From Syntax to Semantics in Systems Biology Towards Automated Reasoning Tools

François Fages

Projet Contraintes, INRIA Rocquencourt,
BP105, Le Chesnay Cedex 78153, France
Francois.Fages@inria.fr
http://contraintes.inria.fr

Mathematical biology has for a long time investigated the dynamics of biomolecular systems by developing numerical models involving (highly non-linear) differential equations and using tools such as Bifurcation Theory for estimating parameters [1]. Mathematical biology provides a firm ground for the numerical analysis of biological systems. However, state-of-the-art quantitative models can hardly be re-used and composed with other models in a systematic fashion, and are limited to a few tenths of variables [2].

Qualitative models of bio-molecular interactions constitute the core of nowadays cell systems biology. Interaction diagrams are the first tool used by biologists to reason about complex systems. The accumulation of knowledge on gene interaction and pathways is currently entered in databases such as KEGG[3], EcoCyc [4], etc. in the form of annotated diagrams. Tools such as BioSpice, Gepasi, GON, E-cell, etc. have been developed for making simulations based on these databases when numerical data is present. Furthermore the interoperability between databases and simulation tools is now possible with standard exchange formats such as the Systems Biology Markup Language SBML [5].

These advances give more acuity to at least three challenges for systems biology:

– One big challenge is the modularity and compositionality of biological models. It is not an easy task today to combine given models of different pathways sharing some molecular components in a given organism, and obtain a mixed model of the complex system. This is a restriction to the re-use of models in systems biology and to their direct use in any application.

– Another challenge is to go beyond simulations and use models to automate various forms of biological reasoning, in purely qualitative models too. Computer aided inference of interaction networks, or computer aided drug target discovery, need non-trivial automated reasoning tools to assist the biologist.

– A third challenge for systems biology will be the possibility to change the way molecular cell biology is taught by making it more formal, putting formal models and tools at the center of the courses. Having a common syntax, one way to approach these challenges is to develop precise semantics of interaction diagrams and build formal methods and tools to reason about them. Our project with the Biochemical Abstract Machine[1] [6], started in 2002, is such an attempt. Based on formal semantics of molecular interactions, Biocham offers:

---

[1] http://contraintes.inria.fr/BIOCHAM

– a compositional rule-based language for modeling biochemical systems, allowing patterns and kinetic expressions when numerical data are available [7];
– a numerical simulator and a non-deterministic boolean simulator;
– a powerful query language based on temporal logic for expressing biological queries such as reachability, checkpoints, oscillations or stability [8];
– a machine learning system to infer interaction rules and parameter values from observed temporal properties [9].

An important characteristic of a language for modeling complex systems is that one may have to consider several semantics corresponding to different attraction levels. It is indeed important to provide the ability to skip from one level of abstraction to another one, and thus to combine several semantics in the language. Perhaps the most realistic semantics is to consider a population of molecules, and consider stochastic simulation as introduced very nicely in the 70s by Gillespie [10]. In Biocham we currently combine two abstraction levels: the molecular concentration semantics and a boolean semantics for reasoning simply about the presence or absence of molecules.

At the concentrations semantics level, a set of reaction rules given with kinetic expressions compiles into a set of (non-linear) ordinary differential equations. At the Boolean semantics, the rules compile into a highly non-deterministic concurrent transition systems which gives an account for all possible competing interactions. The most original feature of BIOCHAM is its use in both cases of a powerful language based on temporal logic to formalize the biological properties of the model. In the (non-deterministic) Boolean semantics, we use the Computation Tree Logic CTL [11], while in the (deterministic) concentration semantics, we use a Linear Time Logic LTL with constraints. We have shown that these temporal logics are sufficiently expressive to formalize a very rich set of biological properties such a reachability, checkpoints, stability or oscillations, either qualitatively or quantitatively[8, 7], and in large models of the cell cycle of up to 500 variables [12]. The machine learning system of Biocham builds on these formal languages and semantics to discover new reaction rules, or fit parameter values, starting from a set of observed properties of the system formalized in temporal logic [9].

Our current work aims at developing a modular modeling discipline for quantitative models based on a full decomposition of interaction rules. Following this approach, we are currently developing a mixed model of the cell cycle and the circadian cycle with applications to cancer therapies.

Acknowledgments

References

1. Segel, L.A.: Modeling dynamic phenomena in molecular and cellular biology. Cambridge University Press (1984)
2. Chen, K.C., Calzone, L., Csik´asz-Nagy, A., Cross, F.R., Gy¨orffy, B., Val, J., Nov`ak, B., Tyson, J.J.: Integrative analysis of cell cycle control in budding yeast. Molecular Biology of the Cell 15 (2005) 3841–3862
3. Kanehisa, M., Goto, S.: KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research 28 (2000) 27–30
4. Keseler, I.M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I.T., Peralta-Gil, M., Karp, P.D.: EcoCyc: a comprehensive database resource for escherichia coli. Nucleic Acids Research 33 (2005) D334–D337
5. Hucka, M., et al.: The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. Bioinformatics 19 (2003) 524–531
6. 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
7. Chabrier-Rivier, N., Fages, F., Soliman, S.: The biochemical abstract machine BIOCHAM. In Danos, V., Sch¨achter, 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
8. 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
9. Calzone, L., Chabrier-Rivier, N., Fages, F., Gentils, L., Soliman, S.: Machine learning bio-molecular interactions from temporal logic properties. In Plotkin, G., ed.: CMSB'05: Proceedings of the third Workshop on Computational Methods in Systems Biology. (2005)
10. Gillespie, D.T.: General method for numerically simulating stochastic time evolution of coupled chemical-reactions. Journal of Computational Physics 22 (1976) 403–434
11. Clarke, E.M., Grumberg, O., Peled, D.A.: Model Checking. MIT Press (1999)
12. Chabrier-Rivier, N., Chiaverini, M., Danos, V., Fages, F., Sch¨achter, V.: Modeling and querying biochemical interaction networks. Theoretical Computer Science 325 (2004) 25–44