A structured approach for modelling of integrated systems in biology

Nicolas Parisey[1*], Marie Beurton-Aimar[12*], S. Randall Thomas[3*]
[*]All authors contributed equally to this work

[1]Mitochondrial Physiopathology Lab. U688 - University Bordeaux 2
[2]LaBRI UMR 5800 - University Bordeaux 1
[3]LAMI UMR 8042 - University Evry

Data and information-flow processes applied to the description of living systems present a new challenge to computational biology. The level of complexity attained by biological knowledge and the development of ever more sophisticated informatics tools for the modelling and simulation of biological processes have revealed the need for methods to structure the manipulated data (where "data" is taken in its most general sense) in a more formal way. Structured approaches like Object Oriented Methods (OOM) allow the definition of data types and relations among these types in order to qualify them semantically. Thus structured, it becomes possible to design computer programs that can automatically interpret the information.

Data types and relations between them can be mapped to arborescences described by a structuration language. Following the emergence of XML (eXtended Markup Language), which is such a language, two XML-languages for biology, SBML (Systems Biology Markup Language) and CellML have attracted a growing community of users in the bioinformatics/ bio-modelling/ systems biology community.

XML provides a means to supplement a text file with any relevant information describing its contents, its origin, its intended use, etc. That is, an XML file is a self-describing ascii file. XML is a structured language with marks (called `tags`) which wrap the file's contents with defining concepts or object classes. SBML [1] and CellML [2] are augmented versions of XML that specify `tags` for descriptions of mathematical models of biological processes and specifically for metabolic networks, kinetic equations, and so on.

The principle of XML is to define a *schema* which specifies the nested `tags` and their hierarchical tree structure. Users are free to define the `tags` as they see fit. However, when a community of users wishes to share files with a common purpose and with a minimum of ambiguity, they agree on a standardized `tag` structure, fixing its definition in a shared formal description, of which there are two standard types: Data Type Definition (DTD) files, and XML Schema. Many thousands of such specifications have been developed for a myriad of applications, of which SBML and CellML are two typical examples. They both

have specific DTD and XML *Schema* definitions, which have gone through several revisions, the latest being available on the respective websites. It is the standardization of these publically available definitions that makes it possible for anyone to design a software program taking one or both of these description formats as input (or producing them as an output file).

As with any standardized language, new situations arise that are not well captured by the current schemas, which leads to continual pressure to extend them to the new situations. Such extensions are a natural part of the evolving use of such systems, but they bring with them the inevitable problem of backward compatibility with previous versions.

Our particular concern is with the modelling of integrated biological systems at not only the cellular and sub-cellular levels (which are essentially the focus of present versions of SBML and CellML) but also of multi-cellular systems, tissues, and even organs. We acutely feel the need to standardize the model-descriptions of such systems using a markup language, but the current versions of SBML and CellML are inadequate to this task. We have thus formed a French Working Group on Markup Languages for Integrated Systems Modelling to characterize these shortcomings and to suggest relevant extensions and improvements to the developers of these two existing MLs (with whom we have established a dialogue).

We present two specific but simple examples drawn from our work (in very different fields) which typify the features we most miss in SBML and CellML: 1) epithelial transport, with transcellular and paracellular transport of water and several solutes; and 2) mitochondrial metabolism modelling. Extension of SBML and CellML to cover these two cases would in fact render them adequate for a much larger class of models which are currently outside their scope.

References

[1] M Hucka and et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, Mar 2003.

[2] Autumn A. Cuellar, Catherine M. Lloyd, Poul F. Nielsen, David P. Bullivant, David P. Nickerson, and Peter J. Hunter. An overview of cellml 1.1, a biological model description language. *SIMULATION: Transactions of The Society for Modeling and Simulation International*, 2003.