

Modélisation, observabilité et expérimentation : étude d'un cas.

- Boucle de rétroaction positive dans un réseau de régulation génétique -

G. Bernot*, J. Michel-Guespin†, A. Zemirline‡, J.P. Comet*, F. Delaplace*, P. Ballet‡, P. Amar*§

Autrans 17 - 21 Mars 2002

Résumé

Nous proposons une méthodologie interdisciplinaire de modélisation biologique qui s'inspire de celle pratiquée pour la conception de grands systèmes informatiques.

Savoir si un modèle pour la biologie peut être entièrement validé par un ensemble d'expériences apparaît comme une contrainte naturelle et nécessaire à sa définition. Définir un modèle doit donc s'accompagner de méthodes et de conditions expérimentales aptes à le valider ou l'invalider. En s'inspirant de la conception de grands logiciels et plus généralement de systèmes informatisés, on distinguera deux étapes de conception (modélisation) :

- spécifier, c'est-à-dire construire un modèle rigoureux du comportement souhaité d'un système informatique futur.
- vérifier *a posteriori* si un système répond à ses spécifications c'est-à-dire au comportement souhaité tel que le décrit le modèle préalablement construit.

Nous souhaitons expérimenter, au travers de la modélisation de la production de mucus par la bactérie *Pseudomonas aeruginosa*, l'application de cette méthode de travail.

Référence de l'article : Séminaires Modélisation et simulation de processus biologiques dans le contexte de la génomique, atelier Observabilité, du 17 au 21 Mars 2002 à AUTRANS (Isère)

1 Introduction

Le terme « modèle » est investi par les biologistes d'un très grand nombre de significations. Même si l'on précise modèle mathématique, on est encore loin de l'unicité du sens.

*Eq. BioInfo, Laboratoire des Méthodes Informatiques, U.M.R C.N.R.S 8042, Univ. Evry

†Laboratoire de Microbiologie du Froid et GR en Biologie Intégrative et Modélisation, Univ. de Rouen

‡EA 2215, Dept Informatique, Univ. de Bretagne Occidentale

§Laboratoire de Recherche en Informatique, U.M.R C.N.R.S 8623, Univ. Paris-Sud, Orsay

Une des difficultés provient de l'interdisciplinarité contenue déjà au niveau de l'expression « un modèle mathématique en biologie ». Qui fait le modèle, qui s'en sert, et avant tout à quoi sert-il ? La plupart des modèles mathématiques en biologie sont faits par des mathématiciens (biomathématiciens), des physiciens, ou des informaticiens (bioinformaticiens). Le modélisateur utilise des données de la littérature (qui ont donc été obtenues dans un autre but). Le modèle résultant, publié dans des revues qui ne sont pas lues par les biologistes, est généralement destiné à expliquer ou rendre compte de certains phénomènes biologiques connus. Au pire, ils remplacent une description phénoménologique par une expression mathématique. Mais ils peuvent aussi donner (ou plutôt suggérer) un nouveau cadre explicatif au processus biologique étudié. Si des biologistes viennent à entendre parler de ce dernier type de modèle, ils peuvent s'y intéresser, mais le plus souvent ne voient pas tellement en quoi ils peuvent être impliqués. Le modèle leur apparaît souvent comme gratuit, et peut susciter toute une gamme de réactions, du rejet violent à l'intérêt poli.

Inversement, et tout aussi révélateur du manque de connaissances interdisciplinaires, certains biologistes pensent qu'en apportant les données qu'ils ont obtenus dans tel ou tel contexte à un modélisateur, celui-ci va les « mettre dans sa moulinette » et va lui sortir l'explication, quand ce n'est pas un outil prédictif thérapeutique... L'interdisciplinarité nécessite non seulement l'apprentissage du travail en commun, mais aussi la mise au point en commun des outils utilisables. Elle exige en outre que chaque partie trouve un intérêt scientifique à la collaboration, autrement dit que celle-ci enrichisse les deux disciplines.

Le cas que nous allons étudier ici met en jeu une imbrication étroite entre deux théories mathématiques et un processus biologique donnant lieu à expérimentations : une première étape de modélisation (déjà publiée), appuyée sur la théorie de la multistationnarité, a rendu plausible une hypothèse novatrice. Une deuxième étape mathématique appuyée sur la logique des systèmes, devrait permettre de déterminer les expériences (biologiques) nécessaires pour valider ou infirmer l'hypothèse. *C'est cette deuxième étape qui fait l'objet du travail de l'atelier.* Son but, en extrapolant les résultats à d'autres systèmes, est de *créer un nouvel outil emprunté à l'informatique* permettant d'une part de mieux impliquer les biologistes dans la modélisation en leur fournissant un instrument de validation des modèles, et d'autre part d'élargir la portée d'un instrument déjà utile en informatique. Cette démarche nécessite évidemment une étroite collaboration entre biologistes et modélisateurs. Le modèle ainsi conçu n'est pas seulement une tentative d'explication *a posteriori* de résultats de la biologie, mais un guide pour l'expérimentation biologique, qui sera *in fine* le critère déterminant.

C'est ici que l'analogie entre la spécification de systèmes informatiques et la modélisation en biologie entre en jeu. En informatique, la conception de systèmes informatisés requiert de :

- spécifier, c'est-à-dire construire un modèle rigoureux du comportement souhaité d'un système informatique futur
- vérifier *in fine* si un système répond à ses spécifications c'est-à-dire au comportement souhaité tel que décrit par le modèle théorique construit préalablement.

Cette dernière activité repose essentiellement sur des méthodes sophistiquées de test de logiciel, fondées sur des théories d'extraction de tests à partir de modèles. L'objectif est alors de proposer un ensemble d'expérimentations sur le logiciel réceptionné qui suffise à prouver par extrapolation que le logiciel testé possèdera toujours un comportement compatible avec

son modèle.

Dans ce cadre, les notions d'opérabilité et d'observabilité jouent un rôle majeur :

- *l'opérabilité* est la capacité à faire s'exécuter (pour les tester) des portions de code interne parfois mises en action dans des configurations rares, complexes ou très spécifiques ; c'est aussi la capacité à faire modifier par le logiciel des variables enfouies au beau milieu de l'ensemble de données souvent gigantesque géré par le logiciel à un instant donné ; tout cela en n'utilisant que les opérations autorisées par l'interface du système sous test.
- *l'observabilité* est la capacité à rendre visible l'effet produit par les manipulations précédentes, dans le but de vérifier leur correction par rapport au modèle de comportement souhaité.

Certains modèles/logiciels ne sont pas testables, par manque d'opérabilité ou d'observabilité. Savoir si un modèle peut être validé par un ensemble d'expériences (de tests) de taille raisonnable est une étape nécessaire à la conception d'un logiciel. Le lecteur peut aisément transcrire cet argumentaire dans le cadre de la modélisation en biologie :

- *opérabilité* : à quoi servirait un modèle par trop détaillé du mode de fonctionnement d'une entité biologique si aucune expérience biologique ne peut conduire à faire s'effectuer spécifiquement ces détails de fonctionnement ?
- *observabilité* : à quoi servirait une expérience qui ne permettrait pas d'observer un comportement révélateur ?

Certains modèles mathématiques pour la biologie apportent fort peu, par manque d'opérabilité ou d'observabilité. Savoir si un modèle pour la biologie peut être validé par un ensemble d'expériences biologiques de coût raisonnable est un préalable nécessaire pour proposer un tel modèle.

Une bonne démarche scientifique impose donc qu'un modèle en bio-informatique soit systématiquement accompagné des méthodes/conditions expérimentales aptes à le valider ou l'invalider. En s'inspirant des théories développées en informatique pour la validation et la vérification de logiciels nous souhaitons expérimenter, au travers du cas que nous abordons ici (section 4), une nouvelle méthode de travail interdisciplinaire pour la modélisation en biologie.

2 Première étape

La théorie sur laquelle est basée l'ensemble du travail de la première étape est celle élaborée par René Thomas [1, 2] et ses collaborateurs, qui montre qu'une boucle de rétroaction positive est une condition nécessaire pour générer une bistabilité. Dans le cours *Epigenesis, multistationarity and positive feedback circuits* [3], cette théorie et ses conséquences en biologie sont présentées, ainsi que la bibliographie correspondante.

Le système biologique choisi est celui de la production de mucus par la bactérie *Pseudomonas aeruginosa*, dans des conditions très particulières qui sont celles des poumons de malades atteints de mucoviscidose. Dans des conditions normales, cette bactérie, qui est très fréquente dans le poumon, ne produit pas ce mucus. Cependant, si on isole des bactéries,

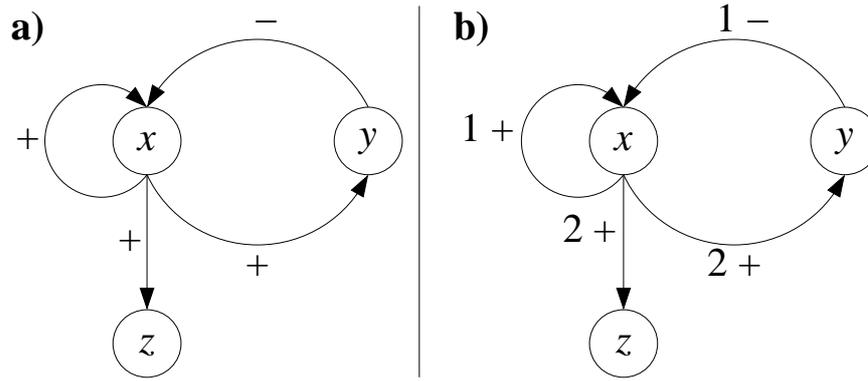
dites mucoïdes, du poumon d'un malade, et qu'on les cultive en laboratoire dans les conditions habituelles, elles continuent de produire du mucus, de façon plus ou moins stable. Les généticiens ont pu montrer, dans le cas des souches dites mucoïdes stables, que c'était la mutation d'un gène qui rendait ces bactéries capables de produire le mucus dans n'importe quelles conditions. L'étude précise des mécanismes moléculaires mis en jeu dans la régulation de la production de mucus, montre que la molécule régulatrice fondamentale (AlgU) est un facteur sigma (cofacteur de la RNA polymérase, permettant la reconnaissance de sites de transcription spécifiques, donc de gènes particuliers), qui est nécessaire à la transcription des gènes impliqués dans la synthèse de mucus. Les mutants mucoïdes sont affectés dans un inhibiteur du facteur sigma AlgU. Le rôle du poumon malade serait donc de favoriser la sélection de mutants mucoïdes. Tout paraît cohérent. Cependant, la protéine AlgU est également nécessaire à sa propre transcription. Il s'agit donc d'une boucle de régulation positive, qui permet de poser la question : et si l'état mucoïde était un état stationnaire alternatif à l'état non mucoïde, déclenché par un signal dans le poumon malade et auto-entretenu lorsque la quantité de facteur sigma spécifique est assez élevée? (la sélection des mutants étant alors un phénomène second, se produisant dans une deuxième étape). L'intérêt de cette hypothèse réside, outre dans la connaissance fondamentale, dans les perspectives thérapeutiques nouvelles qu'elle est susceptible d'ouvrir si elle est vraie.

Un modèle extrêmement simplifié a alors été construit, en tenant compte uniquement des interactions impliquées dans des circuits de rétroaction. Le modèle (Cf. figure 1) contient 3 variables, notées x , y et z . La variable x correspond à la protéine AlgU, la variable y est une abstraction des protéines ayant un effet inhibiteur (antisigma) sur la protéine AlgU, et la variable z est une variable virtuelle représentant la synthèse du mucus. Les 4 arcs¹ représentent : l'auto-régulation de la variable x (arc $x \rightarrow x$), la transcription des gènes impliqués dans la synthèse de mucus (arc $x \rightarrow z$), l'effet inhibiteur des antisigmas sur la protéine AlgU (arc $y \rightarrow x$), et enfin, l'influence positive de AlgU sur la transcription des gènes codant ses propres inhibiteurs (arc $x \rightarrow y$).

Il existe deux circuits de rétroaction, une boucle positive et un circuit négatif, impliquant la molécule AlgU. Il faut insister ici sur le fait que ce modèle est fortement simplifié en raison même de la théorie initiale, qui établit que seuls les circuits de rétroaction sont susceptibles de générer des multistationnarités. Les autres interactions régulatrices sont démontrées d'une importance secondaire et sont donc intentionnellement négligées. Ce modèle peut être étudié par des systèmes d'équations différentielles, ou par l'analyse discrète. C'est cette méthode qui a été choisie et dont l'utilisation est expliquée de façon détaillée dans la publication suivante [4].

En résumé, dans le cas où la variable x influe sur la variable y , la courbe de niveau de y en fonction du niveau de x est une sigmoïde. Cette sigmoïde définit un seuil $S_{(x,y)}$ (Cf. figure 2-a). De la même manière, l'influence de la variable x sur une autre variable z définit un seuil $S_{(x,z)}$ (Cf. figure 2-b). Les deux seuils ($S_{(x,y)}$ et $S_{(x,z)}$), en général différents, définissent à leur tour trois comportements différents de la variable x : soit le niveau de x est inférieur aux deux seuils, soit il est entre les deux, soit supérieur (Cf. figure 2-c). La prise en compte

¹les flèches de la figure 1



Note : la figure de gauche (a) indique l'activation (+) ou d'inhibition (-) exercés entre les différents gènes (resp. protéines) du réseau. A droite (b), figurent sur chaque arc la valeur minimale de seuil pour laquelle cette action (activation ou inhibition) s'exerce. Il s'agit d'une valuation possible. Les variables signifient : x protéine AlgU, y facteurs antisigma, z synthèse du mucus

FIG. 1 – Réseau génétique régulant la production de mucus par une bactérie *Pseudomonas aeruginosa* (modèle simplifié)

des seuils permet d'étudier de manière abstraite l'évolution d'un tel système avec des valeurs discrètes pour les niveaux d'expression.

Dès lors, les seuils correspondent à des valeurs d'interactions entre variables. Chacune de ces valeurs étiquette l'arc du graphe correspondant. On notera sur le graphe $x \xrightarrow{2+} z$ puisque l'on pense que le facteur sigma doit s'exprimer à un niveau élevé (au dessus d'un seuil 2) pour qu'il puisse déclencher la transcription des gènes impliqués dans la synthèse de mucus (Cf. figure 1). Lorsqu'un arc est franchissable, c'est-à-dire lorsque les conditions sur le niveau d'expression de la variable régulatrice sont vérifiées, il reste à décrire comment le système va évoluer. Autrement dit, il faut expliciter le niveau atteint par une variable v en fonction de la présence ou absence des autres variables qui influent sur elle (pour une description précise, voir article [4]). Ces valeurs sont représentées par une fonction notée κ . Dans l'exemple précité, deux cycles cohabitent. Le premier (boucle de rétroaction positive $x \rightarrow x$) impose l'existence de deux états stables : le cas d'auto-entretien de x à un niveau élevé et le cas où le niveau de x vaut toujours 0. L'autre boucle de rétroaction négative va faire basculer le système vers l'un ou l'autre des états stables selon sa force (i.e. selon les valeurs de la fonction K).

L'étude mathématique du modèle fait ressortir que l'hypothèse épigénétique est plausible puisqu'il existe des conditions, non aberrantes d'un point de vue biologique, qui conduisent à des états stationnaires dont les propriétés sont précisément celles du système étudié. Mais, ils disent d'avantage encore. Par exemple, ils prédisent que, si l'hypothèse est vraie, un pulse de production de AlgU peut faire acquérir à la bactérie le phénotype mucoïde de façon relativement stable. Le modèle est donc prédictif en même temps qu'explicatif. Le problème qui se pose alors est : si expérimentalement on réussit une telle expérience, sera-t-elle la preuve que l'hypothèse épigénétique est vraie ? Si, en revanche, cette expérience est négative,

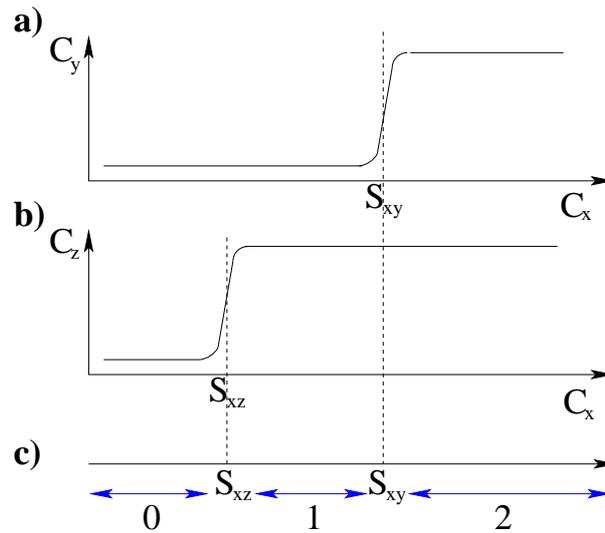


FIG. 2 – Influence de la variable x sur les variables y, z

aura-t-on définitivement écarté cette hypothèse ? C'est là qu'intervient la deuxième étape qui est justement l'objet d'intérêt principal de cet atelier .

3 Modélisation et observabilité

L'activité de modélisation en sciences de la vie consiste, comme en sciences de l'ingénieur, à élaborer à partir d'un ensemble d'observations biologiques nécessairement fini une représentation mathématique d'un phénomène biologique, exprimant un ensemble de comportements *a priori* infini (infinité des valeurs possibles des paramètres des conditions d'expérience, infinité des scénarii possibles de simulation, infinité des valeurs à choisir pour les paramètres du modèle, etc ...). Pour qu'elle soit efficace, une activité de modélisation doit être en mesure d'apporter des réponses à des questions aussi délicates et diverses que les suivantes : Pourquoi tel modèle est de qualité et en quoi est-il meilleur que tel autre ? Étant donné deux modèles qui ne contredisent aucune observation effective mais qui possèdent certains comportements internes différents, quel est le pouvoir explicatif de chacun d'eux ?

Modéliser un phénomène pour bien le comprendre implique nécessairement une abstraction, donc l'approximation de détails et la réduction du nombre d'objets interagissant. Par conséquent, il n'est pas raisonnable d'imaginer qu'un modèle puisse représenter exactement une réalité dont la complexité biologique est bien trop riche, dont le fonctionnement interne est très mal connu et dont les interactions internes et externes, à toutes les échelles, sont non prévisibles. De plus, la plupart des lois élémentaires du modèle et leurs paramètres ne sont pas déterminés *in vivo*, mais sont souvent extrapolés d'un organisme à un autre, et parfois même fixés approximativement. Ainsi, tout modèle d'un phénomène biologique est *faux par construction*. Quel est alors l'intérêt de modéliser un phénomène complexe lorsque le pouvoir de prédiction du modèle est à remettre systématiquement en cause par le seul et élémentaire doute scientifique ? Par ailleurs, qu'un modèle réponde aux comportements pour lesquels on

l'a construit ne prouve rien ; il peut *a priori* échouer face à toute expérimentation biologique future qui mettrait à jour un nouveau comportement. De plus, après remise en cause d'un dogme, comment éviter que le nouveau modèle ait pour effet implicite de remplacer un dogme par un autre ? Certes, ces interrogations montrent que l'action de modéliser impose d'évaluer les limites de toute modélisation. Néanmoins, elles ne doivent pas dissuader le biologiste de faire usage de cet outil qui a fait ses preuves et a acquis une notoriété majeure dans les sciences de l'ingénieur, après bien des échecs, des erreurs, des doutes et un grand nombre de tentatives infructueuses. La construction d'un modèle doit d'abord obéir au principe de Karl Popper, selon lequel la science progresse par conjectures et réfutations [5]. Le critère de scientificité d'une hypothèse est sa falsifiabilité.

Ainsi, si l'on détecte un comportement interne au modèle, non cohérent ou pas en accord avec une expérience biologique, alors on peut remettre en cause le modèle et par là même l'améliorer. Certains comportements d'un modèle sont des conséquences logiques directes de phénomènes connus, biologiquement validés. D'autres sont parfois des extrapolations, introduites par le modélisateur dans le but de construire un modèle qui soit assez général et pas seulement une énumération de cas particuliers connus, ou dans le but de tester une hypothèse scientifique non accessible de façon intuitive. Donc le fait de modéliser implique celui de faire des hypothèses et des conjectures sur le comportement du modèle. Ces hypothèses sont les points faibles des modèles qu'il faut tenter systématiquement de valider, en les attaquant par des expérimentations biologiques ou des simulations bien choisies. Mais elles peuvent aussi être la source d'avancées majeures, permettant de comprendre mieux le processus et d'en mieux prédire les propriétés. Ainsi, même lorsque les seuls résultats produisant une avancée sont des résultats négatifs mettant en évidence des erreurs d'un modèle, il n'en reste pas moins qu'un modèle est d'autant plus intéressant et valable qu'il résiste à des attaques qui tentent de le remettre en cause.

Une bonne modélisation doit donc non seulement expliciter ces hypothèses mais aussi les conditions et les protocoles d'observation des comportements du modèle. Car, tout modèle s'avérerait faux sitôt qu'on augmente le pouvoir d'observation. En effet, bien des propriétés fausses paraissent intrinsèquement irréfutables par manque d'observabilité ! Mais l'observation des comportements d'un modèle est limitée, encore plus en biologie où certaines propriétés ayant pourtant une importance explicative forte ne peuvent être observées. Par conséquent, il faut tenter systématiquement de produire, pour les hypothèses et les conjectures liées au modèle, des conséquences observables et en choisir de judicieuses pour faire les expérimentations, c'est-à-dire celles qui optimisent les chances de réfutation. Une observabilité efficace doit avoir pour objectif la remise en cause de tout modèle, en mettant en évidence ses incohérences internes, en lui trouvant des contradictions avec un autre modèle qui lui est *a priori* compatible, ou encore en exhibant ses comportements qui diffèrent de ceux de la réalité biologique. Après avoir formalisé un modèle et précisé ses hypothèses et ses propriétés intrinsèquement observables, on peut théoriquement, au moyen d'un programme informatique s'appuyant sur la logique formelle, déterminer le choix des expérimentations de réfutation, évaluer le niveau de testabilité de certaines hypothèses, assurer un niveau de couverture ou de complétude de certaines réfutations, pointer des cas génériques non couverts, etc., en s'inspirant et en s'appuyant sur des méthodes et des techniques de test de logiciel. Une forte similitude existe en effet entre le test de logiciel et la réfutation de modèle. Car,

1. L'action de modéliser revient à extrapoler un nombre fini d'expérimentations biologiques et se convaincre que le modèle simule la réalité ; et celle de tester un logiciel revient à extraire d'un nombre fini de tests une certaine confiance en sa conformité avec les spécifications.
2. Les hypothèses de modélisation consistent à supposer certaines régularités ou uniformités des comportements réels ; et les hypothèses de test de logiciel consistent à considérer que chaque test représente tout un domaine de tests partageant son comportement.
3. L'observabilité biologique est limitée par l'instrumentation, l'éthique, etc. ; et l'observabilité d'un logiciel est limitée aux périphériques de sortie (écran) et aux seuls états accessibles en test.

Un test de logiciel procède en sélectionnant d'abord des scénarii de tests potentiels (deux méthodes de test complémentaires sont utilisées : les tests fonctionnels qui portent sur la spécification et qui testent ce que le logiciel est supposé réaliser ; et les tests structurels qui tirent partie de la structure du logiciel et qui testent ce que le logiciel accomplit), ensuite en exécutant les tests retenus et enfin en dépouillant ces tests pour déterminer les parties du logiciel qui doivent être remises en cause et ainsi apporter des corrections au logiciel. L'évaluation de la qualité d'un modèle représentant le phénomène biologique observé trouvera certainement un intérêt à opérer d'une façon analogue au test de logiciel, où le niveau de spécification est le niveau de description biologique. La testabilité des hypothèses d'un modèle nécessite leur accessibilité et leur observabilité et donc la mise en place des conditions d'expériences et/ou de simulation. L'appréciation des résultats des tests subis par le modèle est intimement liée aux niveaux d'abstraction et d'observation fixés. D'où l'importance du principe de modélisation suivant : Choisir judicieusement aussi bien le niveau d'abstraction que le niveau d'observation. Pour chaque niveau de description, préciser également son outillage d'observation et bien cerner ce qui peut être observable. Dans l'analyse et la compréhension d'un phénomène biologique, l'observabilité ne sert-elle pas justement à bien définir les niveaux de description et d'observation ?

4 Le travail de l'atelier

Le système biologique choisi pour illustrer notre approche est celui, déjà largement décrit dans la partie 1, de la production de mucus chez la bactérie *Pseudomonas aeruginosa*. Notre exposé dans cette partie restera largement informel afin de donner les grandes lignes de la méthodologie. Nous serons amenés à introduire quelques notions de logique temporelle. Ainsi, tous les arguments avancés devront être validés par des arguments de logique formelle. Ces travaux sont actuellement en cours. Un modèle sert à établir des propriétés sur un système, à exprimer et manipuler ces propriétés afin d'en extraire d'autres non triviales. Il faut donc formaliser les propriétés de telle manière à ce qu'elles soient manipulables par un ordinateur. L'objectif fixé ici concerne la génération de scénarii d'expériences, dans lesquelles le temps joue un rôle central. On est nécessairement confronté à la notion de temps lorsque l'on veut exprimer des propriétés du système dans le futur. Ces contraintes conduisent naturellement à des logiques temporelles (Cf. [6], [7] pour une description générale des logiques temporelles).

\Rightarrow	0	1
0	1	1
1	0	1

TAB. 1 – Table de vérité de $p \Rightarrow q$

Plus précisément, on veut prouver que, en présence de y , il est possible d'avoir un état récurrent dans lequel du mucus est produit. Par construction du graphe, la présence de y est vérifiée, et le graphe a été validé (biologiquement) du point de vue de la topologie et des signes d'interaction. Seuls les seuils et les valeurs de la fonction K peuvent entraîner plusieurs comportements différents. Reste donc à montrer que les valeurs des seuils et de la fonction K pour l'organisme étudié sont telles que la bactérie peut passer dans un état où z s'exprime de manière récurrente.

Le langage des logiques temporelles comporte non seulement les connecteurs classiques de logique comme par exemple, le "ou", noté \vee , le "et", noté \wedge , l'implication notée \implies , mais aussi des connecteurs particuliers à ce type de logique qui portent sur le temps. On peut par exemple créer le connecteur \mathcal{F}_s , qui signifie que la formule qui suit le connecteur est vraie dans le "futur strict". On appelle ici "futur strict" le futur commençant après un certain délai. Ce délai doit être choisi en fonction des contraintes biologiques du système étudié.

Une fois la logique temporelle correspondant au modèle créée, il faut s'efforcer d'établir la formule que l'on cherche à prouver avec ce formalisme. On veut prouver que dans le modèle que l'on considère, si à un moment donné, la bactérie est dans un état muqueux, alors plus tard (dans un futur strict) elle sera de nouveau dans un état muqueux. Le futur strict signifie ici que l'on veut s'affranchir de la possibilité d'effet à court terme de l'état muqueux, c'est-à-dire que l'on veut s'affranchir d'une phase transitoire éventuelle. Pour des raisons biologiques on sait que le seuil associé à l'interaction $x \rightarrow z$ est égal à 2, et on sait par construction que celui associé à $y \rightarrow x$ vaut 1 (y n'influe que sur x , donc il n'y a qu'un seul seuil pour y). Par contre on ne connaît pas les seuils des arcs $x \rightarrow x$ et $x \rightarrow y$. Autrement dit, on ne connaît pas les quantités relatives de la variable x nécessaires pour obtenir, un effet d'autoentretien, un effet sur y ou un effet combiné. On est amené à vouloir prouver expérimentalement que les forces relatives de ces deux cycles utilisées par la bactérie sont telles qu'il est possible de rendre récurrent l'état ($x=2$), ce qui s'écrit :

$$(x = 2) \implies \mathcal{F}_s(x = 2)$$

Tout scénario qui teste cette formule commence par mettre (artificiellement si nécessaire) x à 2. Dans le cas contraire, la première partie de la formule est fautive, ce qui entraîne que toute la formule est vraie quelque soit la valeur de la deuxième partie de la formule (Cf. table 1). Le scénario d'expériences est donc le suivant :

1. On commence par imposer ($x = 2$).
2. On attend un laps de temps, puis on teste l'état muqueux.
 - Si la bactérie n'est pas dans un état muqueux, alors l'expérience échoue *a priori*.
 - Si la bactérie est à l'état muqueux (donc $x = 2$), on arrête l'expérience car la formule précédente est *ré-entrante*, ce qui veut dire que pour ce nouvel instant on peut de

nouveau appliquer la formule.

Autrement dit, au temps t_0 , on impose $x = 2$, par la formule, l'expérience nous affirme qu'il existe $t_1 = t_0 + d$ (d est un délai), tel que $x = 2$. En faisant un décalage de l'origine du temps, on peut alors affirmer qu'il existe $t_2 = t_1 + d$ tel que $x = 2 \dots$

En itérant cette formule, on montre que l'état muqueux atteint par la bactérie est récurrent. Le scénario suffisant à prouver le théorème est donc le suivant :

1. pulser x à la valeur 2 par une intervention extérieure,
2. attendre suffisamment (pour être sûr qu'on n'est plus dans la condition initiale, c'est-à-dire pour être sûr que la phase de transition dû au pulse initial soit dépassée) pour obtenir à nouveau un état muqueux, c'est-à-dire $x = 2$.

Limites de l'approche - Le graphe sur lequel on fonde le modèle (figure 1) n'est en réalité qu'un sous graphe d'un graphe plus général mettant en scène toutes les variables de l'organisme. Il faudrait donc considérer toutes les interactions avec la partie négligée du graphe général. Avoir négligé les arcs sortants du graphe n'a pas de conséquence puisque l'on ne s'intéresse qu'au sous système entraînant la production de mucus. Par contre, le fait d'avoir négligé les arcs entrants dans ce sous-système peut avoir un impact important. Par construction du graphe, certaines situations peuvent être éliminées.

1. Par définition de z , l'unique arc régulant z est celle que l'on prend en compte : $x \rightarrow z$ et il n'existe donc pas d'autre arc entrant sur z .
2. Tous les arcs qui n'ont pas été considérés dans le modèle mais régulant x ou y ne font pas partie d'un cycle. Le nombre d'états stables ne change pas (Cf. article [8]).
3. S'il existe des arcs entrants sur x et y dont l'influence ne varie pas, la seule conséquence du fait d'avoir extrait un sous-graphe est de décaler les différents seuils associés aux variables x et y . Le système aura d'autres valeurs pour les seuils et éventuellement pour la fonction K , mais les variables seront toujours discrétisées. Les états stables pourront éventuellement ne pas être les mêmes, mais la formule à prouver reste la même.

Un seul cas demeure gênant : il se peut qu'il y ait des régulateurs extérieurs à ce sous-graphe (sur x et y) dont l'influence varie dans le temps. Toute l'étude présentée ici fait l'hypothèse que ces influences sont négligeables. Ces travaux sont cependant valables dans l'hypothèse d'un plongement à influence constante sur les variables x et y .

Enfin, rappelons que le délai mentionné plus haut reste empirique.

5 Conclusion

Le travail interdisciplinaire mené par l'atelier détermine un cadre méthodologique pour définir des modèles comprenant un «kit» de validation/réfutation expérimentale. En ce sens, ce travail renforce résolument l'activité modélisatrice en renforçant sa crédibilité face, avouons le, à la méfiance qu'elle suscite en biologie. En effet en suivant un principe Popperien, cette méthodologie offre l'opportunité de valider/réfuter un modèle par l'expérience, activité centrale en biologie.

Etablir une telle approche nécessite une théorie qui fixe les règles permettant de «raisonner» à partir d'un modèle. La théorie introduite ici est celle de la logique temporelle couramment employée pour l'analyse logique de la dynamique des systèmes. Selon cette théorie, l'étude de cas met en valeur qu'un modèle qualitatif discret d'expression des gènes fondé sur les travaux de René Thomas répond à l'exigence méthodologique mentionnée. Elle permet en effet de déterminer un protocole d'expérimentation pour valider/réfuter l'hypothèse épigénétique que décrit ce modèle (section 4).

Inspirée de celle pratiquée pour analyser les systèmes informatiques, son cadre d'utilisation originel suggère que l'on puisse l'automatiser ; c'est-à-dire fournir des logiciels assistants la conception de tels modèles. Réaliser ces assistants nécessite de poursuivre l'investigation entreprise d'appliquer une méthode dévolue à l'analyse de systèmes formels, les logiciels, à celle du vivant.

Références

- [1] R. Thomas, D. Thieffry, and M. Kaufman, "Dynamical behaviours of regulatory networks - I . biological role of feedback loops and practical use of the concept of feedback loop," *Bulletin of Mathematical Biology*, vol. 57, no. 2, pp. 247–276, 1995.
- [2] R. Thomas and D. Thieffry, "Dynamical behaviours of regulatory networks - II . immunity control in bactériophage lambda," *Bulletin of Mathematical Biology*, vol. 57, no. 2, pp. 277–297, 1995.
- [3] J. Guespin-Michel, "Epigenesis, multistationarity and positive feedback circuits (cours)." Séminaire Modélisation et simulation de processus biologiques dans le contexte de la génomique, Autrans, Mars 2002.
- [4] J. Guespin-Michel and M. Kaufman, "Positive feedback circuits and adaptive regulations in bacteria," *Acta biotheoretica*, vol. 49, pp. 207–218, 2001.
- [5] K. Popper, *conjectures et réfutations*. Taylor and Francis Books, 1969.
- [6] M. R. Huth and M. D. Ryan, *Logic in Computer Science : Modelling and Reasoning about Systems*. Cambridge University Press, 2000.
- [7] R. Lalement, *Logique, réduction, résolution*. Masson, 1990.
- [8] E. Snoussy, "Qualitative dynamics of a piecewise-linear differential equations : a discrete mapping approach," *Dynamical Stab. System*, vol. 4, pp. 189–207, 1989.