raisons décrites ci-dessus ont motivé l'analyse qualitative du réseau de la réponse au stress nutritionnel chez *E. coli* par la méthode présentée dans les sections précédentes (Ropers *et al.*, 2006). Pour entreprendre cette étude, nous nous sommes tout d'abord focalisés sur la partie du réseau de régulation génique qui permet à *E.*

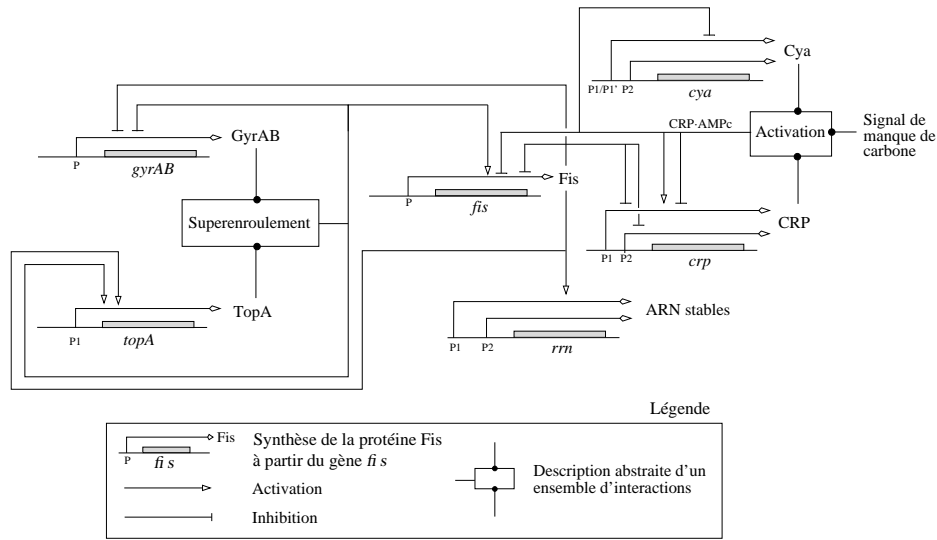


Figure 8. Réseau des gènes, protéines et interactions régulatrices clés impliqués dans la réponse au manque de carbone chez *E. coli*. Bien que simplifiée, la notation suit les conventions graphiques proposées par Kohn (Kohn., 2001). Le contenu des boîtes notées « Activation » et « Superenroulement » est détaillé dans (Ropers et al., 2006).

coli de répondre à un type de stress nutritionnel qui a été particulièrement bien décrit, le manque de source de carbone. L'objectif de l'étude était de simuler la réponse de la bactérie à l'absence ou à la présence de source de carbone dans le milieu de croissance. Dans ce but, un modèle initial simple du réseau a été construit sur la base des données de la littérature. Il inclut six gènes considérés comme jouant un rôle clé dans la réponse au manque de carbone (fig. 8). En particulier, le réseau inclut des gènes codant les protéines dont l'activité dépend de la transduction du signal de manque de carbone (le régulateur global *crp* et l'adénylate cyclase *cya*), des gènes impliqués dans le métabolisme (le régulateur global *fis*), la croissance cellulaire (les gènes *rrn* codant les ARN stables) et le superenroulement de l'ADN, un modulateur important de l'expression génique (la topoisomérase *topA* et la gyrase *gyrAB*).

La représentation graphique du réseau a été traduite en un modèle LPM supplémenté par des contraintes d'inégalités sur les paramètres. Le modèle résultant est constitué de sept variables, une variable de concentration pour le produit de chacun des six gènes, et une variable d'entrée représentant la présence ou l'absence d'un signal de manque de carbone (Ropers et al., 2006). Les 34 paramètres sont contraints par 48 contraintes d'inégalité (24 inégalités de seuil et 24 inégalités focales), dont le choix a largement été déterminé par des données expérimentales.

Pour illustration, l'équation différentielle linéaire par morceaux et les contraintes d'inégalités de la variable x_{topA} sont données ci-dessous.

$$\begin{aligned} \dot{x}_{topA} &= \kappa_{topA} s^+(x_{gyrAB}, \theta_{gyrAB}^2) s^-(x_{topA}, \theta_{topA}^1) s^+(x_{fis}, \theta_{fis}^4) \\ &\quad - \gamma_{topA} x_{topA}, \\ 0 &< \theta_{topA}^1 < \theta_{topA}^2 < max_{topA}, \\ \theta_{topA}^2 &< \kappa_{topA} / \gamma_{topA} < max_{topA}. \end{aligned}$$

L'équation et les inégalités indiquent qu'en présence d'une concentration élevée de Fis ($s^+(x_{fis}, \theta_{fis}^4) = 1$) et d'un niveau élevé du taux de superenroulement de l'ADN ($s^+(x_{gyrAB}, \theta_{gyrAB}^2) s^-(x_{topA}, \theta_{topA}^1) = 1$), la concentration de TopA augmente, pour converger vers une valeur plus élevée ($\kappa_{topA} / \gamma_{topA} > \theta_{topA}^2$). Afin d'obtenir l'équation pour x_{topA} , nous avons fait l'approximation raisonnable que le niveau de superenroulement s'ajuste instantanément, à l'échelle de temps de synthèse et de dégradation des protéines, aux concentrations de TopA et de GyrAB (Ropers *et al.*, 2006). Les inégalités de seuil expriment que TopA a deux seuils de concentration, correspondant aux deux seuils de niveau de superenroulement considérés : un premier seuil pour l'activation du gène *fis* et un deuxième seuil, plus élevé, pour l'inhibition de *gyrAB* et l'activation de *topA*. Les inégalités focales expriment que TopA doit pouvoir atteindre des niveaux de concentration tels qu'elle peut réellement jouer un rôle dans ces interactions.

Le logiciel GNA a été utilisé pour simuler la réponse d'une cellule d'*E. coli* à la privation en source de carbone. En partant de conditions initiales représentant la phase exponentielle de croissance, le système est perturbé par un signal de manque de carbone qui active l'adénylate cyclase. La simulation du réseau prend moins d'une seconde sur un PC (2.4 GHz, 512 Mb) et donne un graphe de transition de 65 états qualitatifs. Nombre de ces états sont associés avec des domaines singuliers que le système traverse de façon instantanée. Étant donné que leur pertinence biologique est limitée, ces états peuvent être éliminés. Cela réduit le graphe de transition à 26 états qualitatifs.

Dans tous les états qualitatifs du graphe de transition, le système part d'un état qualitatif correspondant à des conditions de phase exponentielle, pour atteindre un état d'équilibre qualitatif correspondant à des conditions physiologiques trouvées en phase stationnaire (Ropers *et al.*, 2006). Une séquence d'états qualitatifs typique de ce comportement est présentée dans la figure 9. Le premier événement après la réception d'un signal de manque de carbone est la diminution de la concentration de Fis, suivie d'une diminution de la concentration des ARN stables. L'événement suivant concerne l'augmentation du niveau de CRP. En parallèle, la concentration de GyrAB augmente, tandis que celle de TopA reste constante.

Le niveau des ARN stables est un bon indicateur de la croissance cellulaire. Puisque le niveau de ces ARN est bas dans l'état d'équilibre qualitatif atteint par le système après un signal de manque de carbone, nous concluons que cet état est représentatif d'une cellule en phase stationnaire. Le processus conduisant à l'arrêt de

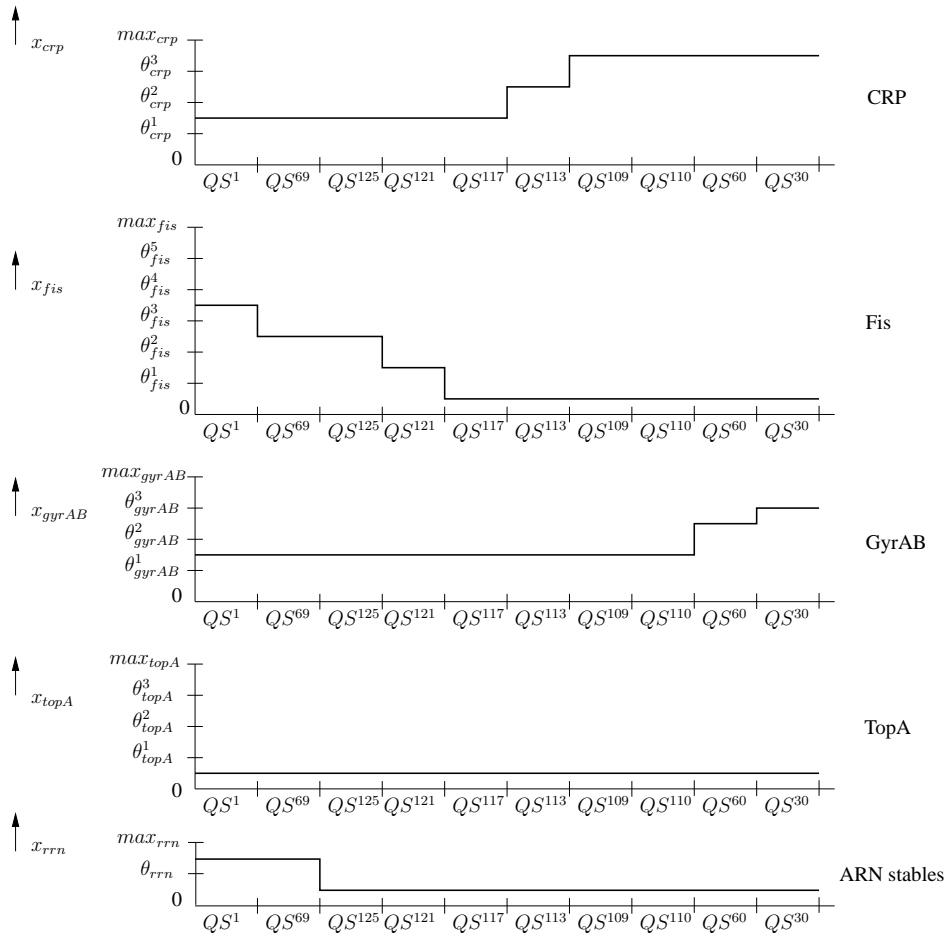


Figure 9. Evolution temporelle d'une sélection de quelques concentrations protéiques dans un comportement qualitatif typique du graphe de transition généré à partir du modèle de réponse au stress nutritionnel chez *E. coli*. Le comportement représente les événements moléculaires accompagnant la transition de la phase exponentielle à la phase stationnaire.

la croissance cellulaire peut être expliquée en reliant le comportement qualitatif au réseau de la réponse au manque de carbone de la figure 8. Lors de la phase exponentielle, l'adénylate cyclase est présente mais inactive. La commutation du signal de manque de carbone active la protéine qui devient capable de produire un métabolite, l'AMPc. La petite molécule se fixe à CRP, qui n'est pas encore abondante, et donne ainsi lieu à une faible concentration du complexe CRP·AMPc. Le niveau du complexe est cependant suffisamment élevé pour commencer à réprimer l'expression de *fis*. Cela provoque une accumulation supplémentaire de CRP et donc, une répression supplémentaire de Fis, par dérépression des promoteurs de *crp* contrôlés par Fis. La diminution de la concentration de Fis provoque l'arrêt de la stimulation de l'expression des gènes *rrn*, en conséquence de quoi le niveau des ARN stables décroît et la cellule entre en phase stationnaire.

Nous concluons de notre modèle du réseau de la réponse au manque de carbone qu'un mécanisme de boucle de rétroaction positive, l'inhibition réciproque de *fis* et de *crp*, joue un rôle clé dans la transition de la phase exponentielle à la phase stationnaire. En présence d'un signal de manque de carbone, il provoque le basculement d'un état avec une concentration élevée de Fis et faible de CRP à un état avec une concentration faible de Fis et élevée de CRP. Le circuit constitué de la boucle de rétroaction positive rend donc la cellule capable d'abandonner la phase exponentielle et d'entrer en phase stationnaire, tout en lui permettant de réentrer en phase exponentielle si une source de carbone est à nouveau disponible. Les boucles de rétroaction positive sont connues plus généralement pour leur rôle important dans les processus de différenciation cellulaire (Thomas *et al.*, 1990).

Les protéines GyrAB et TopA contrôlent le niveau de superenroulement de l'ADN, ce qui contraint sa structure et influence l'expression des gènes (Drlaca, 1990). Suivant la concentration relative des deux enzymes, le superenroulement de l'ADN peut atteindre différents niveaux dans la cellule. En présence de quantités élevées de GyrAB, des tours négatifs sont introduits dans l'ADN, tandis qu'ils sont relâchés lorsque la concentration de TopA est élevée. L'évolution qualitative prédite des concentrations des protéines GyrAB et TopA dans la figure 9 implique que le niveau de superenroulement de l'ADN augmente lors de l'entrée en phase stationnaire. Ce n'est pourtant pas ce qui est observé expérimentalement. Au contraire, il a été montré que le niveau de superenroulement de l'ADN diminue lorsque *E. coli* entre en phase stationnaire (Balke *et al.*, 1987). Cette incohérence entre le niveau de superenroulement prédit et observé suggère que notre représentation du réseau de réponse au stress nutritionnel est incomplète. C'est-à-dire, elle laisse à penser qu'il manque des interactions entre des régulateurs globaux dans la figure 8 ou des régulateurs qui ne sont pas montrés dans la figure. La réalisation d'expériences et l'extension du modèle permettraient d'explorer ces possibilités.

En plus d'une meilleure compréhension de la transition de la phase exponentielle à la phase stationnaire, le modèle a également donné des prédictions sur l'apparition d'oscillations amorties de quelques concentrations protéiques lorsqu'une source de carbone est soudainement disponible (« carbon upshift ») (Ropers *et al.*, 2006). Ces

prédictions, de même que d'autres prédictions sur l'expression des gènes clés du réseau, n'ont jamais été observées expérimentalement. Ceci nous a conduit à débiter des expériences visant à valider les prédictions du modèle. Il s'agit en particulier de mesurer l'évolution temporelle de l'expression des gènes clés du réseau à l'aide de gènes rapporteurs. La comparaison des profils d'expression observés et prédits pourrait conduire à raffiner le modèle, ce qui initierait un nouveau cycle de prédictions et de vérification expérimentale.

6. Discussion

Afin de comprendre comment le fonctionnement et le développement des organismes vivants sont contrôlés, à l'intérieur des cellules ou entre elles, par des réseaux d'interactions entre gènes, protéines et petites molécules, nous avons besoin de méthodes mathématiques et d'outils informatiques. Nous avons insisté en particulier sur le besoin d'approches qualitatives de modélisation, d'analyse et de simulation de réseaux de régulation génique, c'est-à-dire, des approches capables d'inférer des propriétés de la dynamique du système qui soient invariantes pour une large gamme de valeurs de paramètres et de mécanismes réactionnels. L'intérêt pour de telles approches qualitatives découle du fait que, pour la plupart des réseaux d'intérêt biologique, nous ne disposons pas de valeurs numériques précises pour les paramètres et d'informations détaillées sur les mécanismes de réaction (*e.g.*, (Ingolia *et al.*, 2004)). De plus, il est raisonnable de penser que de nombreuses propriétés dynamiques des organismes vivants sont suffisamment robustes pour ne pas être affectées par de petits changements des valeurs de paramètres et des variations des mécanismes réactionnels (Barkai *et al.*, 1997; Eldar *et al.*, 2002; Stelling *et al.*, 2004). Cela ne signifie pas pour autant que les approches qualitatives sont à utiliser en toutes circonstances : il y a en effet des questions biologiques pour lesquelles une précision quantitative est nécessaire et il existe des systèmes pour lesquels une information quantitative détaillée est disponible. En fait, les approches quantitatives et qualitatives devraient être vues comme complémentaires plutôt que mutuellement exclusives.

Dans cet article, nous avons fait une revue de notre travail sur une méthode de modélisation, d'analyse et de simulation qualitative de réseaux de régulation génique qui est bien adaptée aux données expérimentales disponibles à l'heure actuelle et à certaines questions biologiques intéressantes (de Jong *et al.*, 2004b). La méthode est basée sur une classe de modèles d'équations différentielles linéaires par morceaux, introduites à l'origine par Glass et Kauffman (Glass *et al.*, 1973), qui décrit les aspects essentiels de la régulation génique et a des propriétés mathématiques favorables. Nous avons montré que la dynamique qualitative de systèmes LPM peut être représentée par un graphe de transition, constitué d'états qualitatifs et de transitions entre ces états. Le résultat principal sur lequel repose cette méthode est l'invariance du graphe pour une large gamme de valeurs de paramètres, définie par des contraintes d'inégalité qui peuvent être facilement inférées de la littérature. De plus, des règles symboliques simples permettent de calculer le graphe de transition à partir des contraintes d'inéga-

lité. Afin de permettre son application à des réseaux de régulation génique complexes, la méthode a été implémentée dans le logiciel Genetic Network Analyzer (GNA).

L'utilisation des modèles LPM se justifie par l'intuition selon laquelle les gènes peuvent être considérés en première approximation comme des commutateurs (« switches ») logiques qui transforment leurs entrées, par exemple, la concentration en protéines régulatrices, en des sorties discrètes telles que l'état d'activation des gènes (Sugita, 1963; Yuh *et al.*, 1998). Au lieu de développer cette intuition dans des modèles en temps continu et avec des variables de concentration, on peut très bien décider d'utiliser des modèles discrets. Un exemple majeur de cette approche est le formalisme des réseaux booléens (Kauffman, 1969; Kauffman, 1993; Somogyi *et al.*, 1996). L'application des réseaux booléens repose sur l'hypothèse qu'un gène est soit actif, soit inactif, et que les gènes changent leur état d'activation de façon synchrone. Ces hypothèses sont cependant trop fortes si l'on souhaite modéliser des réseaux de régulation génique réels. Pour répondre à ce problème, des formalismes généralisés avec des états d'activation multivalués ont été proposés (Thomas, 1973; Thomas *et al.*, 1990) et appliqués avec succès à l'analyse de réseaux de régulation complexes contrôlant des processus de développement (Mendoza *et al.*, 1999; Sánchez *et al.*, 2003). L'avantage des réseaux booléens et leurs généralisations est qu'ils permettent d'exprimer de façon simple la logique de régulation de l'expression génique. Il leur est cependant difficile de traiter les propriétés dynamiques de réseaux de régulation génique au niveau du seuil d'activation ou d'inactivation d'un gène. L'analyse de la dynamique des modèles LPM dans la section 3 montre pourtant que des points d'équilibre peuvent être localisés sur les seuils. Un résultat similaire, dans cet exemple, est obtenu par l'analyse de modèles d'équations différentielles ordinaires avec des fonctions sigmoïdales raides au lieu des fonctions en escalier, où des points d'équilibre sont retrouvés à proximité des seuils, dans la zone correspondant au point d'inflexion des sigmoïdes.

La méthode présentée dans cette revue préserve la capacité des modèles continus à caractériser la dynamique de systèmes au niveau de leurs seuils, tout en utilisant des abstractions discrètes pour faciliter l'analyse des propriétés dynamiques qualitatives. L'utilisation d'abstractions discrètes pour l'étude des propriétés de systèmes dynamiques a déjà été proposée dans le contexte du raisonnement qualitatif sur des systèmes physiques (de Jong, 2005; Kuipers, 1994; Travé-Massuyès *et al.*, 2003) et l'analyse de systèmes hybrides (Alur *et al.*, 2000; Chutinan *et al.*, 2001; Lygeros *et al.*, 1999). En outre, ces méthodes ont été appliquées à une variété de systèmes modèles procaryotes ou eucaryotes (Alur *et al.*, 2001; Ghosh *et al.*, 2004; Heidtke *et al.*, 1998), certaines applications utilisant les mêmes modèles LPM que dans cet article. Notre méthode emprunte les idées et la terminologie à la fois du raisonnement qualitatif et de l'analyse des systèmes hybrides, mais diffère par un aspect important. Au lieu d'utiliser des méthodes génériques permettant d'analyser une large classe de modèles, nous avons développé une méthode dédiée à l'analyse d'une classe de modèles LPM particulièrement bien appropriés pour les réseaux de régulation génique. De plus, la méthode emploie des algorithmes qui ont été adaptés aux propriétés favorables des modèles LPM, favorisant ainsi la mise à l'échelle de l'analyse à des réseaux complexes.

Au final, il n'y a qu'un seul critère pour évaluer l'utilité d'une méthode pour l'analyse de la dynamique de réseaux de régulation génique : quels résultats d'intérêt biologique la méthode a-t-elle permis d'obtenir ? Jusqu'à présent, notre méthode a été appliquée à un certain nombre de réseaux de régulation génique, tels que l'initiation de la sporulation chez *Bacillus subtilis* (de Jong *et al.*, 2004a), le « quorum sensing » chez *Pseudomonas aeruginosa* (Usseglio Viretta *et al.*, 2004) et la réponse au stress nutritionnel chez *Escherichia coli* (Ropers *et al.*, 2006). La description de cette dernière application dans la section 5 a montré que la méthode de simulation qualitative peut être utilisée pour obtenir des prédictions sur le comportement de réseaux qui ne sont actuellement pas bien compris par les biologistes. Quelques-unes de ces prédictions ont permis de clarifier le rôle de certains mécanismes de régulation (tels que l'inhibition mutuelle de *fis* et *crp*). Par contre, d'autres contredisent les mesures publiées dans la littérature (comme l'augmentation du niveau de superenroulement de l'ADN lors de l'entrée en phase stationnaire) ou concernent des phénomènes qui n'ont pas encore été étudiés expérimentalement (tels que l'apparition d'oscillations amorties lorsqu'une source de carbone est soudainement disponible). Ces deux dernières prédictions sont particulièrement intéressantes d'un point de vue biologique, parce qu'elles soulèvent de nouvelles questions et suggèrent également la réalisation d'expériences.

Les données actuellement disponibles indiquent que le réseau de régulateurs globaux de la transcription impliqués dans la réponse au stress nutritionnel est probablement d'un ordre de grandeur supérieur au réseau présenté dans la figure 8, tandis que la taille du réseau de régulation génique complet d'*E. coli* est même supérieure de plusieurs ordres de grandeur (Babu *et al.*, 2003; Chatterji *et al.*, 2001; Hengge-Aronis, 2000; Ma *et al.*, 2004; Martinez-Antonio *et al.*, 2003; McLeod *et al.*, 2001; Shen-Orr *et al.*, 2002; Thieffry *et al.*, 1998). On peut donc se demander si la méthode présentée dans cet article et plus généralement, les approches qualitatives, peuvent être mises à l'échelle de réseaux bien plus grands que ceux traités jusqu'à présent. La pertinence de cette question repose sur notre expérience de l'analyse des graphes de transition des réseaux de plus de quinze gènes. Ils consistent en général de plusieurs centaines voire plusieurs milliers d'états, ce qui les rend bien trop grands pour être analysés par simple inspection visuelle.

Une façon d'aborder les problèmes posés par la taille des réseaux de régulation génique est d'exploiter les méthodes de vérification formelle. Ces méthodes sont utilisées pour l'analyse automatique et efficace de très grands systèmes de transition discrets rencontrés dans d'autres domaines d'application, comme la vérification de circuits électroniques et de logiciels embarqués. Plus précisément, le simulateur qualitatif pourrait être couplé avec des outils de « model checking » pour permettre la vérification automatique de propriétés dynamiques exprimées en logique temporelle (Clarke *et al.*, 1999; Shults *et al.*, 1997). Afin de réaliser ce couplage, nous avons reformulé le système de transition Σ -QTS, ou plutôt une variante mieux adaptée au type de données disponibles dans notre laboratoire, en une structure de Kripke (Batt *et al.*, 2005a). Cette structure de Kripke peut être exportée de GNA vers des « model checkers » standards, comme NuSMV (Cimatti *et al.*, 2002) et Evaluator, un compo-

sant de CADP (Mateescu *et al.*, 2003), puis analysée au sein de ces outils. Ainsi, nous avons validé le modèle LPM présenté dans la section 5, en vérifiant la cohérence des prédictions avec de nombreuses propriétés de la réponse au manque de carbone chez *E. coli* observées expérimentalement.

Comme nous l'avons montré dans la section 4, Σ -QTS est une approximation conservative de Σ -TS (et donc de Σ). Ceci a pour conséquence que les propriétés dites universelles, c'est-à-dire, devant être vérifiées par tous les chemins dans le graphe de transition, sont préservées. La préservation des propriétés universelles par l'abstraction discrète joue un rôle crucial dans le contexte de la validation de modèles, puisqu'elle permet de garantir qu'un modèle n'est jamais rejeté à tort quand une contradiction entre Σ -QTS et les observations biologiques est trouvée. Par contre, cette garantie n'existe pas pour les propriétés dites existentielles, c'est-à-dire, devant être vérifiées par au moins un chemin dans le graphe de transition. Ces propriétés ne sont pas préservées du fait de la présence possible dans Σ -QTS de chemins qui n'ont pas de correspondants en Σ -TS.

D'autres travaux récents ont confirmé l'intérêt du couplage entre l'analyse qualitative de réseaux biologiques et les techniques de « model checking » (*e.g.*, (Bernot *et al.*, 2004; Chabrier-Rivier *et al.*, 2004)). Ils montrent, par exemple, la diversité de questions biologiquement pertinentes que l'on peut formuler à l'aide de logiques temporelles, ainsi que la possibilité de guider la sélection de modèles par des spécifications en logique temporelle. Toutefois, pour l'analyse de réseaux de taille plus significative, il sera nécessaire d'utiliser des techniques de « model checking » avancées, basées par exemple sur des représentations symboliques plutôt qu'explicitement des graphes de transition (Burch *et al.*, 1990), la minimisation des graphes par équivalence observationnelle (Milner, 1980) ou l'utilisation d'algorithmes de vérification compositionnelle (Kupferman *et al.*, 1998) permettant d'exploiter la structure modulaire des réseaux (voir plus loin). D'un point de vue pratique, il sera également essentiel de développer des environnements informatiques intégrés, avec des interfaces utilisateur adaptées à l'application biologique, permettant des allers-retours rapides entre la formulation du modèle et la vérification de ses propriétés.

Parallèlement au développement d'outils informatiques pour analyser et interpréter les résultats de la simulation qualitative, le défi de la mise à l'échelle demande de réfléchir à l'utilité de construire des modèles toujours plus grands et toujours plus complexes de systèmes cellulaires (Rechenmann *et al.*, 2005). Pour bon nombre de questions biologiques, il n'est pas certain que la prédiction de l'évolution temporelle de la concentration de toutes les protéines dans une cellule d'*E. coli* est la voie la plus sûre pour mieux comprendre la réponse au stress de la bactérie. Dans la plupart des cas, il est plus approprié et informatif de considérer des sous-réseaux du réseau étudié, de décrire individuellement la dynamique de ces sous-réseaux – en utilisant des modèles beaucoup plus abstraits que ceux présentés ici – et de coupler les modèles abstraits de façon à analyser les interactions entre ces sous-réseaux (Bhalla *et al.*, 1999; Kholodenko *et al.*, 2002). Cette approche modulaire se justifie par ce que l'on sait ou soupçonne de la structure des réseaux de régulation biologique (Hartwell

et al., 1999; Thieffry *et al.*, 2004; Wolf *et al.*, 2003). Par exemple, les interactions entre dix à vingt régulateurs globaux de la transcription d'*E. coli* semblent constituer le squelette du réseau de régulation génique de cet organisme, en connectant de façon hiérarchique et en coordonnant le fonctionnement de modules pour assurer des fonctions cellulaires spécifiques (Martinez-Antonio *et al.*, 2003).

Les approches qualitatives pour la modélisation, l'analyse et la simulation de réseaux de régulation génique sont des extensions prometteuses de la boîte à outils du modélisateur, comme l'attestent les résultats obtenus avec la méthode présentée dans cette article ou d'autres méthodes similaires. Leur adoption large passe par la vérification expérimentale des prédictions des propriétés qualitatives de la dynamique du réseau.

Remerciements

Les auteurs remercient les relecteurs pour leurs commentaires sur la première version de l'article. Ils remercient également pour leur soutien financier l'INRIA (ARC, projet GDyn), le Ministère de la recherche (ACI IMPBio, projet BacAttract), ainsi que la Commission européenne (NEST Adventure, projet Hygeia). Cet article est une version traduite et retravaillée d'un chapitre pour le livre *Mathematical and Computational Methods in Biology*, sous la direction d'Elisabeth Pécou, de Salomé Martinez et d'Alejandro Maass, publié dans la série Travaux en cours chez Editions Hermann en 2006.

7. Bibliographie

- Alur R., Belta C., Ivančić F., Kumar V., Mintz M., Pappas G., Rubin H., Schlug J., « Hybrid modeling and simulation of biomolecular networks », in M. D. Benedetto, A. Sangiovanni-Vincentelli (eds), *Hybrid Systems : Computation and Control (HSCC 2001)*, vol. 2034 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, p. 19-32, 2001.
- Alur R., Henzinger T., Lafferriere G., Pappas G., « Discrete abstractions of hybrid systems », *Proceedings of the IEEE*, vol. 88, n° 7, p. 971-984, 2000.
- Arkin A., Ross J., McAdams H., « Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected *Escherichia coli* cells », *Genetics*, vol. 149, n° 4, p. 1633-1648, 1998.
- Babu M., Teichmann S., « Evolution of transcription factors and the gene regulatory network in *Escherichia coli* », *Nucleic Acids Research*, vol. 31, n° 4, p. 1234-1244, 2003.
- Balke V., Gralla J., « Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. », *Journal of Bacteriology*, vol. 169, n° 10, p. 4499-506, 1987.
- Barabasi A., Oltvai Z., « Network biology : Understanding the cell's functional organization », *Nature Reviews in Genetics*, vol. 5, n° 2, p. 101-113, 2004.

- Barkai N., Leibler S., « Robustness in simple biochemical networks », *Nature*, vol. 387, p. 913-917, 1997.
- Batt G., Ropers D., de Jong H., Geiselman J., Mateescu R., Page M., Schneider D., « Validation of qualitative models of genetic regulatory networks by model checking : Analysis of the nutritional stress response in *Escherichia coli* », *Bioinformatics*, vol. 21, n° Suppl 1, p. i19-i28, 2005a.
- Batt G., Ropers D., de Jong H., Geiselman J., Page M., Schneider D., « Qualitative analysis and verification of hybrid models of genetic regulatory networks : Nutritional stress response in *Escherichia coli* », in , M. Morari, , L. Thiele (eds), *Hybrid Systems : Computation and Control (HSCC 2005)*, Lecture Notes in Computer Science, Springer-Verlag, Berlin, 2005b.
- Bernot G., Comet J.-P., Richard A., Guespin J., « Application of formal methods to biological regulatory networks : Extending Thomas' asynchronous logical approach with temporal logic », *Journal of Theoretical Biology*, vol. 229, n° 3, p. 339-348, 2004.
- Bettenbrock K., Fischer S., Kremling A., Jahreis K., Sauter T., Gilles E., « A quantitative approach to catabolite repression in *Escherichia coli* », *Journal of Biological Chemistry*, vol. 281, n° 5, p. 2578-2584, 2005.
- Bhalla U., Iyengar R., « Emergent properties of networks of biological signaling pathways », *Science*, vol. 283, n° 5400, p. 381-387, 1999.
- Botsford J., Harman J., « Cyclic AMP in prokaryotes », *Microbiological Reviews*, vol. 56, n° 1, p. 100-122, 1992.
- Bower J., , Bolouri, H. (eds), *Computational Modeling of Genetic and Biochemical Networks*, MIT Press, Cambridge, MA, 2001.
- Burch J., Clarke E., McMillan K., Dill D., Hwang L., « Symbolic model checking : 10^{20} states and beyond », *Proceedings of the Fifth Annual IEEE Symposium on Logic in Computer Science, LICS'90*, IEEE Computer Society Press, p. 1-33, 1990.
- Casey R., de Jong H., Gouzé J.-L., « Piecewise-linear models of genetic regulatory networks : Equilibria and their stability », *Journal of Mathematical Biology*, vol. 52, n° 1, p. 27-56, 2005.
- Chabrier-Rivier N., Chiaverini M., Danos V., Fages F., Schächter V., « Modeling and querying biomolecular interaction networks », *Theoretical Computer Science*, vol. 325, n° 1, p. 25-44, 2004.
- Chatterji D., Ojha A., « Revisiting the stringent response, ppGpp and starvation signaling », *Current Opinion in Microbiology*, vol. 4, n° 2, p. 160-165, 2001.
- Chutinan A., Krogh B., « Verification of infinite-state dynamic systems using approximate quotient transition systems », *IEEE Transactions on Automatic Control*, vol. 46, n° 9, p. 1401-1410, 2001.
- Cimatti A., Clarke E., Giunchiglia E., Giunchiglia F., Pistore M., Roveri M., Sebastiani R., Tacchella A., « NuSMV2 : An OpenSource tool for symbolic model checking », in , D. Brinksma, , K. Larsen (eds), *14th International Conference on Computer Aided Verification (CAV 2002)*, vol. 2404 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, p. 359-364, 2002.
- Clarke E., Grumberg O., Peled D., *Model Checking*, MIT Press, Boston, MA, 1999.
- Collado-Vides J., , Magasanik, B., , Smith, T. (eds), *Integrative Approaches to Molecular Biology*, MIT Press, Cambridge, MA, 1996.

- Cornish-Bowden A., *Fundamentals of Enzyme Kinetics*, revised edn, Portland Press, London, 1995.
- Coutinho R., Fernandez B., Lima R., Meyroneinc A., « Discrete time piecewise affine models of genetic regulatory networks », *Journal of Mathematical Biology*, vol. 52, n° 4, p. 524-570, 2006.
- de Jong H., « Modeling and simulation of genetic regulatory systems : A literature review », *Journal of Computational Biology*, vol. 9, n° 1, p. 67-103, 2002.
- de Jong H., « Qualitative simulation and related approaches for the analysis of dynamical systems », *Knowledge Engineering Review*, vol. 19, n° 2, p. 93-132, 2005.
- de Jong H., Geiselmann J., Batt G., Hernandez C., Page M., « Qualitative simulation of the initiation of sporulation in *B. subtilis* », *Bulletin of Mathematical Biology*, vol. 66, n° 2, p. 261-299, 2004a.
- de Jong H., Geiselmann J., Hernandez C., Page M., « Genetic Network Analyzer : Qualitative simulation of genetic regulatory networks », *Bioinformatics*, vol. 19, n° 3, p. 336-344, 2003a.
- de Jong H., Geiselmann J., Thieffry D., « Qualitative modelling and simulation of developmental regulatory networks », in , S. Kumar, , P. Bentley (eds), *On Growth, Form, and Computers*, Academic Press, London, p. 109-134, 2003b.
- de Jong H., Gouzé J.-L., Hernandez C., Page M., Sari T., Geiselmann J., « Qualitative simulation of genetic regulatory networks using piecewise-linear models », *Bulletin of Mathematical Biology*, vol. 66, n° 2, p. 301-340, 2004b.
- Drllica K., « Bacterial topoisomerases and the control of DNA supercoiling », *Trends in Genetics*, vol. 6, n° 12, p. 433-437, 1990.
- Edwards R., « Analysis of continuous-time switching networks », *Physica D*, vol. 146, p. 165-199, 2000.
- Edwards R., Siegelmann H., Aziza K., Glass L., « Symbolic dynamics and computation in model gene networks », *Chaos*, vol. 11, n° 1, p. 160-169, 2001.
- Eldar A., Dorfman R., Weiss D., Ashe H., Shilo B.-Z., Barkai N., « Robustness of the BMP morphogen gradient in *Drosophila* embryonic patterning », *Nature*, vol. 419, n° 6904, p. 304-307, 2002.
- Farcot E., « Geometric properties of a class of piecewise affine biological network models », *Journal of Mathematical Biology*, vol. 52, n° 3, p. 373-418, 2006.
- Filippov A., *Differential Equations with Discontinuous Righthand Sides*, Kluwer Academic Publishers, Dordrecht, 1988.
- Gagneur J., Casari G., « From molecular networks to qualitative cell behavior », *FEBS Letters*, vol. 579, n° 8, p. 1867-1871, 2005.
- Ghosh R., Tomlin C., « Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling : Delta-Notch protein signalling », *Systems Biology*, vol. 1, n° 1, p. 170-183, 2004.
- Gillespie D., « A rigorous derivation of the chemical master equation », *Physica D*, vol. 188, p. 404-425, 1992.
- Glass L., Kauffman S., « The logical analysis of continuous non-linear biochemical control networks », *Journal of Theoretical Biology*, vol. 39, n° 1, p. 103-129, 1973.

- Glass L., Pasternack J., « Stable oscillations in mathematical models of biological control systems », *Journal of Mathematical Biology*, vol. 6, p. 207-223, 1978.
- Goodacre R., Vaidyanathan S., Dunn W., Harrigan G., Kell D., « Metabolomics by numbers : Acquiring and understanding global metabolite data », *Trends in Biotechnology*, vol. 22, n° 5, p. 245-252, 2004.
- Gouzé J.-L., Sari T., « A class of piecewise linear differential equations arising in biological models », *Dynamical Systems*, vol. 17, n° 4, p. 299-316, 2002.
- Harman J., « Allosteric regulation of the cAMP receptor protein », *Biochimica et Biophysica Acta*, vol. 1547, n° 1, p. 1-17, 2001.
- Hartwell L., Hopfield J., Leibler S., Murray A., « From molecular to modular cell biology », *Nature*, vol. 402, n° supplement, p. C47-C52, 1999.
- Hasty J., McMillen D., Isaacs F., Collins J., « Computational studies of gene regulatory networks : *In numero* molecular biology », *Nature Review Genetics*, vol. 2, n° 4, p. 268-279, 2001.
- Heidtke K., Schulze-Kremer S., « Design and implementation of a qualitative simulation model of λ phage infection », *Bioinformatics*, vol. 14, n° 1, p. 81-91, 1998.
- Heinrich R., Schuster S., *The Regulation of Cellular Systems*, Chapman & Hall, New York, 1996.
- Hengge-Aronis R., « The general stress response in *Escherichia coli* », in , G. Storz, , R. Hengge-Aronis (eds), *Bacterial Stress Responses*, ASM Press, Washington, DC, p. 161-177, 2000.
- Huisman G., Siegele D., Zambrano M., Kolter R., « Morphological and physiological changes during stationary phase », in , F. Neidhardt, , R. Curtiss III, , J. Ingraham, , E. Lin, , K. Low, , B. Magasanik, , W. Reznikoff, , M. Riley, , M. Schaechter, , H. Umbarger (eds), *Escherichia coli and Salmonella : Cellular and Molecular Biology*, 2nd edn, ASM Press, Washington, DC, p. 1672-1682, 1996.
- Ingolia N., Murray A., « The ups and downs of modeling the cell cycle », *Current Biology*, vol. 14, n° 18, p. R771-R777, 2004.
- Johansson K., Egerstedt M., Lygeros J., Sastry S., « Regularization of Zeno hybrid automata », *Systems and Control Letters*, vol. 38, n° 3, p. 141-150, 1999.
- Kauffman S., « Metabolic stability and epigenesis in randomly constructed genetic nets », *Journal of Theoretical Biology*, vol. 22, p. 437-467, 1969.
- Kauffman S., *The Origins of Order : Self-Organization and Selection in Evolution*, Oxford University Press, New York, 1993.
- Kholodenko B., Kiyatkin A., Bruggeman F., Sontag E., Westerhoff H., Hoek J., « Untangling the wires : A strategy to trace functional interactions in signaling and gene networks », *Proceedings of the National Academy of Sciences of the USA*, vol. 99, n° 20, p. 12841-12846, 2002.
- Kitano, H. (ed.), *Foundations of Systems Biology*, MIT Press, Cambridge, MA, 2001.
- Koch I., Junker B., Heiner M., « Application of Petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber », *Bioinformatics*, vol. 21, n° 7, p. 1219-1226, 2005.
- Kohn K., « Molecular interaction maps as information organizers and simulation guides », *Chaos*, vol. 11, n° 1, p. 84-97, 2001.

- Kuipers B., *Qualitative Reasoning : Modeling and Simulation with Incomplete Knowledge*, MIT Press, Cambridge, MA, 1994.
- Kupferman O., Vardi M., « Modular model checking », *Proceedings of Compositionality : The Significant Difference*, vol. 1536 of *Lecture Notes in Computer Science*, Springer-Verlag, 1998.
- Leloup J.-C., Goldbeter A., « Toward a detailed computational model for the mammalian circadian clock », *Proceedings of the National Academy of Sciences of the USA*, vol. 100, n° 12, p. 7051-7056, 2003.
- Levin M., Carson J., « Fluorescence correlation spectroscopy and quantitative cell biology », *Differentiation*, vol. 72, n° 1, p. 1-10, 2004.
- Lewis J., Glass L., « Steady states, limit cycles, and chaos in models of complex biological networks », *International Journal of Bifurcation and Chaos*, vol. 1, n° 2, p. 477-483, 1991.
- Lockhart D., Winzeler E., « Genomics, gene expression and DNA arrays », *Nature*, vol. 405, n° 6788, p. 827-836, 2000.
- Lygeros J., Pappas G., Sastry S., « An introduction to hybrid system modeling, analysis, and control », 1999, First Nonlinear Control Network (NCN) Pedagogical School, Athens, Greece.
- Ma H.-W., Kumar B., Ditzges U., Gunzer F., Buer J., Zeng A.-P., « An extended transcriptional regulatory network of *Escherichia coli* and analysis of its hierarchical structure and network motifs », *Nucleic Acids Research*, vol. 32, n° 22, p. 6643-6649, 2004.
- Martinez-Antonio A., Collado-Vides J., « Identifying global regulators in transcriptional regulatory networks in bacteria », *Current Opinion in Microbiology*, vol. 6, n° 5, p. 482-489, 2003.
- Mateescu R., Sighireanu M., « Efficient on-the-fly model-checking for regular alternation-free mu-calculus », *Science of Computer Programming*, vol. 46, n° 3, p. 255-281, 2003.
- McAdams H., Arkin A., « Simulation of prokaryotic genetic circuits », *Annual Review of Biophysics and Biomolecular Structure*, vol. 27, p. 199-224, 1998.
- McAdams H., Shapiro L., « Circuit simulation of genetic networks », *Science*, vol. 269, n° 5224, p. 650-656, 1995.
- McLeod S., Johnson R., « Control of transcription by nucleoid proteins », *Current Opinion in Microbiology*, vol. 4, n° 2, p. 152-159, 2001.
- Mendoza L., Thieffry D., Alvarez-Buylla E., « Genetic control of flower morphogenesis in *Arabidopsis thaliana* : A logical analysis », *Bioinformatics*, vol. 15, n° 7-8, p. 593-606, 1999.
- Mestl T., Bagley R., Glass L., « Common chaos in arbitrarily complex feedback networks », *Physical Review Letters*, vol. 79, n° 4, p. 653-656, 1997.
- Mestl T., Plahte E., Omholt S., « A mathematical framework for describing and analysing gene regulatory networks », *Journal of Theoretical Biology*, vol. 176, n° 2, p. 291-300, 1995.
- Mihalcescu I., Hsing W., Leibler S., « Resilient circadian oscillator revealed in individual cyanobacteria », *Nature*, vol. 430, n° 6995, p. 23-24, 2004.
- Milner R., *A Calculus of Communicating Systems*, vol. 92 of *Lecture Notes in Computer Science*, Springer-Verlag, 1980.
- Mockler T., Ecker J., « Applications of DNA tiling arrays for whole-genome analysis », *Genomics*, vol. 85, n° 1, p. 1-15, 2005.

- Monod J., Jacob F., « General conclusions : Teleonomic mechanisms in cellular metabolism, growth, and differentiation », *Cold Spring Harbor Symposium on Quantitative Biology*, vol. 26, Cold Spring Harbor, NY, p. 389-401, 1961.
- Pandey A., Mann M., « Proteomics to study genes and genomes », *Nature*, vol. 405, n° 6788, p. 837-846, 2000.
- Pécou, E., , Martinez, S., , Maass, A. (eds), *Qualitative analysis of the dynamics of genetic regulatory networks using piecewise-linear models*, Editions Hermann, Paris, 2006.
- Phair R., Misteli T., « Kinetic modelling approaches to *in vivo* imaging », *Nature Reviews Molecular Cell Biology*, vol. 2, n° 12, p. 898-907, 2001.
- Plahte E., Kjöglum S., « Analysis and generic properties of gene regulatory networks with graded response functions », *Physica D*, vol. 201, n° 1, p. 150-176, 2005.
- Plahte E., Mestl T., Omholt S., « Global analysis of steady points for systems of differential equations with sigmoid interactions », *Dynamics and Stability of Systems*, vol. 9, n° 4, p. 275-291, 1994.
- Ptashne M., *A Genetic Switch : Phage λ Revisited*, 3rd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2004.
- Rao C., Kirby J., Arkin A., « Design and diversity in bacterial chemotaxis : A comparative study in *Escherichia coli* and *Bacillus subtilis* », *PLoS Biology*, vol. 2, n° 2, p. 0239-0252, 2004.
- Rechenmann F., de Jong H., « Le vivant en équations », *La Recherche*, vol. 383, p. 32-37, 2005.
- Regev A., Silverman W., Shapiro E., « Representation and simulation of biochemical processes using the π -calculus process algebra », in , R. Altman, , A. Dunker, , L. Hunter, , K. Lauderdale, , T. Klein (eds), *Pacific Symposium on Biocomputing, PSB'01*, vol. 6, World Scientific Publishing, Singapore, p. 459-470, 2001.
- Ropers D., de Jong H., Page M., Schneider D., Geiselman J., « Qualitative simulation of the carbon starvation response in *Escherichia coli* », *BioSystems*, vol. 84, n° 2, p. 124-152, 2006.
- Sánchez L., Thieffry D., « Segmenting the fly embryo : A logical analysis of the pair-rule cross-regulatory module », *Journal of Theoretical Biology*, vol. 224, n° 4, p. 517-537, 2003.
- Sauer U., « High-throughput phenomics : Experimental methods for mapping fluxomes », *Current Opinion in Biotechnology*, vol. 15, n° 1, p. 58-63, 2004.
- Shav-Tal Y., Singer R., Darzacq X., « Imaging gene expression in single living cells », *Nature Reviews Molecular Cell Biology*, vol. 5, n° 10, p. 855-861, 2004.
- Shen-Orr S., Milo R., Mangan S., Alon U., « Network motifs in the transcriptional regulation network of *Escherichia coli* », *Nature Genetics*, vol. 31, n° 1, p. 64-68, 2002.
- Shults B., Kuipers B., « Proving properties of continuous systems : Qualitative simulation and temporal logic », *Artificial Intelligence*, vol. 92, n° 1-2, p. 91-130, 1997.
- Simão E., Remy E., Thieffry D., Chaouiya C., « Qualitative modelling of regulated metabolic pathways : Application to the tryptophan biosynthesis in *E. coli* », *Bioinformatics*, vol. 21, n° Suppl. 2, p. ii190-ii196, 2005.
- Smolen P., Baxter D., Byrne J., « Modeling transcriptional control in gene networks : Methods, recent results, and future directions », *Bulletin of Mathematical Biology*, vol. 62, n° 2, p. 247-292, 2000.
- Snoussi E., « Qualitative dynamics of piecewise-linear differential equations : A discrete mapping approach », *Dynamics and Stability of Systems*, vol. 4, n° 3-4, p. 189-207, 1989.

- Somogyi R., Sniegoski C., « Modeling the complexity of genetic networks : Understanding multigenic and pleiotropic regulation », *Complexity*, vol. 1, n° 6, p. 45-63, 1996.
- Stelling J., Sauer U., Szallasi Z., Doyle F., Doyle J., « Robustness of cellular functions », *Cell*, vol. 118, n° 6, p. 675-686, 2004.
- Sugita M., « Functional analysis of chemical systems *in vivo* using a logical circuit equivalent : II. The idea of a molecular automaton », *Journal of Theoretical Biology*, vol. 4, p. 179-192, 1963.
- Szallasi, Z., , Periwal, V., , Stelling, J. (eds), *System Modeling in Cellular Biology : From Concepts to Nuts and Bolts*, MIT Press, Cambridge, MA, 2006.
- Thieffry D., Huerta A., Pérez-Rueda E., Collado-Vides J., « From specific gene regulation to genomic networks : A global analysis of transcriptional regulation in *Escherichia coli* », *BioEssays*, vol. 20, n° 5, p. 433-440, 1998.
- Thieffry D., Sánchez L., « Qualitative analysis of gene networks : Toward the delineation of cross-regulatory modules », in , G. Schlosser, , G. Wagner (eds), *Modularity in Development and Evolution*, University of Chicago Press, Chicago, IL, 2004.
- Thomas R., « Boolean formalization of genetic control circuits », *Journal of Theoretical Biology*, vol. 42, p. 563-585, 1973.
- Thomas R., d'Ari R., *Biological Feedback*, CRC Press, Boca Raton, FL, 1990.
- Travé-Massuyès, L., , Dague, P. (eds), *Modèles et raisonnements qualitatifs*, Hermès, Paris, 2003.
- Tyson J., Chen K., Novak B., « Network dynamics and cell physiology », *Nature Reviews Molecular Cell Biology*, vol. 2, n° 12, p. 908-916, 2001.
- Usseglio Viretta A., Fussenegger M., « Modeling the quorum sensing regulatory network of human-pathogenic *Pseudomonas aeruginosa* », *Biotechnology Progress*, vol. 20, n° 3, p. 670-678, 2004.
- van Kampen N., *Stochastic Processes in Physics and Chemistry*, revised edn, Elsevier, Amsterdam, 1997.
- Wang J., Gilles E., Lengeler J., Jahreis K., « Modeling of inducer exclusion and catabolite repression based on a PTS-dependent sucrose and non-PTS-dependent glycerol transport systems in *Escherichia coli* K-12 and its experimental verification », *Journal of Biotechnology*, vol. 92, n° 2, p. 133-158, 2001.
- Wick L., Egli T., « Molecular components of physiological stress responses in *Escherichia coli* », *Advances in Biochemical Engineering/Biotechnology*, vol. 89, p. 1-45, 2004.
- Wolf D., Arkin A., « Motifs, modules and games in bacteria », *Current Opinion in Microbiology*, vol. 6, n° 2, p. 125-134, 2003.
- Yagil G., Yagil E., « On the relation between effector concentration and the rate of induced enzyme synthesis », *Biophysical Journal*, vol. 11, n° 1, p. 11-27, 1971.
- Yuh C.-H., Bolouri H., Davidson E., « Genomic *cis*-regulatory logic : Experimental and computational analysis of a sea urchin gene », *Science*, vol. 279, p. 1896-1902, 1998.
- Zhu H., Bilgin M., Snyder M., « Proteomics », *Annual Review of Biochemistry*, vol. 72, p. 783-812, 2003.

Grégory Batt a obtenu une maîtrise de biologie moléculaire et cellulaire à l'ENS de Lyon et un DEA en informatique à l'université Joseph Fourier à Grenoble. Il a obtenu une thèse en informatique dans cette même université en 2006, pour des travaux réalisés au sein du projet Helix à l'INRIA Rhône-Alpes. Le sujet de la thèse était la validation de modèles qualitatifs de réseaux de régulation génique à l'aide de techniques de model checking. Il est actuellement post-doctorant au Center for Information and Systems Engineering et au Center for BioDynamics à l'université de Boston (États-Unis).

Hidde de Jong Après des études de gestion, d'informatique et de philosophie des sciences à l'université de Twente (Pays-Bas), Hidde de Jong a obtenu une thèse en informatique (cum laude) en 1998, également à l'université de Twente. Il est entré la même année en tant que chargé de recherche à l'INRIA et a été nommé directeur de recherche en 2005. Au sein du projet Helix, il s'intéresse aux méthodes d'analyse mathématique et informatique de réseaux d'interactions cellulaires, ainsi que leur application aux systèmes biologiques.

Johannes Geiselman a obtenu une thèse en biologie moléculaire à l'université d'Oregon (États-Unis) en 1989. Après un post-doctorat à l'Institut Pasteur à Paris et un poste de professeur associé à l'université de Genève, il est devenu professeur à l'université Joseph Fourier en 1998 où il dirige l'équipe Contrôle de l'expression génique au sein du laboratoire Adaptation et Pathogénie des Microorganismes. Sa formation en biologie moléculaire et en biophysique l'a amené à explorer des approches interdisciplinaires pour l'étude des mécanismes moléculaires impliqués dans le contrôle de l'expression génique chez *Escherichia coli*.

Jean-Luc Gouzé est directeur de recherche à l'INRIA Sophia-Antipolis et responsable du projet Comore. Il a soutenu une thèse sur la modélisation mathématique en neurobiologie à l'université Paris XI en 1983 et il est entré à l'INRIA l'année suivante. Il applique les méthodes de l'automatique (théorie des systèmes) à la modélisation mathématique en biologie. Ses intérêts scientifiques principaux sont la modélisation mathématique en biologie, les biomathématiques, les études qualitatives des systèmes dynamiques, l'étude des systèmes positifs en biologie et l'identification et contrôle de systèmes biologiques.

Michel Page a obtenu une thèse en informatique à l'Institut National Polytechnique de Grenoble en 1991. Il est actuellement maître de conférences à l'université Pierre Mendès-France à Grenoble et enseignant à l'IAE de Grenoble. Il est également membre du projet Helix à l'INRIA Rhône-Alpes et s'intéresse à la modélisation de systèmes d'information, aux techniques de simulation et aux systèmes d'aide à la décision.

Delphine Ropers a obtenu une thèse en biologie moléculaire et cellulaire à l'université Henri Poincaré à Nancy en 2003 au cours de laquelle elle a mené des travaux expérimentaux et de modélisation mathématique chez le virus HIV-1. Elle a ensuite rejoint le projet Helix pour un stage post-doctoral sur la modélisation mathématique des réseaux de régulation génique bactériens. En 2006, elle est devenue chargée de recherche à l'INRIA Rhône-Alpes.

Tewfik Sari est professeur à l'université de Haute Alsace à Mulhouse et membre du Laboratoire de Mathématiques. Il s'intéresse aux aspects asymptotiques et géométriques dans la théorie des

équations différentielles et des systèmes dynamiques, ainsi que leurs applications, principalement en biologie.

Dominique Schneider *Après un doctorat en génétique moléculaire à l'université Henri Poincaré à Nancy en 1993, Dominique Schneider a fait un stage post-doctoral au John Innes Institute à Norwich (Angleterre). Il a obtenu un poste de maître de conférences à l'université Joseph Fourier en 1996 et a été nommé professeur dans la même université en 2004. Il dirige l'équipe Dynamique et évolution de génomes bactériens au sein du Laboratoire Adaptation et Pathogénie des Microorganismes.*