

Les réseaux de régulation biologique : rencontre entre biologie et informatique

Production de mucus chez *P. aeruginosa*

1. Introduction

Pour élucider les principes qui gouvernent la complexité en biologie, la modélisation bioinformatique doit dépasser les explications *ad hoc* et faire émerger de nouveaux concepts abstraits [HUA 01]. Cela impose de comprendre la structure des systèmes biologiques, la dynamique de ces systèmes et les méthodes de mise au point de modèles pour prendre en compte les propriétés observées [KIT 02, WOL 01].

Les réseaux de régulation génétique modélisent les interactions entre objets biologiques (gènes et protéines essentiellement). Les biologistes établissent des modèles moléculaires représentant l'ensemble des relations connues pour un réseau donné, qui correspondent aux gènes et aux protéines régulatrices qui, en se liant avec ces gènes favorisent (effet positif) ou empêchent (effet négatif) leur « expression » c'est-à-dire *in fine* la synthèse des protéines codées par ces gènes. Une même protéine peut être régulatrice de plusieurs gènes et avoir un effet positif ou négatif selon le gène considéré. De plus, une protéine peut réguler son propre gène, on parlera d'autorégulation. On voit qu'il est possible de considérer de tels modèles comme des graphes dont les nœuds représentent ces objets et les arcs leurs interactions (figures 1 et 3). Alors que les biologistes se sont intéressés essentiellement jusqu'à présent à l'aspect statique des relations entre les gènes et les protéines, il est possible à chaque instant d'associer à chaque nœud une valeur numérique qui décrit le niveau de concentration de son produit dans la cellule. Les évolutions temporelles de ces niveaux de concentration constituent la dynamique du système.

Dans les années 1980, René Thomas, après avoir introduit une approche booléenne (entité allumée ou éteinte) qui capture qualitativement la dynamique [THO 78], a montré que la présence de circuits de rétroaction dans ces réseaux avaient une importance capitale pour déterminer leur comportement dynamique [THO 81]. En effet, il a montré qu'une condition nécessaire à l'apparition de multistationnarité est la présence d'un circuit positif (constitué par un nombre pair d'interactions négatives) alors qu'un circuit négatif (nombre impair d'interactions négatives) était une condition nécessaire d'homéostasie, unique état d'équilibre vers lequel le système converge ou autour duquel il oscille. Bien que l'approche discrète ne soit pas obligatoire pour étudier de

tels réseaux, elle est souvent plus favorable, ne serait-ce que parce que les données biologiques obtenues sont rarement quantitatives. Pour lui permettre de décrire les comportements dynamiques dans leur ensemble, René Thomas l’a généralisée à des niveaux multivalués [THO 95].

Cet article s’appuie sur un exemple concret, la régulation de la sécrétion de mucus par la bactérie *P. aeruginosa* [GUE 01]. Cette bactérie, très répandue dans la nature, sécrète ce mucus uniquement lorsqu’elle se trouve dans les poumons des malades atteints de mucoviscidose (ce qui en fait la cause principale de mortalité ou morbidité chez ces malades). Lorsqu’on prélève ces bactéries d’un poumon malade et qu’on les cultive en laboratoire, elles continuent en général à produire ce mucus de manière plus ou moins stable. La majorité de ces bactéries ont subi une mutation que les biologistes ont tendance à interpréter comme la cause de la production de mucus. Cependant, le réseau de régulation qui contrôle la production de mucus a été élucidé et montre qu’une protéine (AlgU) est à la fois le régulateur positif des gènes qui codent les enzymes permettant la synthèse de ce mucus, le régulateur d’un complexe protéique qui inhibe son activité et qu’il s’agit aussi d’un autorégulateur positif (régulateur de sa propre synthèse). Il y a donc une boucle de régulation positive qui rend possible une dynamique à deux états stables (un où la protéine s’exprime peu et l’autre où elle s’exprime beaucoup, entraînant la production de mucus), ce qui correspondrait, du point de vue biologique, au passage épigénétique (stable mais non dû à une mutation) de l’état non mucoïde à l’état mucoïde. *La modélisation dynamique de ce système cherche donc à répondre à la question : le passage à l’état mucoïde des bactéries dans le poumon malade est-il dû à une modification épigénétique ou bien à une mutation ?* (Dans la première hypothèse, la mutation observée serait simplement sélectionnée parce que la présence du complexe inhibiteur est toxique pour la bactérie). Cette question n’a pas qu’un intérêt académique, car des stratégies thérapeutiques nouvelles pourraient découler de l’hypothèse épigénétique. La modélisation peut contribuer à répondre à cette question qui devient formellement : *existe-t-il un modèle de la bactérie compatible avec les résultats biologiques connus et qui présente une « bistationnarité » dans laquelle un état stable produit régulièrement du mucus et l’autre non ?* Le *model checking* y répond positivement (section 4).

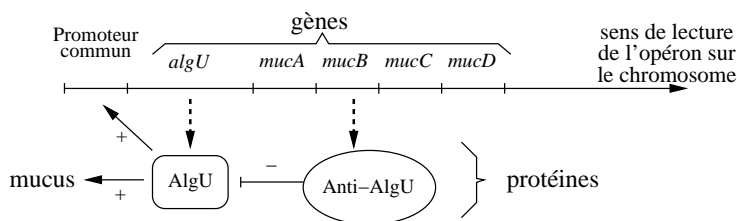


Figure 1. Modèle moléculaire simplifié du réseau de régulation de la production de mucus chez *P. aeruginosa*

Les résultats biologiques connus sont les suivants (figure 1). Le régulateur principal de la production de mucus est le gène *algU*. Il code pour la protéine AlgU qui régule positivement tous les gènes impliqués dans la synthèse du mucus ainsi que l'opéron qui contient outre le gène *algU* lui-même (autorégulation positive), des gènes qui codent pour un complexe inhibiteur de la protéine AlgU. Ce modèle biologique, assez simple, suffit à illustrer l'intérêt des méthodes formelles pour les systèmes biologiques.

2. Réseaux de régulation biologique

La variable x (e.g. un gène) de la figure 2 agit positivement sur y et négativement sur z . Les courbes représentent le niveau de concentration de la protéine produite par y ou z respectivement, en fonction de celui de x (après le délai d'action).

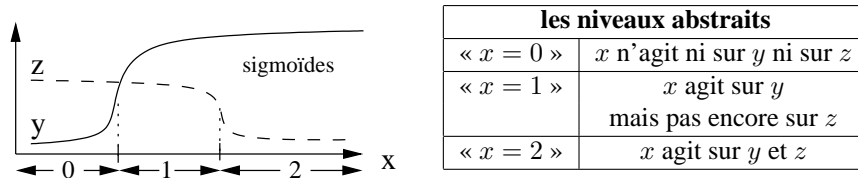


Figure 2. Discretisation des niveaux de concentration d'une variable

On voit que le nombre de niveaux abstraits pertinents d'une variable est entièrement déterminé par les variables sur lesquelles elle agit.

Les réseaux de régulation biologique sont alors représentés par des graphes orientés étiquetés où les nœuds sont les variables. Une interaction entre variables, arc $x \rightarrow y$, est étiquetée par son signe et le niveau abstrait de x à partir duquel elle est effective.

Sur la figure 3 la variable x abstrait le gène AlgU de la figure 1 et sa protéine ; y abstrait *mucB* et l'anti-AlgU, et le troisième nœud abstrait le mécanisme de production de mucus déclenché par x . Les arcs représentent respectivement :

- $x \rightarrow x$: action de AlgU sur son propre promoteur (auto-induction),
- $y \rightarrow x$: effet inhibiteur de l'anti-AlgU sur AlgU,
- $x \rightarrow y$: induction de *mucB* par AlgU (opéron).

On a postulé [GUE 01] que la production de mucus apparaît au second et dernier niveau de concentration de x , mais on ne connaît pas l'ordre des seuils de $x \rightarrow x$ et de $x \rightarrow y$, ce qui conduit à deux graphes de régulation possibles (figures 3a et 3b). Nous avons vu l'importance des cycles ; remarquons ici le cycle positif ($x \rightarrow x$) et le cycle négatif ($x \rightarrow y \rightarrow x$).

Si une variable v reçoit ses arcs entrants des variables $u_1 \cdots u_k$, à chaque instant v tend vers un niveau de concentration déterminé par l'ensemble des u_i qui passent ou

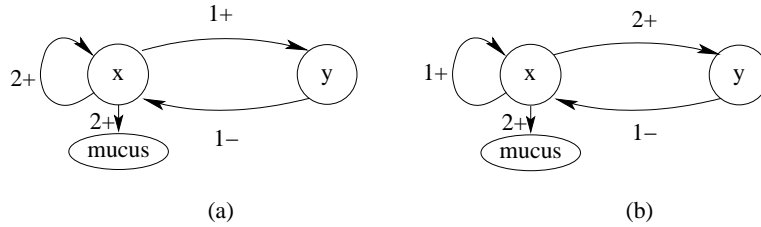


Figure 3. Graphes de régulation possibles pour la mucoïdie chez *P. aeruginosa*

non leur seuil. On note $k_{v,\omega}$ ce niveau « attracteur » de v où ω est le sous-ensemble de $\{u_1, \dots, u_k\}$ formé, d’une part, des inducteurs de v qui passent leur seuil, et d’autre part, des répresseurs de v qui ne le passent pas. ω est l’ensemble des ressources de v à un instant donné. On appelle *réseau de régulation* un graphe de régulation muni des paramètres k_{\dots} précédents. Par exemple, chaque graphe de la figure 3 nécessite six paramètres $k_{x,\{y\}}, k_{x,\{y\}}, k_{x,\{x\}}, k_{x,\{x,y\}}, k_{y,\{x\}}$ et $k_{y,\{x\}}$. Comme x peut prendre trois valeurs et y deux valeurs, $3^4 \times 2^2$ réseaux peuvent être associés à chacun des graphes, soit 648 modèles. Notons que par construction, $k_{mucus,\{y\}} = 0$ et $k_{mucus,\{x\}} = 1$, ce qui permettra d’ignorer le nœud *mucus* dans la suite.

3. Dynamique des réseaux de régulation biologique

Si les variables du réseau sont $\{x_1, \dots, x_p\}$ alors l’ensemble des états possibles (p-uplets des niveaux de concentration des variables) constitue une grille fine de dimension p . Chaque état $(n_{x_1}, \dots, n_{x_p})$ détermine de manière unique pour chaque x_i l’ensemble de ses ressources ω_i (selon les prédécesseurs de x_i passant ou non leur seuil). Ainsi le *graphe d’états synchrone* possède un unique arc sortant de chaque état $(n_{x_1}, \dots, n_{x_p})$ vers l’état attracteur $(k_{x_1,\omega_1}, \dots, k_{x_p,\omega_p})$, voir figure 4 pour un exemple de graphe synchrone issu de la figure 3b avec des paramètres arbitraires.

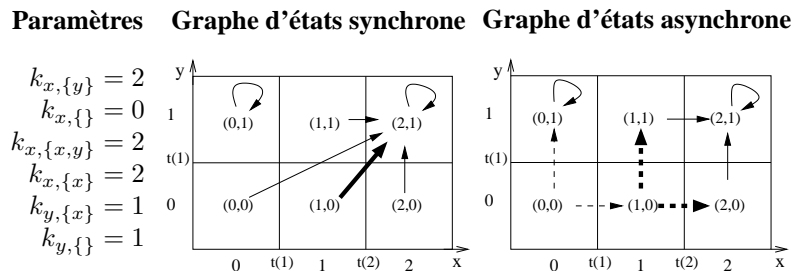


Figure 4. Construction de la dynamique à partir des paramètres

La figure 4 montre également comment construire la véritable dynamique, *asynchrone*, d’un réseau de régulation en suivant deux idées principales [THO 01] :

– en biologie, la probabilité que deux variables passent leurs seuils en même temps est négligeable mais on ne sait pas laquelle est plus rapide. On remplace donc tout arc synchrone diagonal par l'ensemble des arcs qui ne modifient qu'une seule des variables à la fois. Par exemple, $(1, 0) \rightarrow (2, 1)$ est remplacé par les arcs $(1, 0) \rightarrow (2, 0)$ et $(1, 0) \rightarrow (1, 1)$ dans la figure 4 (arcs en gras). Ces deux arcs sont concurrents ;

– comme on modélise un phénomène continu, la cible d'un arc synchrone n'est qu'un attracteur, et la véritable dynamique est reflétée par un arc s'arrêtant sur un état voisin. Dans la figure 4, $(0, 0) \rightarrow (2, 1)$ donne lieu à $(0, 0) \rightarrow (1, 0)$ au lieu de $(0, 0) \rightarrow (2, 0)$.

4. CTL et *model checking*

Souvent le graphe de régulation résulte des connaissances biologiques mais les k_{\dots} sont non mesurables, et constituent de fait l'élément déterminant pour ajuster la dynamique du modèle avec le comportement *in vivo* et en prédire de nouveaux. L'hypothèse épigénétique mentionnée en introduction s'exprime : *existe-t-il des paramètres k_{\dots} pour un des graphes de la figure 3 pour lesquels le graphe d'états asynchrone présente deux états stables, l'un produisant du mucus de manière récurrente, l'autre jamais ?*

Ce type de propriétés peut s'exprimer en CTL (Computation Tree Logic [EME 90]) avec des atomes de la forme $(v = n)$ où v est une variable du réseau et n un naturel. Parmi les modalités de CTL, rappelons que $AX\varphi$ signifie que φ est vraie dans tous les états suivants, $AF\varphi$ dans un état futur, et $AG\varphi$ pour tous les états futurs, voir [HUT 00].

Concernant l'hypothèse épigénétique sur *P. aeruginosa*, nous avons d'abord prouvé que pour tous les modèles « mucus = 1 de manière récurrente » est équivalent à « $x = 2$ de manière récurrente ». Les deux états stables s'écrivent alors :

$$\varphi_1 : (x = 2) \Rightarrow AX AF(x = 2) \quad \text{et} \quad \varphi_2 : (x = 0) \Rightarrow AG(\neg(x = 2))$$

où $AX AF(x = 2)$ signifie que pour tous les futurs possibles, sans prendre en compte le présent, l'état muqueux réapparaîtra à un moment donné.

Une fois passée la difficulté d'écrire correctement en CTL les propriétés biologiques, le *model checking* identifie l'ensemble des modèles qui les satisfont [HUT 00]. Notre environnement de manipulation de réseaux biologiques SMBioNet montre que huit modèles de *P. aeruginosa* parmi les 648 survivent à ces deux formules.

On sait déjà que φ_2 (état non mucoïde) est satisfaite *in vivo*, donc φ_1 constitue la clef du problème. Sa syntaxe suggère un plan d'expérience biologique : porter x à saturation par un signal extérieur à la bactérie, et après la phase transitoire due à ce pulse, voir si la production de mucus persiste [GUE 01].

NOTE. – Des outils automatiques de génération de tests fonctionnels sont capables d'engendrer ce plan d'expérience. Nous avons également montré que le succès de

cette expérience (en cours au Laboratoire de Microbiologie du Froid, Rouen) suffit à prouver l'hypothèse épigénétique.

5. Conclusion

L'interdisciplinarité exige que chaque discipline trouve un intérêt scientifique à la collaboration. Dans un premier temps, une étape de modélisation non formelle appuyée par la théorie selon laquelle la multistationnarité nécessite un circuit de rétroaction positif, a rendu plausible une hypothèse biologique novatrice (initiation épigénétique de la mucoidie chez *P. aeruginosa*). Dans un deuxième temps, les méthodes formelles en ont démontré la cohérence par *model checking*. Dans un troisième temps, elles ont validé les expériences biologiques nécessaires pour confirmer ou infirmer l'hypothèse. L'objectif de l'atelier *observabilité* de genopole[®] est de mieux articuler modélisation et expérimentation en fournissant aux biologistes des instruments de validation des modèles, et en élargissant à la biologie la portée de théories issues de la recherche en informatique. Les modèles ainsi conçus ne sont pas seulement des tentatives d'explications *a posteriori* de résultats de la biologie, mais des guides pour les expérimentations biologiques dont le succès sera *in fine* le critère déterminant.

CTL n'est pas le seul cadre formel pouvant supporter cette démarche pluridisciplinaire. La relative simplicité du système de production de mucus chez *P. aeruginosa* en fait un exemple bien calibré pour tester d'autres approches formelles et celui-ci pourrait servir de « benchmark » par exemple dans le cadre du groupe AFADL.

Remerciements

Les nombreuses discussions au sein de l'atelier *observabilité* de genopole[®] à Evry sont à l'origine de notre collaboration. Ces ateliers ont été rendus possibles grâce à H. Pollard et P. Tambourin. Nous remercions également S. Pérès, V. Bassano et A. Richard pour leurs participations aux logiciels développés. Mentionnons enfin que les expériences biologiques liées à ces travaux sont réalisées à Rouen en coopération avec une équipe du CHU de Grenoble et financées par l'association « vaincre la mucoviscidose ».

6. Bibliographie

- [EME 90] EMERSON E., « *Handbook of theoretical computer science, Volume B : formal models and semantics* », chapitre Temporal and modal logic, p. 995–1072, MIT Press, 1990.
- [GUE 01] GUESPIN-MICHEL J., KAUFMAN M., « Positive feedback circuits and adaptive regulations in bacteria. », *Acta Biotheor.*, vol. 49, n° 4, 2001, p. 207-218.
- [HUA 01] HUANG S., « Genomics, complexity and drug discovery : insights from Boolean network models of cellular regulation. », *Pharmacogenomics.*, vol. 2, n° 3, 2001, p. 203-222.

- [HUT 00] HUTH M., RYAN M., *Logic in Computer Science : Modelling and reasoning about systems*, Cambridge University Press, 2000.
- [KIT 02] KITANO H., « Looking beyond the details : a rise in system-oriented approaches in genetics and molecular biology. », *Curr. Genet.*, vol. 41, n° 1, 2002, p. 1-10.
- [THO 78] THOMAS R., « Logical analysis of systems comprising feedback loops. », *J. Theor. Biol.*, vol. 73, n° 4, 1978, p. 631-656.
- [THO 81] THOMAS R., « On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations », *Springer Ser. Synergetics*, vol. 9, 1981, p. 180-193.
- [THO 95] THOMAS R., THIEFFRY D., KAUFMAN M., « Dynamical behaviour of biological regulatory networks - I. », *Bull. Math. Biol.*, vol. 57, n° 2, 1995, p. 247-276.
- [THO 01] THOMAS R., KAUFMAN M., « Multistationarity, the basis of cell differentiation and memory. I. & II. », *Chaos*, vol. 11, 2001, p. 170-195.
- [WOL 01] WOLKENHAUER O., « Systems biology : the reincarnation of systems theory applied in biology ? », *Brief Bioinform.*, vol. 2, n° 3, 2001, p. 258-270.

Janine Guespin
Laboratoire de Microbiologie du Froid
Université de Rouen
Janine.Guespin@univ-rouen.fr

Gilles Bernot, Jean-Paul Comet
LaMI, Université d'Évry
{bernot, comet}@lami.univ-evry.fr