Bioinformatics Advance Access published November 17, 2009

Annotation and merging of SBML models with semanticSBML

Falko Krause, Jannis Uhlendorf, Timo Lubitz, Marvin Schulz, Edda Klipp, and Wolfram Liebermeister*

Humboldt-Universität zu Berlin, Theoretische Biophysik, Invalidenstr. 42, D-10115 Berlin

Associate Editor: Prof. Alfonso Valencia

ABSTRACT

Summary: SBML is the leading exchange format for mathematical models in Systems Biology. Semantic annotations connect model elements with external knowledge via unique database identifiers and ontology terms, enabling software to check and process models by their biochemical meaning. Such information is essential for model merging, one of the key steps toward the construction of large kinetic models. The tool semanticSBML helps users to check and edit MIRIAM annotations and SBO terms in SBML models. Using a large collection of biochemical names and database identifiers, it supports modellers in finding the right annotations and in merging existing models. Initially, an element matching is derived from the MIRIAM annotations and conflicting element attributes are categorised and highlighted. Conflicts can be resolved automatically or manually, allowing the user to control the merging process in detail.

Availability: SemanticSBML is free software written in Python and released under the GPL 3. A Debian package, a source package for other Linux distributions, a Windows installer, and an online version of semanticSBML with reduced functionality are available at www.semanticsbml.org. A preinstalled version can be found on the Linux Live DVD SB.OS, available at www.sbos.eu.

Contact: wolfram.liebermeister@biologie.hu-berlin.de

1 INTRODUCTION

One of the big challenges of Systems Biology - and an important example of model reuse - is the construction of complex models from smaller existing models (Snoep *et al.* (2006); Liebermeister (2008)). A variety of Systems Biology models have been published in model repositories (Olivier and Snoep (2004); Le Novère *et al.* (2006)) in the exchange format SBML (Systems Biology Markup Language (Hucka *et al.* (2003))). To merge several SBML models, duplicate elements have to be matched and contradicting statements need to be found and resolved (Fig. 1). For this purpose, the biochemical meaning of the model elements should be specified in machine-readable form. Several scientific journals and research consortia have adopted the MIRIAM guidelines (Le Novère *et al.* (2005)), which put forward such semantic annotations. A fact like "The species `H2O` in model XYZ represents water" can be declared within the model by MIRIAM-compliant annotations that link model elements to the entries of data repositories (e.g. *via* the ChEBI entry `CHEBI:15377` for water). The `SBOterm` attributes, which point to the Systems Biology Ontology (Le Novère (2006)), can provide additional semantic information, e.g. on kinetic equations and on the biochemical roles of substances therein.

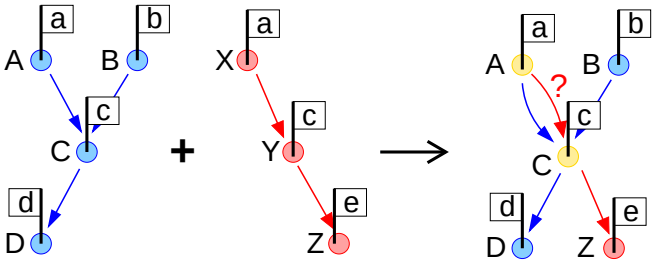

Fig. 1. Model merging. To merge two pathway models (circles: substances; arrows: chemical reactions), duplicate elements have to be joined (magenta circles). Elements are matched by their annotations (shown in little flags) because their names (beside the circles) can differ between the models. Some decisions depend on the modeller's interpretation of the models. For instance, the reaction a → c appears in both models. If it is catalysed by different isoenzymes, it should appear twice in the merged model. Otherwise, it should be treated as a duplicate and be merged.

Computer tools could simplify the tedious work of model merging: they can display SBML models in a human-readable form, execute routine steps like the normalisation of names, and check the merged model for syntactic and mathematical correctness. Some decisions, however, depend on the modeller's interpretation of the models and therefore require user interaction (see Fig. 1). Thus, a software for model merging should not just merge models automatically, but serve as an editor that supports modellers in taking the right decisions.

2 APPROACH

Although tools for model annotation, checking, and building (e.g., SBMLeditor, COPASI, CellDesigner) as well as for merging (SBMLmerge, the predecessor of semanticSBML) already exist, semanticSBML is the first comprehensive software combining all these abilities for SBML models with a focus on semantic annotations. Its graphical user interface provides a simple

*to whom correspondence should be addressed

and intuitive way to edit SBO terms and MIRIAM-compliant RDF annotations (Le Novère *et al.* (2005, 2006)), supporting the BioModels.net qualifiers (e.g. `versionOf`) and element identifiers from all data repositories specified in the MIRIAM resources (Laibe and Le Novère (2007)). For a number of databases (including KEGG, ChEBI, and Gene Ontology), identifiers can be retrieved by the corresponding names. A query for the name "water", for instance, will yield a list of database identifiers that could be used for annotation. Annotations can be copied between models. If an SBML file contains known names (for instance, within the `name` attributes), annotations can also be set automatically. Furthermore, the user can check annotated SBML models for semantic validity (e.g. the conservation of atom numbers within a reaction, or the physical separation of compartments) and display their network structure. Fully annotated SBML models can be built from a list of chemical reactions or KEGG reaction identifiers.

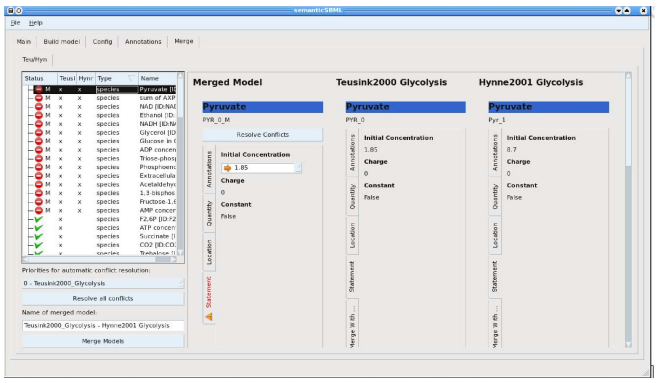

Fig. 2. Model merging with semanticSBML (screenshot). On the left, elements from the input models and the output model (in its preliminary form) are aligned to each other. Details about selected elements are shown on the right. The user can edit their properties and change the matching between elements until the models are ready for merging.

In model merging, semanticSBML compares the MIRIAM annotations of two or more input models and suggests a preliminary version of the merged model, which provides a starting point for manually completing the element matching. To resolve conflicting element properties, e.g. different initial concentrations for a species, conflicts are highlighted and categorised. This allows the user to recognise the severeness of a conflict. The user can navigate through the models, change the matching of model elements, check the conflicts between them, and decide how they should be resolved. Alternatively, the software can resolve all conflicts automatically according to a model priority list defined by the user.

3 DISCUSSION

Mathematical modelling is an art rather than just a stereotypical work. Models are made for specific purposes, based on different physical and biochemical assumptions, and intentionally simplified if the data do not suffice for a more detailed model. If models have been built by different researchers with different assumptions and intentions, merging them can be hard. In particular, models that contain lumped substances or reactions may simply not fit. Thus, merging them can be a difficult task and the final model might have to be refitted to data in order to produce comparable results. A software like semanticSBML can help the human modeller to detect syntactic and semantic conflicts, to perform uncritical routine steps, and to handle the different naming conventions.

Currently we are extending the annotation abilities of our tool by supporting more data repositories. For a reuse within other applications, we are planning to split semanticSBML into several seperate modules. A long-term aim in model merging is to translate more and more of the modellers' knowledge, intuition, and intentions into formal validity criteria. Programs could check if a model is syntactically or mathematically sound, if it agrees with physical laws and biochemical knowledge, and if it is detailed enough to serve its specific purpose (Liebermeister (2008)).

4 CONCLUSION

To facilitate their reuse, SBML models need to contain semantic annotations. SemanticSBML allows users to set and edit annotations easily. Model merging starts with an initial version of the merged model, which the user can later correct and refine. This makes model merging faster, easier, and safer than merging by hand. However, a computer software can only support, but not replace the modeller in building biochemically meaningful models because it cannot handle the implicit assumptions and intentions on which any model is based.

ACKNOWLEDGEMENTS

We thank Anselm Helbig for his work on semanticSBML and the entire SBML community for their strong commitment and support.

Funding: This work was supported by European Commission [grant LSHG-CT-2006-037469] and by SysMo Translucent (funded by the German Federal Ministry of Science and Education).

REFERENCES

Hucka, M., Sauro, A. F. H., Bolouri, H., Doyle, J., Kitano, H., and the rest of the SBML Forum (2003). The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**(4), 524–531.

Laibe, C. and Le Novère, N. (2007). MIRIAM Resources: tools to generate and resolve robust cross-references in Systems Biology. *BMC Systems Biology*, **1**(1), 58.

Le Novère, N. (2006). Model storage, exchange and integration. *BMC Neuroscience*, **7**(Suppl 1), S11.

Le Novère, N., Finney, A., Hucka, M., Bhalla, U., Campagne, F., Collado-Vides, J., Crampin, E., Halstead, M., Klipp, E., Mendes, P., Nielsen, P., Sauro, H., Shapiro, B., Snoep, J., Spence, H., and Wanner, B. (2005). Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotech.*, **23**(12), 1509–1515.

Le Novère, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., Sauro, H., Schilstra, M., Shapiro, B., Snoep, J., and Hucka, M. (2006). Biomodels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res*, **34**, Database Issue: D689–91.

Liebermeister, W. (2008). Validity and combination of biochemical models. *Proceedings of 3rd International ESCEC Workshop on Experimental Standard Conditions on Enzyme Characterizations*.

Olivier, B. and Snoep, J. (2004). Web-based kinetic modelling using JWS online. *Bioinformatics*, **20**(13), 2143–2144.

Snoep, J., Bruggeman, F., Olivier, B., and Westerhoff, H. (2006). Towards building the silicon cell: A modular approach. *Biosystems*, **83**, 207–216.