The Systems Biology Markup Language (SBML): A Standard for the Exchange of Biochemical Network Models

Andrew Finney, Michael Hucka, Herbert Sauro,
Hamid Bolouri, John Doyle, Hiroaki Kitano
ERATO Kitano Systems Biology Project
Control and Dynamical Systems, MC 107-81
California Institute of Technology, Pasadena, CA
sysbio-team@caltech.edu

ABSTRACT

Over the years, a variety of biochemical network modeling packages have been developed and used by researchers interested in understanding cellular networks. Unfortunately, there have been few attempts so far to devise a common standard for exchanging biochemical network models.

The lack of a common standard for exchanging models poses two problems for researchers. First, models cannot be shared directly between tools because the file format used by each tool is unique and nontransferable. This makes it difficult if not impossible for a researcher to exchange models. Second, when a simulator is upgraded or is no longer supported, models developed in the "old" system become stranded and unusable in any other simulator. This later scenario has already happened on a number of occasions, with the resulting loss of the model to the scientific community. With the recent proliferation of new biochemical simulators, this situation can only get worse.

In an attempt to address these issues, together with our collaborators, the groups developing the simulation packages BioSpice [1], DBsolve [4], E-Cell [9], Gepasi [6], Jarnac [7], VCell [8] and StochSim [2], we propose a standard model representation language for storing biochemical models. We call this proposed standard the *Systems Biology Markup Language* (SBML). SBML is based on XML [3]. We have kept the base definition of the language as simple as possible, so that simulator developers will not find it too difficult to implement support for SBML in their tools. This base definition, called *SBML Level 1*, is the result of merging modeling-language features from a number of existing software packages, and encompasses the minimal information required for biochemical models.

SBML contains XML elements to represent compartments, species and reactions. Reaction elements contain a rate law and lists of products and reactants with their respective stoichiometry. Species are located in specific compartments. Optional features include unit definitions, parameters and rules (constraints).

Additional structures and facilities will be added to SBML in subsequent levels. By freezing subsets of features in SBML definitions at incremental levels, we hope to provide the community with stable standards to which software authors can design to, while at the same time allowing the simulation community to gain experience with the language definitions before introducing new elements. Documents describing SBML are currently available from our project web site at `http://www.cds.caltech.edu/erato/`.

1. REFERENCES

[1] A. P. Arkin. *Simulac* and *Deduce*. `http://gobi.lbl.gov/~aparkin/Stuff/Software.html`, 2001.

[2] D. Bray et al. *StochSim*. `http://www.zoo.cam.ac.uk/comp-cell/StochSim.html`, 2001.

[3] T. Bray et al. Extensible Markup Language (XML) 1.0, W3C Recommendation 10-February-1998, `http://www.w3.org/TR/1998/REC-xml-19980210`, 1998.

[4] I. Goryanin et al. Mathematical simulation and analysis of cellular metabolism and regulation. *Bioinformatics*, 15(9):749–758, 1999.

[5] M. Hucka et al. The ERATO Systems Biology Workbench: An Integrated Environment for Multiscale and Multitheoretic Simulations in Systems Biology. In *Foundations of Systems Biology*, chapter 6. 2001. In press.

[6] P. Mendes. *Trends in Biochemical Sciences*, 22:361–363, 1997.

[7] H. M. Sauro. Jarnac: A system for interactive metabolic analysis. In J.-H. S. Hofmeyr et al, editors, *Animating the Cellular Map: Proceedings of the 9th International Meeting on BioThermoKinetics*. Stellenbosch University Press, 2000.

[8] J. Schaff et al. Physiological modeling with the Virtual Cell framework. In M. Johnson and L. Brand, editors, *Methods in Enzymology*, volume 321, pages 1–23. Academic Press, San Diego, 2000.

[9] M. Tomita et al. *Bioinformatics*, 15(1):72–84, 1999.