

Avant-propos

Cédric LHOSSAINE^{1,2} et Élisabeth REMY³

¹ CRIStAL, CNRS, Université de Lille, Lille, France

² Centrale Lille, Lille, France

³ I2M, CNRS, Aix Marseille Université, Marseille, France

La biologie des systèmes, ou biologie systémique, est une approche de la biologie qui propose d’englober toute la complexité des interactions entre les entités biologiques dans un tout systémique dans le but de comprendre l’émergence de propriétés physiologiques ou fonctionnelles. Elle est souvent opposée à l’approche dite « réductionniste » qui consiste à appréhender le vivant par le biais de ses parties étudiées isolément, en réduisant le « tout » en composants de base. La biologie des systèmes s’est particulièrement développée au début des années 1990 avec l’avènement des techniques expérimentales issues des biotechnologies. En effet, le grand plan de séquençage du génome humain et la production de données omiques (génomiques, protéomiques, métabolomiques, etc.) ont fortement contribué à l’émergence de ce domaine par l’identification de nombreuses nouvelles interactions moléculaires. Un des grands défis qui s’est alors posé est le problème de l’intégration de toutes ces informations. Le recours aux outils et méthodes mathématiques et informatiques s’est avéré incontournable, et la biologie des systèmes est intimement liée aux approches interdisciplinaires.

Outil privilégié de la biologie des systèmes, la modélisation permet de saisir la complexité inhérente des systèmes qu’elle étudie. Aux côtés des mathématiques, l’informatique joue un rôle fondamental dans l’exploitation de cet outil. Elle offre en effet une panoplie importante de méthodes de construction de modèles à partir des données, d’analyses et de simulation, jusqu’à la prédiction de nouvelles hypothèses qui suggéreront elles-mêmes de nouvelles expériences.

Cet ouvrage présente des apports des méthodes formelles issues de l'informatique pour la modélisation de la dynamique des systèmes biologiques. Les méthodes qui nous intéressent principalement ici sont les méthodes symboliques, c'est-à-dire celles qui permettent d'établir des propriétés qualitatives des modèles. Ces approches permettent, autant que possible, de s'affranchir des valeurs quantitatives des paramètres des modèles qui sont souvent sujettes à interprétation du fait de la grande variabilité des données expérimentales. L'utilisation des paramètres est encore plus problématique quand ceux-ci ne sont tout simplement pas identifiables à cause d'un nombre trop important de paramètres ou de données insuffisantes. Malgré une perte incontournable de précision, ces méthodes permettent d'extraire les différents comportements qualitatifs robustes de la dynamique du système. Les chapitres 1 (A. Bonnaffoux) et 4 (J. Béal *et al.*) explorent l'inférence de modèles, c'est-à-dire la construction de modèles qualitatifs à partir de différents types de données (qualitatives ou quantitatives, à l'échelle de la population ou de l'individu...).

Ces considérations sur les données et la robustesse des propriétés nous amènent à plusieurs analogies avec les systèmes informatiques, analogies qui ont d'ailleurs souvent inspiré les méthodes symboliques développées pour les systèmes biologiques. Les programmes informatiques sont en effet des objets dynamiques (dans leur exécution) qui dépendent de données (les entrées des programmes) et qui sont potentiellement en interaction avec d'autres programmes. Garantir la bonne exécution des programmes, c'est-à-dire l'absence de bugs ou la conformité avec leurs spécifications, quelles que soient leurs entrées, c'est aussi vérifier des propriétés robustes des programmes vis-à-vis de leurs données. La sémantique des langages de programmation est ce domaine de l'informatique qui a permis de développer de nombreux formalismes de spécification et de preuves de telles propriétés. Parmi ceux-ci, citons la logique de Hoare et les logiques temporelles, utilisées pour inférer des modèles à partir de données cinétiques dans les chapitres 8 (G. Bernot *et al.*) et 3 (T. Ribeiro *et al.*) garantissant ainsi automatiquement la satisfaction des traces expérimentales par les modèles obtenus. La sémantique étudie également le degré d'abstraction des différents langages de modélisation à l'aide des outils puissants d'interprétation abstraite. Cela est notamment abordé dans le chapitre 5 (L. Paulevé) qui introduit un nouveau mode de mise à jour des réseaux booléens leur permettant ainsi d'abstraire les sémantiques d'équations différentielles. L'interprétation abstraite est aussi largement exploitée dans le chapitre 9 (J. Féret) pour étudier statiquement les interactions et liaisons potentielles entre les protéines modélisées dans le langage Kappa.

Étroitement liés à la sémantique des langages de programmation et aux mathématiques, les formalismes logiques apparaissent naturellement, et de façon récurrente, dans les méthodes de modélisation symbolique. On retrouve, déjà citées plus hauts, les logiques de Hoare et temporelles, mais également les logiques booléennes et multi-valuées utilisées non pas pour spécifier des propriétés, mais pour modéliser directement la dynamique des systèmes. L'inférence ainsi que l'analyse de tels modèles

engendrent une explosion combinatoire qui nécessite de s'attarder sur les aspects algorithmiques et de complexité. Le chapitre 2 (C. Frioux et A. Siegel) aborde cette problématique par la programmation par ensembles réponses (ASP). Des logiques beaucoup plus expressives peuvent aussi faire l'objet à la fois de langages de modélisation et de langages de propriétés. C'est en particulier le cas de celles qui sont issues de la logique linéaire et plus généralement des logiques constructives comme le montre le chapitre 6 (E. De Maria *et al.*). Ce constructivisme nous mène aux modèles de calcul largement étudiés en informatique. Dans le chapitre 7 (F. Fages et F. Molina), les auteurs s'interrogent sur le pouvoir calculatoire, au sens informatique, des langages de modélisation issus des réactions.

Ainsi, ces neuf chapitres présentant différentes approches liées à la sémantique, le langage, la modélisation et leur lien avec les données, nous permettent d'aborder des problèmes fondamentaux et défis auxquels nous confronte la biologie des systèmes. Cet ouvrage est organisé en deux parties : la première regroupe les contributions qui s'appuient sur les diverses données accessibles pour construire des modèles, et la seconde des travaux autour des questions de la sémantique et des méthodes formelles.