La machine abstraite biochimique BIOCHAM

François Fages

Projet Contraintes
INRIA Rocquencourt
BP 105 F-78153 Le Chesnay Cedex
Francois.Fages@inria.fr

A la fin des années 90, le front de la recherche en bioinformatique a évolué, passant de l'analyse de la séquence génomique à l'analyse de données diverses produites en masse par les technologies dites "post-génomiques", telles que l'expression des ARN, la production des protéines et les intéractions protéine-protéine. La biologie moléculaire s'est engagée dans un effort systématique d'élucidation des processus cellulaires de haut niveau, en termes de leur base biochimique à l'échelle moléculaire, qui soulève le besoin d'un effort parallèle important sur la représentation formelle des processus biologiques. La biologie systémique est un domaine multidisciplinaire où biologistes, informaticiens, logiciens, mathématiciens et physiciens se rassemblent pour tenter de passer les murs de complexité permettant de raisonner sur les systèmes biologiques dans leur globalité. Afin de passer à l'échelle, il y a un sentiment général que la biologie a besoin d'être en partie réinventée, et que, au delà des outils qu'elle peut offrir aux biologistes, l'informatique théorique a beaucoup a apporter en termes de concepts et de méthodes.

En 2002, à la suite de travaux pionniers de Regev, Silverman et Shapiro [1] utilisant le π-calcul de Milner pour modéliser des voies de signalisation dans la cellule, une action de recherche collaborative ARC CPBIO [1] sur "calculs de processus et biologie des réseaux moléculaires", a été lancée visant à :

- identifier dans la famille des langages venant de la théorie de la concurrence (π-calcul et ses dérivés) et de la programmation logique (langages concurrents avec contraintes et leurs extensions au temps discret et continu) les ingrédients d'un langage adapté à la représentation modulaire et multi-échelle des processus biologiques ;
- fournir une série d'exemples de systèmes bio-moléculaires transcrits dans des langages formels, et un ensemble de questions biologiques sur ces modèles ;
- concevoir et appliquer des outils automatisés de raisonnement formel pour la simulation, l'analyse et l'interrogation des modèles.

La machine abstraite biochimique BIOCHAM [2] [2] issue de ces travaux, fournit une sémantique précise aux réseaux d'intéractions bio-moléculaires dans la cellule, à deux niveaux d'abstraction (s'affrichissant de l'espace) : le niveau quantitatif des concentrations moléculaires, et le niveau qualitatif booléen décrivant la présence ou l'absence de molécules. En se fondant sur des sémantiques de Kripke dans les deux cas [3, 4], le logiciel BIOCHAM offre :

- un langage simple de règles de réactions bio-moléculaires, autorisant des schémas de réaction, ainsi que des expressions cinétiques continues lorsque les données numériques sont disponibles ;
- un simulateur numérique (implantant la méthode de Rosenbrock pour les systèmes hautement non-linéaires) et un simulateur non-déterministe booléen ;
- un langage de requête original fondé sur la logique temporelle [5] (*Computation Tree Logic, CTL,* pour les modèles qualitatifs, et *Linear Time Logic, LTL,* avec contraintes pour les modèles numériques) permmetant d'exprimer des propriétés biologiques telles que l'accessibilité, les points de passage obligés (*checkpoints*), les oscillations et la stationnarité ;

[1] `http ://contraintes.inria.fr/cpbio`
[2] `http ://contraintes.inria.fr/BIOCHAM`

– enfin, un système d'apprentissage automatique d'inférence de règles de réactions bio-moléculaires à partir de propriétés temporelles observées [6].

BIOCHAM partage plusieurs similitudes avec le système *Pathway Logic* [7] implanté en Maude. Les deux systèmes ont une syntaxe algébrique et sont des langages de règles. Une différence est l'utilisation dans BIOCHAM de la logique CTL qui permet d'exprimer une plus grande variété de requêtes biologiques, ainsi que la prise en compte des dynamiques continues dans les modèles quantitatifs. BIOCHAM est complétement implanté en Prolog et interfacé au vérificateur symbolique de modèles NuSMV [8] pour évaluer les requêtes CTL dans des modèles de grande taille, comprenant plusieurs centaines de variables. Les modèles BIOCHAM peuvent être importés et exportés dans un format standard de biologie des systèmes : le *Systems Biology Markup Language SBML*. La bibliothèque de modèles de systèmes biologiques en cours de constitution, CMBSlib[3] [9], permet de comparer les modèles tout comme les formalismes.

L'expressivité et l'efficacité du système BIOCHAM ont été montrées dans [3, 10] en fournissant une contrepartie formelle de la carte de Kohn [11] décrivant le contrôle du cycle cellulaire chez les mammifères et en important d'autres modèles de la littérature. Ceci permet de transformer une connaissance statique des capacités d'intéraction en un système de transition concurrent incorporant une dynamique qualitative dans le cas des modèles booléens, ou en un système d'équations différentielles (non linéaires) dans le cas des modèles cinétiques. L'efficacité des outils de vérification symbolique de modèles dans ce contexte a été montrée sur l'exemple du cycle cellulaire des mammifères, comprenant 147 schémas de règles, 2733 règles expansées, 165 protéines et gènes élémentaires, 500 variables et 2^{500} états.

Le niveau d'abstraction booléen permet de raisonner sur tous les comportements possibles du système, en l'absence de modèle cinétique. L'interrogation en logique temporelle de la dynamique qualitative du système permet ainsi d''analyser des systèmes pour lesquels on ne connaît pas tous les paramètres numériques nécessaires à l'élaboration d'un modèle cinétique. Par ailleurs, les propriétés temporelles du système biologique étudié (dans des organismes sauvages ou mutés) peuvent être également formalisées en CTL, et tournées en des spécifications formelles attendues du modèle. C'est cette voie que nous explorons dans [6], en développant un algorithme d'apprentissage ad-hoc de règles de réactions à partir de propriétés temporelles formelles, et en l'appliquant à la complétion ou à la correction d'un modèle partiel, ainsi qu'à la recherche de cibles thérapeutiques.

Références

[1] Regev, A., Silverman, W., Shapiro, E.Y. : Representation and simulation of biochemical processes using the pi-calculus process algebra. In : Proceedings of the sixth Pacific Symposium of Biocomputing. (2001) 459–470

[2] Fages, F., Soliman, S., Chabrier-Rivier, N. : Modelling and querying interaction networks in the biochemical abstract machine BIOCHAM. Journal of Biological Physics and Chemistry **4** (2004) 64–73

[3] Chabrier-Rivier, N., Chiaverini, M., Danos, V., Fages, F., Schächter, V. : Modeling and querying biochemical interaction networks. Theoretical Computer Science **325** (2004) 25–44

[4] Chabrier-Rivier, N., Fages, F., Soliman, S. : The biochemical abstract machine BIOCHAM. In Danos, V., Schächter, V., eds. : CMSB'04 : Proceedings of the second Workshop on Computational Methods in Systems Biology. Volume 3082 of Lecture Notes in BioInformatics., Springer-Verlag (2004) 172–191

[5] Clarke, E.M., Grumberg, O., Peled, D.A. : Model Checking. MIT Press (1999)

[3] `http ://contraintes.inria.fr/CMBSlib`

[6] Calzone, L., Chabrier-Rivier, N., Fages, F., Gentils, L., Soliman, S. : Machine learning biomolecular interactions from temporal logic properties. In Plotkin, G., ed. : CMSB'05 : Proceedings of the third Workshop on Computational Methods in Systems Biology. (2005)

[7] Eker, S., Knapp, M., Laderoute, K., Lincoln, P., Meseguer, J., Sönmez, M.K. : Pathway logic : Symbolic analysis of biological signaling. In : Proceedings of the seventh Pacific Symposium on Biocomputing. (2002) 400–412

[8] Cimatti, A., Clarke, E., Enrico Giunchiglia, F.G., Pistore, M., Roveri, M., Sebastiani, R., Tacchella, A. : Nusmv 2 : An opensource tool for symbolic model checking. In : Proceedings of the International Conference on Computer-Aided Verification, CAV'02, Copenhagen, Danmark (2002)

[9] Soliman, S., Fages, F. : CMBSlib : a library for comparing formalisms and models of biological systems. In Danos, V., Schächter, V., eds. : CMSB'04 : Proceedings of the second Workshop on Computational Methods in Systems Biology. Volume 3082 of Lecture Notes in BioInformatics., Springer-Verlag (2004) 231–235

[10] Chabrier, N., Fages, F. : Symbolic model cheking of biochemical networks. In Priami, C., ed. : CMSB'03 : Proceedings of the first Workshop on Computational Methods in Systems Biology. Volume 2602 of Lecture Notes in Computer Science., Rovereto, Italy, Springer-Verlag (2003) 149–162

[11] Kohn, K.W. : Molecular interaction map of the mammalian cell cycle control and DNA repair systems. Molecular Biology of the Cell **10** (1999) 703–2734