Systems biology markup language: Level 2 and beyond

A. Finney*[1] and M. Hucka†

*Science and Technology Research Centre, University of Hertfordshire, Hatfield Campus, College Lane, Hatfield, Hertfordshire AL10 9AB, U.K., and †Control and Dynamical Systems, Mail Code 107-181, California Institute of Technology, Pasadena, CA 91125-8100, U.S.A.

Abstract

The SBML (systems biology markup language) is a standard exchange format for computational models of biochemical networks. We continue developing SBML collaboratively with the modelling community to meet their evolving needs. The recently introduced SBML Level 2 includes several enhancements to the original Level 1, and features under development for SBML Level 3 include model composition, multistate chemical species and diagrams.

Introduction

A range of biochemical network modelling packages is used by researchers interested in understanding biochemical networks. This diversity of tools brings with it several problems, including difficulties in moving models between packages and a lack of mechanisms for publishing models in electronic form. In an attempt to overcome these problems, we have developed an exchange language, the SBML (systems biology markup language) [1], with the help of a community of software developers.

We are only aware of one other language, CellML [2], that is designed for the exchange of biochemical network models. CellML is built around an approach of composing systems of equations. By contrast, SBML provides constructs that are similar to the object models used in packages specialized for simulating and analysing biochemical networks. These differences notwithstanding, the SBML and CellML efforts share much in common, and the development of SBML has benefited from discussions with the developers of CellML.

SBML is being developed in *levels*, where each level extends the set of features of the language. By freezing SBML development at incremental levels, software authors can work with stable standards and gain experience with the standard before further development. The separate levels of SBML are intended to coexist.

Changes introduced in Level 2

The success of Level 1 has led to requests for new language features from the community of users and developers. The final form of Level 2 was agreed upon by the community in June 2003 (see http://www.sbml.org/). The main features distinguishing SBML Levels 1 and 2 are the following.

In SBML Level 1, mathematical formulae are expressed using text strings. Level 2 uses MathML (the mathematical markup language) [3] instead, enabling the expression of more complex mathematical formulae and providing greater compatibility with CellML.

SBML Level 2 provides facilities for adding metadata to a model, using the same approach as in CellML.

Level 2 provides a facility for defining named mathematical functions; these can then be used in mathematical expressions throughout the model. By contrast, Level 1 does not support the ability to define new functions and instead specifies a limited dictionary of predefined functions.

Level 2 provides a construct for specifying delay functions. These are useful for representing biological processes having a delayed response, but where the details of the processes and the actual delay mechanism are not relevant to the operation of the model.

Level 2 provides a facility for defining discrete events that can occur at defined transitions in a model's state and affect the values of model variables.

In addition to these major changes, several other smaller refinements are present in Level 2.

Proposals for Level 3

In partnership with the modelling community, we are now developing Level 3. Although SBML has been successfully adopted by many groups developing systems biology software, there exist packages that support classes of models which presently cannot be encoded in Level 2. Level 3 is intended to provide support for these tools. The community plans to introduce features in Level 3 that will add support for: (i) composing models from component submodels; (ii) describing states and interaction of components of species in terms of rules rather than explicit enumerations of all possible combinations; (iii) describing two-and three-dimensional spatial geometries; (iv) describing model diagram layouts; (v) enabling parameter and initial condition values to be

Key words: collaboration, language, model, network, simulation, standard.

Abbreviation used: SBML, systems biology markup language.

[1]To whom correspondence should be addressed (e-mail A.Finney@herts.ac.uk).

defined separately from models and (vi) allowing for alternative mathematical representations of reactions.

Discussion

One measure of success for a data format is its degree of acceptance among software developers, and by this measure, SBML has been quite successful. SBML is supported by many software packages, including Cellerator [4], Cytoscape [5], Gepasi [6], Jarnac [7], JigCell [8], NetBuilder [9], Systems Biology Workbench [10], StochSim [11] and Virtual Cell [12]. Further the Defense Advanced Research Program Agency BioSPICE and International *E. coli* Association consortia are currently using SBML as their *de facto* standard model definition language.

The development of SBML has been funded by the Japan Science and Technology Corporation's Exploratory Research for Advanced Technology program and the U.K. Biotechnology and Biological Sciences Research Council. We developed SBML Level 2 with the help of many people including members of the BioSPICE Model Definition Language, sysbio and sbml-discuss mailing lists.

References

1 Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A. et al. (2003) Bioinformatics **19**, 524–531
2 Hedley, W.J., Nelson, M.R., Bullivant, D.P. and Nielson, P.F. (2001) Phil. Trans. R. Soc. Lond. A **359**, 1073–1089
3 Sandhu, P. (2002) The MathML Handbook, Charles River Media Hingham
4 Shapiro, B.E. and Mjolsness, E.D. (2001) in The Proceedings of the Second International Conference on Systems Biology (ICSB2001) (Kitano, T.-M. et al., eds.), pp. 342–351, Omnipress, Madison
5 Ideker, T., Ozier, O., Schwikowski, B. and Siegel, A.F. (2002) Bioinformatics **18** (suppl. 1), S233–S240
6 Mendes, P. (1997) Trends Biol. Sci. **22**, 361–363
7 Sauro, H.M. (2000) in Animating the Cellular Map: Proceedings of the 9th International Meeting on BioThermoKinetics (Snoep, J.-H. et al., eds.), pp. 221–228, Stellenbosch University Press, Stellenbosch
8 Vass, M., Shaffer, C.A., Tyson, J.J., Ramakrishnan, N. and Watson, L.T. (2003) Bioinformatics, in the press
9 Brown, C.T., Rust, A.G., Clarke, P.J.C., Pan, Z., Schilstra, M.J., De Buysscher, T., Griffin, G., Wold, B.J., Cameron, R.A., Davidson, E.H. and Bolouri, H. (2002) Dev. Biol. **246**, 86–102
10 Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C. and Kitano, H. (2002) in Pacific Symposium on Biocomputing 2002 (Klein, R.B., et al., eds.), pp. 450–461, World Scientific, River Edge
11 Morton-Firth, C.J. and Bray, D. (1998) J. Theoret. Biol **192**, 117–128
12 Schaff, J., Slepchenko, B. and Loew, L.M. (2000) Methods Enzymol. **321**, 1–23

Received 16 June 2003