
Editorial

Depuis plusieurs années, on assiste à une accumulation spectaculaire d'informations moléculaires sur la structure, le développement et le fonctionnement de nombreux organismes vivants. Outre le séquençage de génomes complets déjà terminé ou en cours d'accomplissement, des méthodes à haut débit de caractérisation fonctionnelle des génomes ont été développées. Ainsi, par exemple, les puces à ADN permettent maintenant de mesurer simultanément le niveau d'expression de plusieurs milliers de gènes. Ce nouveau type d'informations qui constitue un complément inestimable aux données produites par les projets de séquençage génomique a amené la biologie fonctionnelle à changer d'échelle.

Les informations moléculaires n'ont pas seulement augmenté en quantité mais aussi en complexité. En particulier, l'étude expérimentale des mécanismes de régulation génétique conduit à la constatation que ces régulations forment des réseaux complexes d'interactions enchevêtrées souvent difficiles à interpréter. D'autre part, pour de nombreux systèmes, la connaissance des interactions demeure encore très fragmentaire. Dans ce contexte, le besoin d'acquérir plus d'informations concernant les interactions, de développer des méthodes de description, de structuration et d'analyse rigoureuses, se fait de plus en plus pressant et exige une démarche pluridisciplinaire regroupant biologistes et théoriciens.

Dans cette perspective, en associant différentes méthodes et techniques issues de l'informatique et de la bio-informatique, des mathématiques et de la physique, la biologie des systèmes cherche à mettre en évidence les propriétés dynamiques et fonctionnelles de réseaux moléculaires aussi complexes que les réseaux de régulation génétique, les réseaux métaboliques, les réseaux de communication intra- et extracellulaire. Elle permet d'aider à comprendre les mécanismes à la base du fonctionnement de ces réseaux de régulation, d'étudier leur stabilité et réponses à différents types de fluctuations et perturbations, et de suggérer des stratégies pour leur contrôle. Elle permet aussi de tester différentes hypothèses pour concevoir de nouvelles expériences. Cette biologie des systèmes qui s'intéresse à la fois à la

structure du système, à sa dynamique et à la construction de modèles «prédictifs», fait largement appel à la modélisation et à la simulation.

L'objectif de ce numéro spécial n'est pas de couvrir toutes les problématiques de la biologie des systèmes mais de présenter, au travers de quelques approches et applications biologiques se rapportant à différents domaines, des travaux de recherche actuels en modélisation et simulation pour la post-génomique.

A l'heure actuelle, la représentation et la description formalisée des interactions entre gènes, leur assemblage en réseaux de régulation, la modélisation et la simulation de leur fonctionnement sont des étapes indispensables à la compréhension de la régulation des gènes et de leur fonction dans le contrôle des processus biologiques. L'approche la plus classique est, sans conteste, basée sur l'utilisation d'équations différentielles. Cependant, le manque actuel de données biologiques précises et quantitatives rend souvent la construction de modèles différentiels difficile et suggère l'adoption de méthodes d'analyse et modélisation qualitatives mais néanmoins rigoureuses. Ainsi, l'« Analyse qualitative de la dynamique de réseaux de régulation génique par des modèles linéaires par morceaux », présentée dans l'article de G. Batt *et al.*, permet de simuler, d'analyser et de prédire le comportement qualitatif d'un réseau de régulation génétique dans différentes conditions, et d'utiliser des outils de « model-checking » pour vérifier que le modèle proposé rend bien compte des propriétés dynamiques connues du système biologique étudié.

Dans leur article « Langages formels dans la machine abstraite biochimique Biocham », L. Calzone, N. Chabrier-Rivier, F. Fages, L. Fosse et S. Soliman montrent également comment la représentation des systèmes biochimiques à l'aide d'une machine abstraite basée sur des systèmes de réécriture, rend possible la vérification automatique de propriétés temporelles exprimées dans une logique temporelle. Ces formalisations peuvent permettre d'utiliser des techniques d'apprentissage automatique pour inférer de nouvelles règles de réaction, estimer les valeurs des paramètres cinétiques, et corriger ou compléter les modèles semi-automatiquement. Il s'agit d'assister le biologiste dans la construction de son modèle, surtout lorsque les connaissances sont qualitatives et incomplètes. Dans ce contexte, la programmation logique par contraintes permet d'inférer des modèles à partir des comportements observés et de déterminer des caractéristiques dynamiques qualitatives comme les états stationnaires, les séparatrices entre bassins d'attraction, les cycles... Ce paradigme de programmation s'applique particulièrement bien dans

le cadre de l'abstraction qualitative des réseaux géniques, et permet d'inférer les valeurs des paramètres pour qu'un système satisfasse certaines propriétés. Cette approche est présentée dans l'article « Modélisation de réseaux biologiques discrets en programmation logique par contraintes », de F. Corblin, E. Fanchon et L. Trilling.

Cependant pour certains systèmes biologiques, l'abstraction qualitative n'est pas aisée à construire ou ne permet pas de répondre aux questions du biologiste. La simulation est alors d'une grande aide dans le processus de compréhension du système puisqu'elle permet de tester *in silico* certaines hypothèses et de montrer comment des propriétés locales peuvent conduire à l'émergence de propriétés globales. Par exemple, dans le cas de l'hématopoïèse qui implique de nombreux processus interdépendants, les mécanismes de régulation de l'auto-renouvellement des cellules souches *in vitro* sont encore mal connus. Vu les difficultés liées aux expériences *in vitro*, la simulation du comportement de ce réseau d'interactions peut aider à comprendre le fonctionnement de ce système biologique complexe. Dans leur article « Modélisation et simulation de la régulation de l'hématopoïèse précoce grâce aux réseaux de Petri hybrides fonctionnels », S. Troncale , F. Tahi , D. Campard et J.-P. Vannier montrent que ces réseaux de Petri sont bien adaptés lorsqu'il faut décrire la production et consommation de ressources et distinguer les processus continus et les événements discrets.

D'autres niveaux d'organisations du vivant font apparaître des phénomènes difficiles à comprendre pour lesquels la simulation peut apporter un nouveau regard. Par exemple, l'organisation tridimensionnelle de structures multiprotéiques, peut être abordée par des modèles multi-agents intégrant une description des déplacements libres des molécules et un modèle énergétique des interactions entre agents, voir « Modélisation cellulaire pour l'émergence de structures multi-protéiques auto-organisées », de A. Coulon, H. Soula, O. Mazet, O. Gandrillon et G. Beslon. A un niveau macroscopique, P. Ballet et P. Tracqui se sont intéressés à la modélisation biomécanique multi-agents de la migration cellulaire dans le cadre de la « migration de cellules virtuelles déformables ». L'approche multi-agents permet ici d'étudier l'influence des interactions mécaniques entre cellules et substrat sur la motilité des cellules. Les déformations et les déplacements cellulaires sont ainsi déterminés par les forces d'adhésion mises en jeu au cours de la migration, ainsi que par la régulation du recrutement de récepteurs transmembranaires.

La biologie des systèmes s'attache aussi à intégrer les données issues des techniques récentes à haut débit, comme les puces à ADN, ChIP on chip, la protéomique... Pour exploiter tout le potentiel de ces mesures à grande échelle et en extraire les caractéristiques biologiquement pertinentes, il est nécessaire de résoudre le problème que pose le grand nombre d'entités par rapport au nombre de données disponibles et de développer des méthodes pour identifier les relations de causalité entre les entités et reconstruire les réseaux de régulation impliqués dans le contrôle de processus cellulaires importants. En particulier, les données d'expression des gènes qui sont générées par les puces à ADN sont un bon point de départ pour la construction de réseaux de régulation transcriptionnelle, comme le montre l'article « Fouille de données pour l'extraction de grands réseaux de régulation génétique » de M. Elati, F. Radvanyi, et C. Rouveirol. Est-il possible d'inférer, à partir de telles données relatives à un contexte biologique, les régulations effectives ? Quels sont les facteurs de transcription qui influencent la transcription d'un gène, et surtout quels sont les modes de coopération/concurrence qui opèrent lorsque plusieurs facteurs de transcription agissent sur un même promoteur ? Les réseaux Bayésiens combinés avec des techniques d'extraction de connaissances permettent d'obtenir des résultats prometteurs.

Mentionnons enfin que l'étude des voies métaboliques est une étape essentielle pour comprendre le comportement d'un organisme. L'utilisation des modes élémentaires qui permettent d'analyser les propriétés d'un tel réseau est cependant limitée par l'explosion combinatoire de leur nombre. L'article « classification des modes élémentaires : application au métabolisme énergétique mitochondrial » de S. Pérès, M. Beurton-Aimar et J.-P. Mazat, propose une première méthode de classification de ces modes élémentaires : l'agglomération de motifs communs permet d'interpréter de grands ensembles de modes élémentaires et de retrouver les fonctions biologiques partagées.

Les systèmes biologiques mettent en jeu un grand nombre d'éléments reliés entre eux par des interactions qui sont non seulement multiples mais souvent fortement non linéaires. De nombreuses études montrent cependant que les réseaux de régulations comportent des groupes de gènes co-exprimés ainsi que des ensembles d'éléments qui présentent un niveau de connectivité significativement supérieur à la valeur moyenne. Ces systèmes sont donc modulaires car les éléments de ces ensembles participent souvent à une fonction biologique commune. Ces modules sont cependant très interdépendants car certains éléments appartiennent à plusieurs modules : une fonction peut être assurée par différents éléments, et inversement, un

même élément peut assurer plusieurs fonctions. Cette particularité des systèmes biologiques est fondamentale. Ainsi, la question de la compositionnalité des «modules» dans un sens plus large et plus souple que son acception en informatique constitue un enjeu important pour la modélisation, la simulation et finalement la compréhension de la complexité du vivant.

Jean-Paul Comet
Université d'Evry Val d'Essonne
IBISC, Evry

Marcelline Kaufman
Université Libre de Bruxelles
Unité de biologie théorique et computationnelle