

# Les réseaux de régulation génétique : un exemple de modélisation en biologie

Jean Della-Dora - Laurent Tournier

L.M.C. - I.M.A.G.  
BP 53 38041 Grenoble, France  
{Jean.Delladora, Laurent.Tournier}@imag.fr

**Résumé** La modélisation mathématique de phénomènes physiques et biologiques complexes est un véritable défi posé aux mathématiciens. Devant l'arrivée en masse de données biologiques (notamment en génétique), il devient important de fournir des modèles permettant d'exploiter ces données afin d'aider à comprendre les phénomènes qui rentrent en jeu. Dans cet article, nous exposons le problème de la régulation génétique et proposons une modélisation à l'aide de systèmes hybrides.

## 1 Quelques remarques initiales sur la modélisation

La prise en compte de la modélisation mathématique dans plusieurs concours importants à travers des épreuves comme les TIPE pour les concours ENSI, l'épreuve de modélisation à l'Agrégation pose certainement beaucoup de problèmes de préparation. Dans ce texte nous allons développer quelques remarques conséquences de notre participation à plusieurs des jurys précédemment cités. Aujourd'hui que la presse grand public se fait l'écho de problèmes liés à la place des mathématiques dans l'enseignement général, je crois que nous devons non pas réagir comme nous le faisons trop souvent (et bien souvent avec des arguments très conservateurs) mais proposer nous-même des solutions.

D'abord il faut savoir de quoi on parle. Et l'expérience montre que ce n'est pas simple! Il suffit de regarder les textes proposés dans plusieurs concours pour voir que la démarche de modélisation qui commence par **l'établissement du modèle** est complètement obérée. Elle est remplacée par un exercice de mathématique classique sur un système d'équations convenu. C'est donc un exercice au sens habituel, habillé des couleurs d'une discipline très adjacente.

Pour cela il faut situer correctement la modélisation. Nous avons beaucoup appris à la lecture du texte de Giorgio Israel intitulé **La mathématisation du réel** [2].

On y verra, en particulier, une très pertinente critique de l'éternelle (pour les concours) lutte des requins et des sardines et de son intérêt des plus limité. Commencer la modélisation en biologie par cette étude ne manque pas de sel!

Une fois cette saine réflexion entreprise on peut chercher les textes accessibles et vivants (c'est à dire non axés sur un formalisme mathématique figé). Le livre de C.C Lin et L.A.Segel [4] est très instructif en particulier le chapitre 6 consacré à "Simplification, Dimensional Analysis and Scaling". Comme cela semble une

lacune générale dans nos programmes de mathématique on y apprendra l'importance des échelles et de l'analyse dimensionnelle. On pourra objecter que tout cela doit se faire en Physique et que ce sont des considérations bien élémentaires ; ce serait une grossière erreur car d'une part la compréhension des ordres de grandeur ouvre la voie à l'asymptotique et d'autre part cette analyse est à la base de l'étude des équations différentielles par des méthodes de symétrie qui, bien que peu enseignée en France, est à la base de nombreux travaux actuels en Physique. Historiquement on pourra consulter le très intéressant ouvrage de Garret Birkhoff "Hydrodynamics" (A study in logic, fact, and similitude) [5]. Les exemples de [4] contiennent aussi beaucoup d'applications à la biologie.

Si l'on veut comprendre plus en profondeur ce qu'est un modèle et comment on l'étudie (comment on doit le désacraliser . . .) on lira l'ouvrage de Rutherford Aris [6] "Mathematical modeling techniques". On verra, sur des exemples importants, comment un modèle mathématique doit être critiqué, modifié, amendé et comment l'on apprend beaucoup à le simplifier. Cette démarche "à la physicienne" me semble fondamentale pour deux raisons : elle apprend à ne pas considérer un modèle comme un absolu mais un outil opératoire comme un autre (cela devrait empêcher de numériser n'importe quoi, n'importe comment), d'autre part la simplification d'un modèle permet l'utilisation d'outils mathématiques "analytiques" qui justifient les heures passées à étudier les solutions de telles ou telles équations particulières.

Naturellement, et malgré la très grande difficulté de la démarche (que l'on peut très bien ne pas cautionner) il faut essayer de suivre la démarche de René Thom exposée dans [7].

En conclusion nous pensons que l'on met la traditionnelle charue avant le tracteur de l'histoire et qu'il serait bon d'abord de discuter avec de vrais praticiens sur les méthodologies à la base du processus de modélisation. On y verra sûrement qu'il ne peut être question de proclamer l'impérieuse nécessité d'enseigner telle ou telle discipline absolument indispensable aujourd'hui (plus indispensable que moi tu meurs . . .) : théorie des fonctions d'une variable complexe pour ingénieur avant hier, analyse numérique hier, statistique aujourd'hui, etc mais qu'il est plus important d'avoir une culture mathématique solide, généreuse et à large spectre. Tous les ingénieurs vous le diront ! (Ce qui n'empêche pas de réfléchir aux programmes mais peut être avec une plus grande ouverture d'esprit).

## 2 De la modélisation en biologie

L'idée de cet article est que si l'on veut faire dans le classique, alors il vaut mieux s'arrêter là et se précipiter (comme beaucoup de générateurs de textes de concours) sur la somme de J.D.Murray "Mathematical biology" [1] c'est un livre fascinant et très complet. On pourra éventuellement le compléter sur quelques points techniques par le livre de P.Grindrod sur "Patterns and Waves" [8]. Ceci pour les généralités.

On peut aussi essayer (suivant les suggestions du paragraphe précédent) de se

faire sa propre opinion. Pour cela on dispose des merveilleux ouvrages de Stephen Jay Gould, en particulier de [3] “Aux racines du temps”. En effet la première chose importante en biologie ce sont les échelles de temps : le temps profond de Jay Gould a été une notion monstrueusement difficile à appréhender (et pour cause) vues les racines religieuses et historiques du problème. Temps linéaire ou temps cyclique ? Ces questions historiques et biologiques conduisent alors au coeur d’un premier problème important en modélisation biologique : dans les systèmes vivants l’évolution et le mécanisme de fonctionnement d’un être vivant font cohabiter la plus large gamme d’échelles de temps possible. Nous venons de parler du temps “profond” qui se mesure (au moins) par l’âge de la terre, mais une cellule étant le siège de réactions chimiques incessantes et multiples, c’est l’échelle de la femto seconde qui est nécessaire. Comment construire des modèles dans ce cadre ? Est-il raisonnable de penser à modéliser le fonctionnement d’une cellule compte tenue du fait que la dimension historique est une des caractéristiques du vivant ?

Cette remarque étant faite nous allons proposer dans la suite de cet article un exemple actuel de problème de modélisation en biologie moléculaire. Cet exemple, bien qu’historique, est toujours d’actualité.

### 3 La notion de réseau génétique

Dans cette partie, nous allons tenter de décrire brièvement ce que l’on entend par réseau de **régulation** génétique. Pour cela, nous allons d’abord nous familiariser avec certaines notions de base de la biologie cellulaire, notamment la mécanique enzymatique et le contrôle génétique. Pour une description plus détaillée de ces phénomènes, on se reportera à [12], ainsi qu’à [13].

#### 3.1 Éléments de biologie cellulaire

La biologie cellulaire est une science très récente, puisque la découverte de la molécule d’ADN date seulement de 1868, et ce n’est qu’au milieu du vingtième siècle que l’on a compris que cette molécule était effectivement le *matériel génétique*. Cette science cherche en fait à expliquer les différents mécanismes qui se déroulent à l’échelle moléculaire et qui, s’enchaînant de manière complexe les uns aux autres, constituent l’**activité** ou **vie cellulaire**. En d’autres termes, elle tente de comprendre comment une cellule (c’est-à-dire une entité élémentaire de la vie organique) fait pour s’alimenter, se diviser, et organiser la multitude de tâches nécessaires à sa survie ?

Nous savons aujourd’hui que chaque cellule possède un bagage génétique qui porte une quantité gigantesque d’information. C’est cette masse d’information qui sert à diriger le métabolisme de la cellule, exactement comme une usine dans laquelle un *centre de commande* dirige une *machinerie*. Grossièrement, les **gènes** sont *traduits* ou *exprimés* pour produire les protéines qui exécutent les fonctions cellulaires : c’est le **dogme** central de la biologie cellulaire. Néanmoins, même si l’on arrive à décoder petit à petit le génôme humain, nous sommes encore

incapables de décrire avec précision tous les liens complexes qui existent entre le métabolisme cellulaire et l'information génétique.

Nous allons à présent étudier la cellule à plus petite échelle. Tous les mécanismes internes peuvent s'interpréter alors comme des interactions chimiques entre les différents composants présents dans le milieu, commandées par la présence d'agents particulièrement important : les enzymes.

**la mécanique enzymatique.** Les briques de base de la matière organique sont les **protéines**. Ce sont de grosses molécules constituées d'une séquence plus ou moins longue d'acides aminés. Ce séquençage a pour effet de donner à la protéine une forme spatiale complexe qui détermine sa **fonction**. Certaines de ces molécules, les **enzymes**, ont la propriété d'accélérer (on préférera le verbe *catalyser*) un certain type de réaction biochimique. Ces réactions étant à l'état naturel extrêmement lentes, la présence d'un catalyseur est fondamentale car c'est elle qui détermine si la réaction a lieu ou non. Nous voyons ainsi que les enzymes sont des acteurs directs de la vie cellulaire : par leur présence ou leur absence, ils contribuent à effectuer telle ou telle tâche.

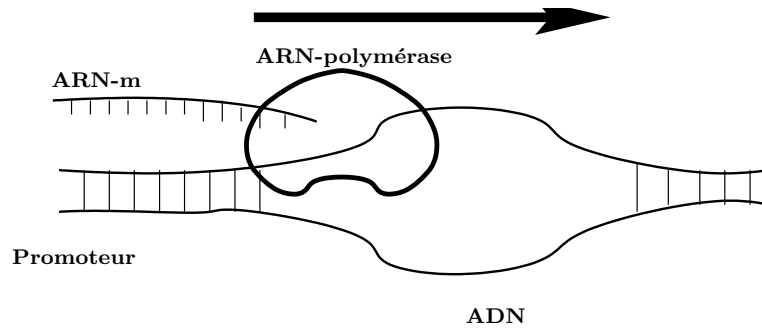
Pour mieux comprendre ce phénomène, prenons l'exemple de l'alimentation cellulaire d'une bactérie. Pour survivre, la première tâche de la cellule est de produire de façon constante de l'énergie. Pour se faire, elle doit dégrader les molécules de glucose dont elle dispose et récupérer l'énergie qui résulte de cette dégradation. Or, la bactérie peut être amenée à manquer de glucose (si par exemple le milieu extérieur en est dépourvu). Elle pourra alors se tourner vers une autre source d'énergie, le lactose. Ce sucre est plus difficile à exploiter car il demande un traitement particulier : il faut le laisser entrer dans la cellule, puis le transformer en glucose. Ces deux opérations demandent la présence de deux enzymes respectifs : la perméase et la  $\beta$ -galactosidase. Si la cellule veut survivre en cas de carence en glucose, il faudra donc qu'elle se mette à produire ces deux enzymes rapidement de manière à s'**adapter** à ses nouvelles conditions de vie.

**le contrôle génétique.** une *fonction* ou *tâche* cellulaire peut donc s'interpréter comme une réaction ou une chaîne de réactions biochimiques, commandées par un ou plusieurs enzymes. Lorsque les conditions l'exigent, une cellule doit donc produire en quantité suffisante les substances dont elle a besoin pour survivre. Nous devons comprendre comment se fait cette production.

Ce processus se fait par l'**expression** de l'information contenue dans l'ADN, par l'intermédiaire de deux étapes : la *transcription* et la *traduction*. Grossièrement, l'étape de transcription peut être vue comme une recopie de l'information génétique sur un autre support : une molécule d'ARN, tandis que la traduction consiste en l'interprétation de cette information dans le langage des acides aminés.

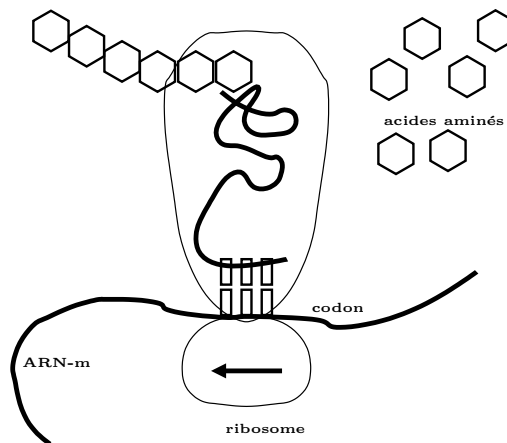
Ainsi lorsqu'un certain type de protéine devient nécessaire, le gène qui la code est transcrit rapidement de manière à augmenter la concentration en l'espèce voulue. Par l'intermédiaire d'une grosse enzyme : l'**ARN-polymérase**, le segment d'ADN concerné est recopié sur plusieurs segments d'ARN (c'est la *transcription* : voir figure 1), puis, ces molécules d'ARN se déplacent dans le milieu

intracellulaire pour être traduites en protéines (c'est la *traduction*).



**Fig.1.** schéma simplifié de la transcription

L'alphabet de l'information génétique est constitué de quatre lettres :  $\{A, T, C, G\}$  qui correspondent à des **bases** ou **nucléotides** qui s'enchaînent le long de la molécule d'ADN. L'alphabet des protéines est quant à lui constitué d'une vingtaine de lettres qui correspondent aux acides aminés contenus dans la nature. La traduction se fait donc par l'intermédiaire de triplets de bases, ou **codons**, qui sont lus par des agents appelés ribosomes. La figure 2 présente un schéma simplifié de la traduction.



**Fig.2.** schéma simplifié de la traduction

**la régulation génétique.** Aujourd'hui le mécanisme de l'expression génétique est globalement connu. Mais si l'on comprend comment la cellule produit les protéines dont elle a besoin et comment ces protéines interagissent pour accomplir la tâche voulue, il est encore difficile de comprendre comment ces mécanismes s'enchaînent et comment ils sont régulés à l'échelle de la cellule.

Grâce à de nombreux travaux, notamment ceux menés par F. Jacob et J. Monod, nous savons qu'il existe des protéines particulières, spécifiquement dédiées à la régulation. Ces **protéines régulatrices**, de par leur structure, se lient à la molécule d'ADN sur des sites particuliers appelés **opérateurs** et ont pour effet de bloquer ou d'accélérer la transcription.

Nous observons donc un phénomène de **rétroaction** : les gènes codent pour des protéines, or certaines de ces protéines interviennent dans l'expression génétique elle-même (les protéines régulatrices et également certains enzymes, notamment la polymérase).

Ce sont ces boucles de rétroaction qui sont à l'origine de la notion de réseau génétique que nous allons aborder à présent. Nous cherchons à comprendre, d'un point de vue mathématique, comment peuvent se modéliser de tels réseaux. Ce sont des notions relativement récentes, et les modèles mathématiques qui ont été proposés sont encore souvent incomplets.

### 3.2 Un exemple de régulation : l'opéron lactose (Jacob & Monod)

Afin de mieux comprendre ce que l'on entend par réseau de régulation génétique, nous allons considérer l'exemple de l'opéron lactose, proposé par F. Jacob et J. Monod. Ces deux chercheurs français ont reçu en 1965 le prix Nobel de médecine pour leurs travaux sur la régulation génétique (on pourra par exemple se reporter à [14]). Ils ont notamment travaillé sur le métabolisme du lactose chez la bactérie *escherichia coli* ou *e-coli* et ils ont mis en évidence la structure de régulation qui le commande.

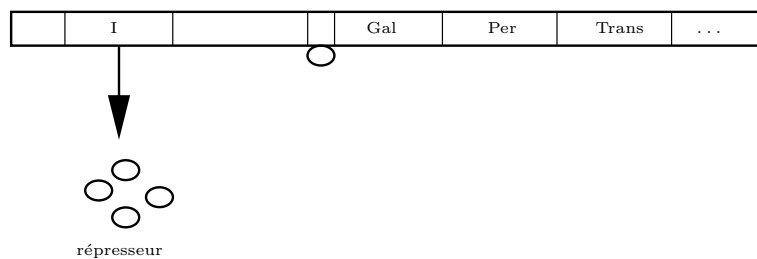
Comme nous l'avons expliqué plus haut, en cas de carence en glucose, la bactérie (qui est un organisme mono-cellulaire) doit se tourner vers une solution de remplacement pour continuer à produire de l'énergie. Si le milieu extérieur est riche en lactose, la solution choisie est d'exploiter ce sucre. En effet, Jacob et Monod ont observé que lorsqu'une culture de bactéries est privée de glucose et en présence de lactose, la concentration en perméase et en  $\beta$ -galactosidase subit une brusque augmentation à l'intérieur des cellules, alors qu'en temps normal ces deux enzymes sont quasiment absents. Ce sont justement ces deux molécules dont la bactérie a besoin pour traiter le lactose.

Ce phénomène constitue une **bascule génétique**, c'est-à-dire le passage brusque d'un état où les gènes des deux enzymes ne sont quasiment pas exprimés à un état où la production est au contraire très intense. Dans ce cas précis on parle également d'**induction**. Le terme bascule correspond à un changement qui se fait en l'espace de quelques minutes seulement. C'est donc un évènement relativement brutal dans l'échelle de temps de la vie cellulaire.

Voyons à présent comment s'explique cette bascule à l'échelle moléculaire.

**description de l'opéron.** Chez la bactérie *e-coli*, les deux gènes qui codent pour les deux enzymes sont côte-à-côte le long de l'unique molécule d'ADN (voir figure 3). Juste avant le début du premier gène se trouve un site particulier, appelé opérateur. Cet opérateur, de par sa séquence tout-à-fait particulière de nucléotides, présente une affinité naturelle avec une protéine spécifique produite en amont par un autre gène : I. Cette molécule est appelée **répresseur** car la liaison avec l'opérateur a pour conséquence de bloquer la liaison de l'ARN-polymérase avec l'ADN et donc d'empêcher la transcription des deux gènes.

En régime normal, il existe une quantité à peu près constante de répresseur dans la cellule (relativement peu) et de cette manière la transcription des enzymes ne se fait pas. L'opéron est alors bloqué (c'est le cas de la figure 3).



**Fig.3.** opéron lactose bloqué

**déclenchement de la bascule.** Lorsque la bactérie est en présence de lactose, la quantité résiduelle de perméase suffit à faire pénétrer quelques molécules à l'intérieur de la cellule. Le lactose, qui est l'**inducteur** de l'opéron se lie alors au répresseur et lui ôte ainsi son affinité avec l'opérateur. La polymérase n'est alors plus bloquée et la transcription peut s'amorcer. L'opéron est dit débloqué ou **induit**. Grâce à la perméase produite, le lactose rentre en masse dans la cellule et il commence à être traité.

Cette boucle de rétroaction est appelé boucle négative de l'opéron. Son rôle est de bloquer l'opéron tant que le milieu est dépourvu de lactose (puisqu'alors les deux enzymes ne sont pas utiles). Il existe également une boucle positive au rôle sensiblement différent. Comme nous l'avons déjà évoqué, la dégradation lactique est une opération qui n'est nécessaire que dans le cas où la cellule ne dispose plus de glucose. Pour répondre à cette demande, il existe une molécule au rôle **activateur** : la CAP (*catabolite activator protein*). Cette molécule forme un complexe avec une autre substance : l'*AMP cyclique* (AMP-c). Ce complexe se lie à la molécule d'ADN et cette liaison a pour effet d'augmenter la transcription. Or, l'AMP-c n'est présente que si le glucose vient à manquer.

Par cette double boucle de régulation, l'opéron contrôle la production des deux enzymes. S'il y a du lactose, ces deux enzymes sont utiles et l'opéron est

induit, mais la transcription reste faible. Si en outre le glucose vient à manquer, les deux enzymes deviennent indispensables et l'induction est fortement accélérée (expérimentalement d'un facteur 50).

*Remarque : Le fonctionnement décrit ici est en fait incomplet. Il existe un autre gène qui suit ceux de la perméase et la galactosidase le long de la molécule d'ADN. Il code pour un autre enzyme, la transacétylase, dont le rôle dans la régulation est aujourd'hui encore mal connu.*

## 4 Les différents types de modélisations

Les différents modèles qui ont été proposés pour tenter de donner une approche mathématique des réseaux génétiques se rangent pour l'essentiel dans deux catégories : les systèmes discrets et les systèmes continus. Après une brève description de ces deux modélisations, nous présenterons un modèle hybride, qui décrit l'évolution continue des réseaux tout en prenant en compte l'aspect discret qui les caractérise.

### 4.1 L'idéalisation booléenne

Les systèmes discrets semblent assez bien adaptés aux systèmes biologiques car ces derniers présentent des particularités discrètes intéressantes, notamment le phénomène de bascule précédemment décrit. Dans l'idéalisation booléenne des réseaux génétiques, on modélise un gène (ou tout autre *stimulus* biologique) par une variable booléenne. Un gène est donc soit *allumé* (transcrit), soit *éteint* (non transcrit). On représente ensuite les influences positives ou négatives d'un gène sur les autres par des fonctions booléennes. L'avantage d'un tel modèle est bien sûr sa simplicité de calcul. Nous pouvons ainsi simuler des réseaux de taille relativement importante et répondre par exemple à des questions générales comme le nombre de types cellulaires. L'un des grands noms de cette modélisation est Stuart Kauffman (voir [16]). On pourra également consulter les travaux de René Thomas (voir [17]).

Plus précisément, un réseau génétique booléen se modélise comme un réseau booléen. C'est un **graphe orienté**  $G = (V, F)$  où  $V = \{v_1, v_2, \dots, v_n\}$  est un ensemble fini de **nœuds** et où  $F$  est un ensemble de **fonctions booléennes** qui définit une topologie d'arêtes.  $n$  est appelé *taille* ou *dimension* du réseau.

Un nœud représente soit un gène, soit un **stimulus biologique**, c'est-à-dire un facteur physique ou chimique qui fait partie du réseau mais qui n'est ni un gène, ni le produit d'un gène (par exemple une irradiation, ou tout simplement une quantité extérieure qui rentre dans le système). A chaque nœud  $v$ , on associe une valeur booléenne  $x(v)$  qui représente le **niveau d'expression** du stimulus correspondant. Ainsi pour un gène, la valeur de  $x$  sera 0 si le gène est *éteint* (ou non transcrit) et 1 si le gène est *allumé* (ou transcrit). De même, pour une



espèce chimique, la valeur sera 1 si l'espèce est considérée comme *présente* (c'est-à-dire si elle est au-dessus d'un certain seuil) et elle sera 0 si l'espèce est *absente* (au-dessous du seuil).

Un tel réseau fonctionne en temps discret : une **itération** se formalise par l'application d'une fonction  $\psi$  :

$$(x_1^{(t+1)}, \dots, x_n^{(t+1)}) = \psi(x_1^{(t)}, \dots, x_n^{(t)})$$

où chaque  $\psi_i$  est une fonction booléenne à  $k$  entrées ( $k$  est appelée connectivité de la fonction).

Malheureusement l'approche discrète, bien que relativement simple, est trop restrictive. Représenter une variable biologique par une valeur booléenne ne rend pas compte de la réalité. En revanche, elle permet par sa simplicité de calcul de manipuler de gros réseaux et d'apporter ainsi des moyens informatiques pour traiter les masses de données expérimentales provenant des nouvelles technologies, notamment celle des biopuces ou *microarrays* (voir [15]). A titre d'exemple, on consultera l'article [18] qui expose le principe d'un algorithme d'identification de réseau booléen à partir de séries de données temporelles.

#### 4.2 Un exemple de modélisation continue : le modèle linéaire par morceaux

Pour approcher de manière plus fine le fonctionnement temporel d'un réseau génétique, une autre approche consiste à considérer l'évolution *continue* du système : les variables sont les concentrations en les différentes espèces du milieu et l'évolution globale de l'ensemble est régie par un système d'équations différentielles. Cette modélisation est beaucoup plus proche de la réalité mais elle est de manière générale plus difficile à manipuler, surtout lorsque la dimension augmente. Ceci est dû essentiellement au caractère non linéaire des équations différentielles de régulation.

Une solution consiste à simplifier les équations et à considérer par exemple un système **linéaire par morceaux** ([19], [20]). L'espace des phases est divisé en *boîtes* délimitées par des seuils correspondant à chaque variable. A l'intérieur de chaque boîte, on simplifie le système différentiel en considérant le système linéaire :

$$\frac{dx_i}{dt}(t) = \gamma_i(\mathcal{B}, x) - \nu_i(\mathcal{B}, x)x_i(t)$$

où  $\mathcal{B}$  est la boîte considérée et  $x$  le vecteur des concentrations.

Le terme  $\gamma_i$  est un terme de production (que l'on suppose constant à l'intérieur de la boîte  $\mathcal{B}$ ) et le terme  $-\nu_i x_i$  est un terme de dégradation linéaire qui correspond à la dégradation naturelle de  $i$  ( $\nu_i$  est donc généralement considérée comme une constante qui ne dépend que de la nature de la substance).

La résolution de tels système peut se faire de manière **exacte** à l'intérieur de chaque boîte et l'on peut donc simuler des réseaux relativement gros et observer des comportements dynamiques intéressants (cycles, points fixes, ...)

### 4.3 L'approche hybride

Comme nous l'avons déjà évoqué, le modèle discret et le modèle continu sont deux approches très différentes du même problème. Pourtant, elles présentent toutes les deux plusieurs avantages. Il est certain qu'une représentation purement booléenne est trop éloignée de la réalité et on lui préférera une modélisation prenant en compte le caractère *continu* des variables biologiques. Pourtant, le phénomène de bascule que nous avons décrit fait passer le système d'une configuration dynamique à une autre configuration complètement différente en un temps très court, ce qui n'est pas sans rappeler les transitions de l'idéalisation booléenne. L'idée d'allier les deux modèles en utilisant la théorie des systèmes hybrides paraît donc tout-à-fait justifiée.

Nous allons tenter de rappeler brièvement ce qu'est un système hybride (on trouvera la définition complète dans [21] et [22]). Il s'agit globalement d'un système dynamique dont le second membre dépend d'une configuration discrète. En d'autres termes, les variables évoluent de façon continue selon une loi qui dépend de l'état dans lequel se trouve le système. Lorsqu'elles atteignent la frontière de cet état, elles *sautent* dans l'état suivant et suivent alors de nouvelles équations différentielles. Le lien que nous faisons avec les réseaux génétiques est simple : une bascule se modélise par une transition, qui fait passer notre système différentiel d'une configuration à la suivante. Remarquons que par définition, les transitions d'un système hybride se font en temps nul, ce qui n'est bien sûr pas le cas dans un système biologique. Cependant, la rapidité des bascules relativement à l'évolution temporelle générale de la cellule semble rendre cette approximation acceptable.

Afin de nous familiariser avec ce modèle, nous allons appliquer directement la définition des systèmes hybrides au cas de l'opéron lactose.

**les variables.** Tout d'abord, Il nous faut isoler un ensemble de **variables continues** dont l'évolution temporelle nous intéresse : il s'agit des concentrations en les différentes espèces qui composent le système. En étudiant chacune de ces espèces, nous devons déterminer lesquelles ont un rôle régulateur, c'est-à-dire celles dont la valeur de la concentration est importante pour déterminer la configuration du système.

Dans le cas de l'opéron lactose, nous choisissons les concentrations en glucose et en lactose internes comme variables de régulation. Les autres variables sont les concentrations des autres espèces composant le système : les enzymes et le lactose extérieur.

**les états.** Il nous faut ensuite déterminer les **états discrets** (ou configurations) de notre système. Nous définissons ces états à l'aide des variables régulatrices que nous avons déterminé plus haut, selon leurs positions relativement à un seuil. Dans notre cas, nous aurons donc quatre états discrets, que nous noterons  $GL$ ,  $G\bar{L}$ ,  $\bar{G}L$ ,  $\bar{G}\bar{L}$ .

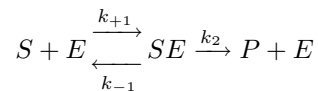
- $G\bar{L}$  : c'est l'état le plus fréquent. La cellule dispose de glucose, elle n'a donc pas besoin de lactose : l'opéron est bloqué.

- $\bar{G}L$  : il s'agit de l'état inverse, cette fois la cellule n'a plus de glucose et dispose de lactose. L'opéron est donc débloqué et accéléré par l'AMP cyclique. On parle d'induction activée.
- $GL$  : c'est le cas où la cellule dispose des deux sucres. l'opéron est débloqué mais la transcription est freinée par la présence de glucose. On parle d'induction modérée.
- $\bar{G}\bar{L}$  : la cellule ne dispose d'aucun sucre. Dans notre modèle, la cellule reste dans cet état dans l'attente d'un apport extérieur.

**les équations.** La prochaine étape consiste à déterminer le **système différentiel** qui régit la dynamique du système. Pour cela, nous utilisons les équations différentielles données par la cinétique chimique. il nous faut donc établir les interactions entre les différents composants présents dans le milieu.

Il nous faudra également introduire des termes de production sous forme de constantes de production  $\gamma_{esp}$  pour modéliser la fabrication de protéines à partir de l'ADN. Enfin, on prévoira des termes de dégradation (termes linéaires avec un taux  $\nu_{esp}$ ) qui correspondent à l'usure naturelle des molécules organiques au cours du temps. Ces derniers seront présents pour toutes les protéines et dans toutes les équations ; en revanche, les termes de production dépendront de l'état dans lequel le système se trouve.

Afin de mieux comprendre comment interviennent les termes d'interaction, nous allons rappeler brièvement les équations de la catalyse enzymatique définies en premier par Michaelis et Menten (1913). Dans une réaction mono-enzymatique, l'enzyme et son substrat se lient au niveau du site catalytique de l'enzyme pour créer un complexe enzyme/substrat qui est à l'origine de l'apparition du produit. Schématiquement, la catalyse se représente par l'équation-bilan :

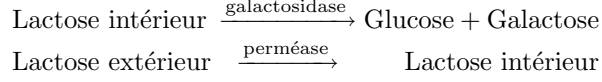


Grâce aux lois de la cinétique chimique, nous arrivons au système différentiel suivant :

$$\begin{cases} \frac{ds}{dt} = -k_1es + k_{-1}c & , \quad s(0) = s_0 \\ \frac{de}{dt} = -k_1es + (k_{-1} + k_2)c & , \quad e(0) = e_0 \\ \frac{dc}{dt} = k_1es - (k_{-1} + k_2)c & , \quad c(0) = 0 \\ \frac{dp}{dt} = k_2c & , \quad p(0) = 0 \end{cases}$$

où  $s$ ,  $e$ ,  $c$  et  $p$  représentent les concentrations en substrat, enzyme, complexe enzyme/substrat et produit.

Dans notre exemple, il y a deux réactions mono-enzymatiques à prendre en compte :



Il nous faut donc introduire deux variables supplémentaires  $c_1$  et  $c_2$  correspondant aux complexes formés respectivement par ces réactions.

Afin de résumer toutes ces idées, nous présentons ici le système différentiel correspondant à l'état  $GL$  d'induction modérée :

$$\begin{cases} \dot{x} = -\nu x + \gamma - k_1^{(1)}lx + (k_{-1}^{(1)} + k_2^{(1)})c_1 \\ \dot{y} = -\nu y + \gamma - k_1^{(2)}l_e y + (k_{-1}^{(2)} + k_2^{(2)})c_2 \\ \dot{g} = -K_{gluc}g + k_2^{(1)}c_1 + G \max(G_M - g, 0) \\ \dot{l} = k_2^{(2)}c_2 - k_1^{(1)}lx + k_{-1}^{(1)}c_1 \\ \dot{l}_e = k_{-1}^{(2)}c_2 - k_1^{(2)}l_e y + L \max(L_M - l_e, 0) \\ \dot{c}_1 = k_1^{(1)}lx - (k_{-1}^{(1)} + k_2^{(1)})c_1 \\ \dot{c}_2 = k_1^{(2)}l_e y - (k_{-1}^{(2)} + k_2^{(2)})c_2 \end{cases}$$

où  $x$ ,  $y$  représentent les concentrations de galactosidase et de perméase,  $g$ ,  $l$ ,  $l_e$  représentent les concentrations de glucose intérieur, de lactose intérieur et de lactose extérieur.

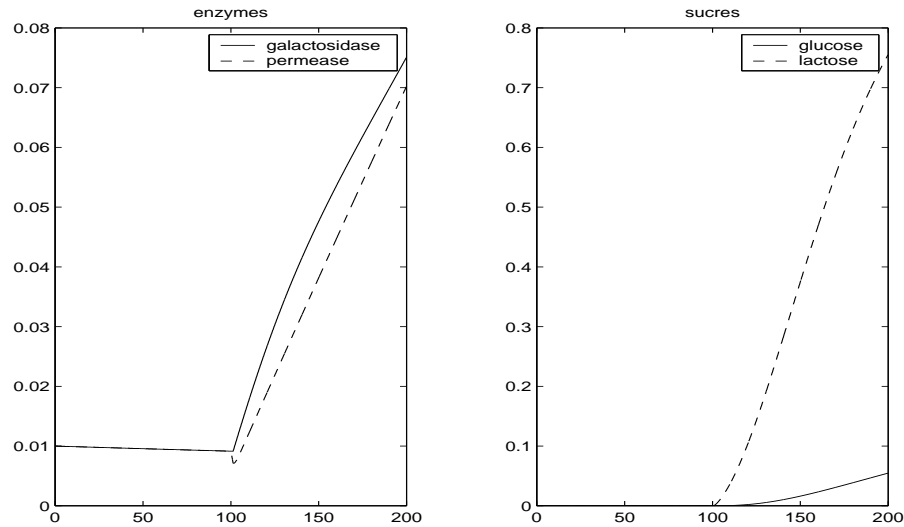
Les fonctions  $G$  et  $L$  sont des fonctions à valeurs dans  $\{0, 1\}$  et représentent l'apport extérieur en sucre. Les constantes  $G_M$  et  $L_M$  permettent de modérer cet apport (afin de ne pas faire exploser les variables  $g$  et  $l_e$ ).  $G$  et  $L$  exercent en fait un **contrôle** sur le système, elles sont fixées par l'utilisateur qui détermine ainsi l'environnement extérieur du système hybride.

Enfin, la constante  $K_{gluc}$  représente la dégradation naturelle du glucose dans la cellule afin d'en tirer de l'énergie (glycolyse).

**les transitions.** Il nous reste à déterminer les transitions de notre système hybride : dès que la trajectoire parvient à la frontière d'un état, c'est-à-dire lorsqu'un des deux sucres traverse le seuil, elle *saute* dans l'état suivant en prenant pour conditions initiales les valeurs des variables au moment où la frontière a été atteinte.

#### 4.4 résultats

Dans cette partie, nous allons présenter quelques courbes auxquelles nous a conduit ce modèle. Notons que ces résultats donnent lieu à des interprétations qualitatives, mais ne permettent pas pour le moment de tirer de conséquences quantitatives. Il faudrait pour cela disposer de plusieurs constantes chimiques (notamment des constantes de cinétique) avec précision. Toutefois, les résultats obtenus permettent d'illustrer les phénomènes de régulation que nous avons décrits, notamment le phénomène de bascule (voir fig 4).



**Fig.4.** Simulation du phénomène d'induction. Au temps  $t = 100$ , le milieu est mis en présence de lactose (l'inducteur). Les courbes des enzymes grimpent alors rapidement et le lactose rentre en masse dans la cellule. La consommation de lactose commence. Notons ici que l'induction est de plus activé par le fait qu'il n'y a pas de glucose.

## Conclusion

L'exemple précédent est typique des espoirs et des doutes que fait naître la démarche de modélisation en biologie. Les systèmes vivants sont incroyablement complexes et hétérogènes. Il faut donc isoler des sous systèmes (au risque de donner une trop grande place à certaines entités comme les gènes) et essayer d'en faire un modèle mathématique. Traditionnellement, le problème était le manque de mesures nécessaires pour caler les modèles. Cependant, un changement important est en cours avec l'arrivée d'outils provenant du programme du Génome Humain. Ces outils (issus de la bioinformatique, des techniques robotisées de manipulation, des biopuces etc. . .) nous fournissent une très grande quantité de données nouvelles. Une étape naturelle de biostatistique est en cours pour interpréter ces données, viendra ensuite l'étape de modélisation "dynamique" qui devrait conduire à l'apparition de modèles plus réalistes.

## Références

1. J.D.Murray *Mathematical Biology*. Second, Corrected Edition. Biomathematics Texts, Volume 19. Springer.
2. Giorgio Israel *La mathématisation du réel*. Essai sur la modélisation du réel. Seuil.
3. Stephen Jay Gould *Aux racines du temps*. Le livre de poche biblio essais. Grasset.
4. C.C.Lin, L.A.Segel *Mathematics Applied to Deterministic Problems in Natural Sciences* C.L.A.S.S.I.C.S in applied Mathematics. SIAM Philadelphia.
5. Garrett Birkhoff. *Hydrodynamics*. A study in logic, fact and similitude. Princeton University press for university of Cincinnati. 1950.
6. Rutherford Aris *Mathematical Modelling Techniques*. Dover Publications, Inc.
7. René Thom. *Stabilité structurelle et morphogénèse*. Essai d'une théorie générale des modèles. W.A.Benjamin,Inc.
8. Peter Grindrod *Patterns and Waves*. The theory and applications of reaction-diffusion equations. Oxford applied mathematics and computing science series. Clarendon Press. Oxford.
9. John Guckenheimer *Simulation and beyond*. Computing in the 21st century.
10. J.L.Lions *World mathematical year 2000 and computer sciences*.
11. John von Neumann *Method in the physical science*.
12. D. Robert, B. Vian *Eléments de biologie cellulaire*. Doin ed. 1994.
13. site web : <http://esg-www.mit.edu:8001/esgbio/chapters.html>
14. F. Jacob, J. Monod *Genetic regulatory mechanisms in the synthesis of protein* Journal of molecular biology, 3, 318-356. 1961.
15. J.P. Fitch, B. Sokhansanj *Genomic Engineering : moving beyond DNA sequence to function*. proceedings of the IEEE. 2000.
16. S. A. Kauffman *The origins of order*. Oxford university press. 1993.
17. R. Thomas, R. D'Ari *biological feedback* CRC press. 1990.
18. P. D'haeseleer, S. Liang, R. Somogyi *Tutorial : Gene expression data analysis and modeling*. Session on gene expression and genetic networks, Pacific symposium on biocomputing, Hawaii. 1999.
19. L. Glass, J. S. Pasternak *Prediction of limit cycles in mathematical models of biological oscillations*. Bulletin of mathematical biology, volume 40. 1978.
20. E. Farcot *Réseaux de régulation génétique, le modèle linéaire par morceaux*. Rapport de DEA, disponible sur la page : [http://www-lmc.imag.fr/CF/page\\_mash/Accueil.html](http://www-lmc.imag.fr/CF/page_mash/Accueil.html). 2001.
21. J. Della Dora, A. Maignan, M. Mirica-Ruse, S. Yovine *Hybrid computation*. In ISSAC'01, Ontario, Canada. juillet 2001.
22. A. Girard *Etude de systèmes dynamiques hybrides affines par morceaux*. Rapport de DEA, disponible sur la page : [http://www-lmc.imag.fr/CF/page\\_mash/Accueil.html](http://www-lmc.imag.fr/CF/page_mash/Accueil.html). 2001.