

Des réseaux de Pétri stochastiques pour les réseaux génétiques

De nombreux processus biologiques reposent sur des molécules en faible nombre. Cette constatation revêt une importance nouvelle depuis l'apparition de méthodes permettant de procéder à des mesures à l'échelle cellulaire. Que ce soit au niveau de la transcription ou à celui de l'engagement dans une voie de développement aux dépens d'autres voies possibles, de nombreux phénomènes aléatoires ont été rapportés [1-3]. Ces observations macroscopiques à l'échelle cellulaire doivent être rapprochées du comportement intrinsèquement aléatoire des faibles populations moléculaires qui en sont à l'origine. Un gène ce n'est guère plus d'une ou quelques molécules par cellule, bon nombre d'ARN et protéines n'excèdent pas quelques dizaines ou quelques centaines de molécules par cellule.

La prise en compte des fluctuations moléculaires pour la compréhension des phénomènes cellulaires pose des problèmes particuliers. Le gros « défaut » des processus aléatoires par rapport aux modèles déterministes, c'est qu'à une condition initiale éventuellement connue, ne correspond pas un état final unique, mais une probabilité pour chacun des états possibles du système. Ce « défaut » a toutes sortes de conséquences qui rendent la manipulation de ce type de modèles particulièrement difficile. Il devient impossible d'avoir une intuition du comportement du modèle. Du fait des interactions fortes des molécules biologiques entre elles, il est fréquent de se représenter l'évolution d'un système biologique de manière binaire. Sans faire appel à aucune méthode de calcul, il est possible de comprendre, et dans une certaine mesure de prédire, l'évolution de bon nombre de systèmes biologiques.

Ce type d'approximation n'est plus possible lorsque l'on prend en compte la stochasticité des interactions à l'échelle moléculaire ; le nombre de possibilités ne permet plus d'avoir une vision globale du système. Lorsqu'on cherche à calculer leur évolution, les processus aléatoires se révèlent, pour la même raison, extrêmement lourds. Pratiquement, cela signifie que l'évolution de modèles même relativement simples peut tout simplement être hors de la portée des moyens de calculs dont on dispose aujourd'hui.

Un outil conceptuel, les SPN

Un certain nombre d'auteurs ont néanmoins étudié des modèles aléatoires de divers phénomènes biochimiques ou cellulaires comme la décision lyse/lysogénie du phage lambda, la transduction du signal, l'induction d'un gène, etc. La portée de ces travaux a été très limitée par le manque d'outil permettant de procéder à ce type d'analyse. Le formalisme employé peu standardisé rend la lecture des travaux difficile. L'étude se limite généralement à la réalisation d'un programme informatique spécifique pour faire des simulations du

système. Or de tels programmes posent à la fois des problèmes de communication des résultats et de performance. Personne n'a envie d'aller mettre son nez dans un programme bricolé par quelqu'un d'autre et donc le principal résultat du travail n'est pas réellement communicable. Par ailleurs, la réalisation d'un programme d'analyse de processus stochastique relève d'une compétence particulière qu'il ne faut pas minimiser.

Il nous a donc semblé prioritaire de rechercher des outils génériques permettant l'étude du comportement aléatoire des grands réseaux génétiques (ensemble des interactions moléculaires décrivant le fonctionnement coordonné de plusieurs gènes). Mais avant de trouver l'outil informatique, il était nécessaire de trouver l'outil conceptuel adéquat. Les réseaux de Pétri stochastiques (SPN), une formalisation mathématique développée en informatique, ont retenu notre attention [4] car, d'une part, la traduction des termes utilisés en biochimie dans la terminologie SPN est immédiate (Tableau I) et, d'autre part, les SPN ont une représentation graphique très proche de celle habituellement utilisée en bio-

Tableau I	
INTERPRÉTATION MOLÉCULAIRE DE LA TERMINOLOGIE SPN	
Terminologie SPN	Interprétation moléculaire
Place	Espèce moléculaire
Jeton	Molécule
Marquage	Nombre de molécules
Transition	Réaction
Place d'entrée	Réactif
Place de sortie	Produit
Fonction de pondération	Vitesse de réaction
Transition déclenchable	Réaction possible
Déclenchement d'une transition	Événement réactionnel entre molécules

DERNIÈRE HEURE

chimie. L'exemple simple de l'induction d'un gène permet d'illustrer l'analogie qui existe entre réseaux génétiques et SPN. Le réseau de Pétri correspondant à ce système est représenté sur la *figure 1*. On considère que le gène peut être dans deux états (actif et inactif) en fonction, par exemple, de l'état de son promoteur. Sur la *figure 1* les transitions (= réactions) sont symbolisées par des rectangles. L'activation du gène correspond au rectangle λ , et l'inactivation à celui marqué μ . Les deux états possibles du gène sont représentés par les deux cercles marqués « gène actif » et « gène inactif ». La synthèse de protéines est exprimée par la transition v et sa dégradation par la transition δ . Comme la dégradation de la protéine conduit à des produits qui ne sont pas pris en compte dans le modèle, il n'y a pas de flèche qui ressort de la transition δ . L'état initial du système est indiqué par le point dans le cercle « gène inactif ». Au temps 0, le gène est inactif et il n'y a aucune molécule de protéine dans le système. Outre l'analogie graphique et terminologique entre les SPN et les réseaux génétiques, il convient de souligner que la dynamique d'un SPN, c'est-à-dire ses probabilités de transition, peut être définie de

manière cohérente avec la dynamique des réactions chimiques à faible concentration. Le concept est donc intéressant; est-il utile?

Un outil informatique, UltraSAN

Pour qu'il le fût, il fallait trouver des logiciels capables d'étudier le comportement de tels modèles. Il existe un très grand nombre de logiciels du fait que les SPN sont notamment utilisés pour la modélisation des architectures informatiques (ordinateurs, réseaux, etc.). Bon nombre d'entre eux sont gratuits pour une utilisation académique et ils tournent très généralement en environnement UNIX. Nous avons utilisé UltraSAN qui répondait le mieux à nos besoins. Les modèles peuvent être définis à l'aide d'un éditeur graphique. Outre la définition du réseau lui-même, le logiciel invite l'utilisateur à définir ses variables d'intérêt et les études numériques qu'il veut entreprendre. La définition des variables d'intérêt est une étape capitale. On a souvent tendance à se limiter à l'observation d'une variable d'état (le nombre de molécules de telle protéine, par exemple) mais, dans bien des cas, des variables plus complexes peuvent se révéler pertinentes et plus simples

à interpréter. On peut, par exemple, comparer le nombre de molécules de deux espèces différentes, calculer le temps passé avant que tel seuil de concentration soit atteint, etc. Après avoir procédé à toutes les définitions nécessaires, il devient possible de commencer les calculs. Deux grandes familles de calculs sont disponibles, la résolution numérique de la distribution des variables d'intérêt et les simulations. La première famille permet un calcul explicite de la distribution de probabilité des variables alors que la deuxième famille utilise la génération de nombres aléatoires pour estimer cette distribution. Chaque fois qu'elle est possible, la première option est préférable. En fait, c'est surtout la taille du système, le nombre de ses états possibles, qui est limitante. Au-delà de quelques centaines de milliers d'états, seule la simulation est possible. Dans chaque famille, il est possible de procéder aussi bien à des études transitoires (à temps fini) qu'à la caractérisation de l'état stationnaire (équilibre) quand il y en a un. UltraSAN offre plusieurs sortes de solveurs utilisant des techniques numériques différentes parmi lesquelles il faut choisir la plus adaptée selon les caractéristiques fines du modèle. Enfin, UltraSAN se charge de répartir la charge de calcul sur plusieurs stations de travail.

Nous avons utilisé cet outil entre autre pour étudier une version stochastique d'un modèle déterministe de la répliation des plasmides de la famille ColE1 (*figure 2*). Ce modèle comporte dix espèces moléculaires dont 7 formes différentes de complexes impliquant le plasmide, et 18 réactions. Nous avons pu, par simulation, estimer l'évolution temporelle de la distribution du nombre de plasmides au cours d'une génération cellulaire. Mais nous avons aussi pu estimer la fréquence d'événements rares comme l'absence de répliation du plasmide au cours d'une génération. L'étude de l'effet de différents paramètres a permis de mettre en évidence des combinaisons ayant des nombres moyens de plasmides égaux mais des variances différentes ce qui peut avoir des conséquences pour la stabilité du plasmide. UltraSAN permettant

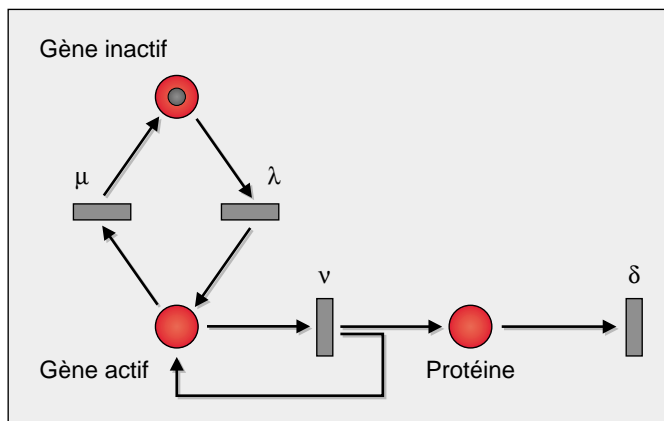


Figure 1. **Analogie entre réseaux génétiques et SPN.** Les représentations SPN sont symbolisées par des cercles, les réactions (transitions) par des rectangles. L'activation du gène correspond au rectangle λ , et l'inactivation à celui marqué μ . La synthèse de protéines est exprimée par la transition v et sa dégradation par la transition δ . L'état initial du système est indiqué par le point (une seule copie du gène) dans le cercle « gène inactif ». Au temps 0, le gène est inactif et il n'y a aucune molécule de protéine dans le système. Comme la dégradation de la protéine conduit à des produits qui ne sont pas pris en compte dans le modèle, il n'y a pas de flèche qui ressort de la transition δ .

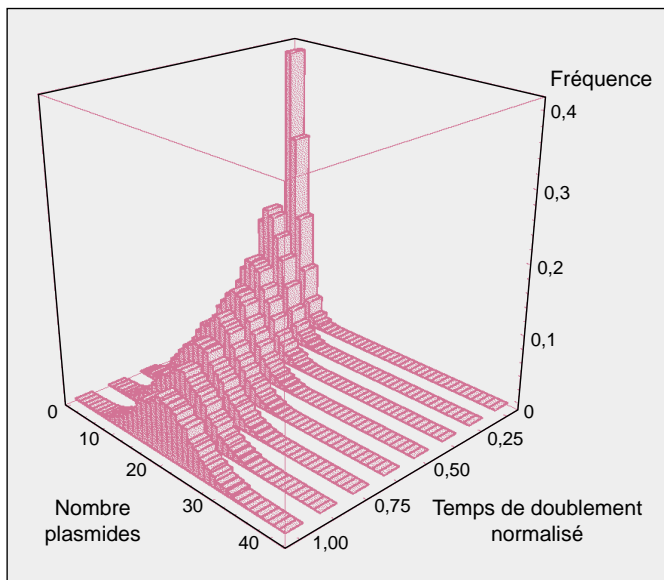


Figure 2. **Distribution du nombre de copies de plasmides dans une bactérie unique au cours de la première génération (intervalles de 10 minutes).** Ces histogrammes sont le résultat de simulations. Le nombre moyen de copies augmente avec le temps mais aussi sa variance. À la fin de la première génération, le nombre moyen de plasmides est de 18,6 et l'écart-type de 5,4. La probabilité que le plasmide ne se reproduise pas au cours de la première génération est intéressante au point de vue biologique car elle est liée à la probabilité pour le plasmide de s'établir dans la population. Elle peut être définie comme variable d'intérêt. Dans ce cas elle est estimée à $4,0 \cdot 10^{-4}$ avec un intervalle de confiance au niveau 95% de $[1,3 \cdot 10^{-4}, 1,1 \cdot 10^{-3}]$.

d'archiver les modèles ; les nôtres sont disponibles sur un site web (www.timc.imag.fr/spn/). Leur étude peut être reproduite ou prolongée par quiconque a accès à UltraSAN ■

RÉFÉRENCES

1. Peccoud J, Viville S. Aspects aléatoires de la dynamique de la différenciation cellulaire. *Med Sci* 1994; 10: 877-83.
2. Kupiec J, Sonigo P. Du génotype au phénotype: instruction ou sélection? *Med Sci* 1997; 13 (suppl 8/9): I-VI.
3. Kupiec J. Un mécanisme de hasard-sélection pourrait expliquer la différenciation cellulaire. *Med Sci* 1998; 14: 115-21.
4. Goss P, Peccoud J. Quantitative modeling of stochastic systems in molecular biology using stochastic Petri nets. *Proc Natl Acad Sci USA* 1998; 95: 6750-5.

Jean Peccoud

University of Washington, Department of Electrical Engineering, Box 352500, Seattle, WA, 98195-2500, États-Unis.

TIRÉS À PART

J. Peccoud.

SOCIÉTÉ DE BIOLOGIE

Cent-cinquantième de la Société de Biologie

19 novembre 1998 (14 h - 18 h)

Séance des membres correspondants de la Société de Biologie

Harold Kalant (Professeur à l'Université de Toronto)

Rôle de la Science dans la formulation de la politique sociale sur la drogue

Christian Doutremépuich (Professeur à l'Université de Bordeaux)

Polymorphisme de l'ADN, applications médico-légales

Ba Denian (Président de l'Académie des Sciences Médicales de Chine)

La recherche en immunologie du cancer en Chine

Chen Zhu (Directeur de l'Institut d'Hématologie de Shanghai II)

Le génome humain en Chine

Susan Hollán (Professeur à l'Institut National d'Hématologie et d'Immunologie de Budapest)

Déficiences en enzymes glycolytiques et neurodégénérescence

Gustav Born (Professeur à l'Institut de Recherche William-Harvey de Londres)

Les effets d'hormones et de médicaments sur la fixation de protéines plasmatiques athérogènes sur les parois artérielles

Hannes Stähelin (Professeur à l'Hôpital Cantonal de l'Université de Bâle)

Vitamine D plasmatique et force musculaire chez les sujets âgés

Francis Karst (Professeur à l'Université de Poitiers)

Biosynthèse et transport des stérols chez la levure

Collège de France

11, place Marcelin-Berthelot – 75005 Paris, France