Merging of systems biology models with semanticSBML

Wolfram Liebermeister, Falko Krause, Edda Klipp
Max Planck Institute for Molecular Genetics
Ihnestr. 63-73, 14195 Berlin, Germany

Abstract

Model repositories like BioModels.net or JWS online provide a large number of biochemical pathway models in the data format SBML (Systems Biology Markup Language). We have developed semanticS-BML, an interactive model merging tool that helps users build, annotate, check, and combine SBML models. It supports the current standard for biological annotations within SBML and uses a large collection of synonyms and database identifiers for the matching and comparison of model elements. During model merging, semanticS-BML automatically detects the various kinds of syntactic and semantic conflicts and guides the user in resolving them. Here we demonstrate its use with a simple example: following a work by Snoep et al., we merge three models describing the glycolytic pathway, the production of glycerol, and the degradation of the toxic byproduct methylgly-oxal. The example shows typical issues that arise in automatic model merging and emphasises the importance of extensive and detailed biological annotations. SemanticSBML can be freely downloaded at `http://semanticsbml.sourceforge.net`.

1 Introduction

Many biochemical pathway models are published in a computer-readable form. In particular, a large number of models are available in model repositories like BioModels [1] or JWS online [2] in the widely used data format SBML [3] (Systems Biology Markup Language). In bottom-up modelling, several

models describing subsystems or individual biochemical reactions are merged into larger, more comprehensive models [4, 5, 6, 7]. To demonstrate this modelling strategy, Snoep et al. [6] have manually merged three metabolic pathway models describing glycolysis [8], the glycerol side branch [9], and the glyoxylase pathway [10] in the yeast *S. cerevisiae*. The glycolysis model itself had been constructed before, also in a bottom-up approach, from individual enzymatic reactions kinetics obtained from *in vitro* enzyme assays.

Model checking and merging by hand is tedious. Yet the process can be greatly simplified if uncritical routine work and validity checks are performed automatically by computer programs. We developed semanticSBML, an interactive tool for the annotation, checking, and merging of SBML models. The main steps in model merging are (i) to detect duplicate elements and (ii) to detect and resolve conflicts between them. In both steps, the tool proposes choices based on semantic information contained in the models, but the user is always free to revise choices made by the program. Suggestions by the program are based on the biological meaning of model elements as specified by biological annotations in the SBML files.

In this article, we introduce the main features of semanticSBML and demonstrate its use with the example presented by Snoep et al. [6]. We show how the models need to be prepared, what conflicts arise in the merging process, how they can be resolved, and which difficulties still remain.

2   Overview of semanticSBML

SemanticSBML is a program suite for the simultaneous, semi-automatic merging of an arbitrary number of SBML models. In addition, it allows users to create, display, check, and annotate SBML models. For merging, the program first proposes an automatically merged model as a starting point; the user can refine the model based on his own demands by manual manipulation or by applying certain priority rules. Biological annotations are a key element for automated entity recognition and for the detection of conflicts during model merging. Thus, semanticSBML provides functions to find, insert, remove and modify annotations in the format used in the BioModels model repository (which we shall call here "MIRIAM/BioModels annotations"). SemanticSBML is open source software released under the GNU public license and is hosted on sourceforge (project name `semanticsbml`).

SemanticSBML can be installed locally or accessed *via* a web interface.

The graphical user interface allows the user to build, display, check, annotate, and merge several SBML models. Models can be created easily from lists of chemical reactions by the "build" function. The network structure of models can be displayed graphically using network layout algorithms from the graphviz library ("display" function). The "check" function tests a number of validity criteria, most of them based on the MIRIAM/BioModels annotations in the model. The annotation and merging functions are described in Figures 1 and 2.

3   Merging of metabolic pathway models for yeast

As a test case for model merging, Snoep et al. [6] have merged a model of glycolysis in yeast cells [8] with models of the glycerol branch [9] and the glyoxylase branch [10]. Here, we shall repeat the same exercise to demonstrate how semanticSBML is used to annotate, check, and merge existing SBML models. The original models were downloaded from the public SBML repositories BioModels [1] and JWS online [2].

4   Models of glycolysis, glycerol pathway, and glyoxylase pathway

We downloaded the glycolysis and the glycerol model (annotated with MIRIAM / BioModels annotations) from the BioModels repository [1]. As the glyoxylase model was not available at BioModels, we downloaded a version without annotations from JWS online [2] and used semanticSBML to add MIRIAM/RDF annotations (see Table 1). An inspection of the three SBML models showed a couple of issues that needed to be resolved during merging:

1. **Glycolysis model** The glycolysis model by Teusink et al. [8] represents the glycolytic pathway from glucose down to pyruvate; it includes the import of glucose from the extracellular space as well as the production of glycerol, ethanol, and succinate. The concentrations of glyceraldehyde and DAPH (dihydroxyacetone phosphate) are represented by a lumped element called `Triose phosphates`; in the annotation of the
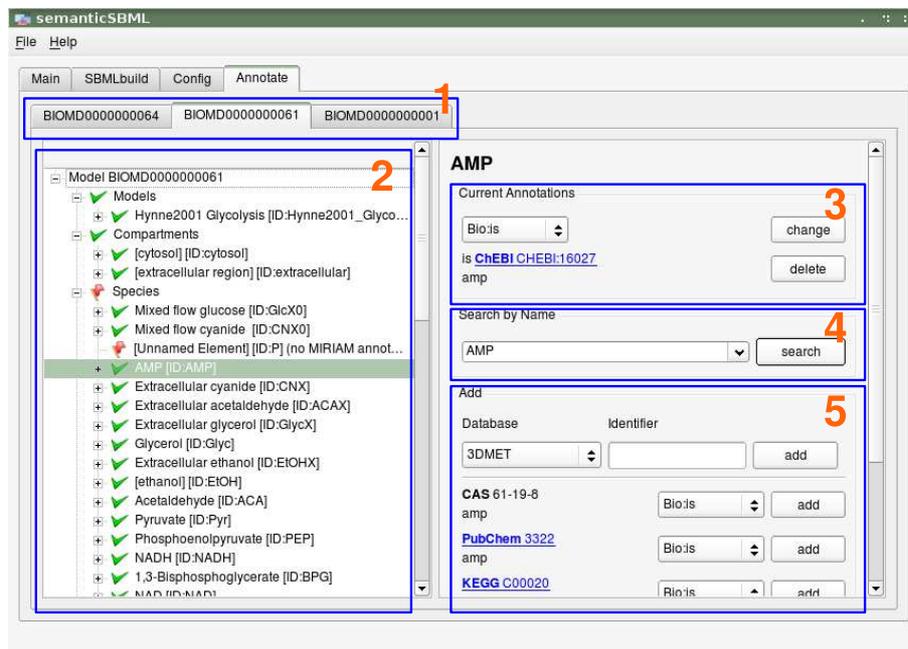
Figure 1: Annotation window of semanticSBML. The framed area 1 contains different tabs, each representing a model. In area 2, all annotatable SBML elements from one model are shown in a tree view. Green icons indicate that an element has MIRIAM annotations, red icons indicate missing MIRIAM annotations. When an element is selected, an annotation menu for this element is shown on the right. Area 3 shows the current annotation and allows the user to modify the annotation qualifiers or to delete annotations. Below, in area 4, a search string can be entered to find new annotations. In area 5, new annotations can be added directly or chosen from the list of suggestions. One such list is shown on initialisation; it is replaced with the search results after a manual search (area 4). When a root node in the element tree has been selected (e.g. `species` in area 2), the annotation menu allows to add annotations automatically to all of its child elements.

element, the individual substances are listed with `hasVersion` qualifiers. In addition, the model contains two elements representing adenosine phosphates and high energy phosphates; effectively, these elements represent the concentrations of AMP, ADP, and ATP in a rather complicated manner; they are semantically overlapping [7], which is also
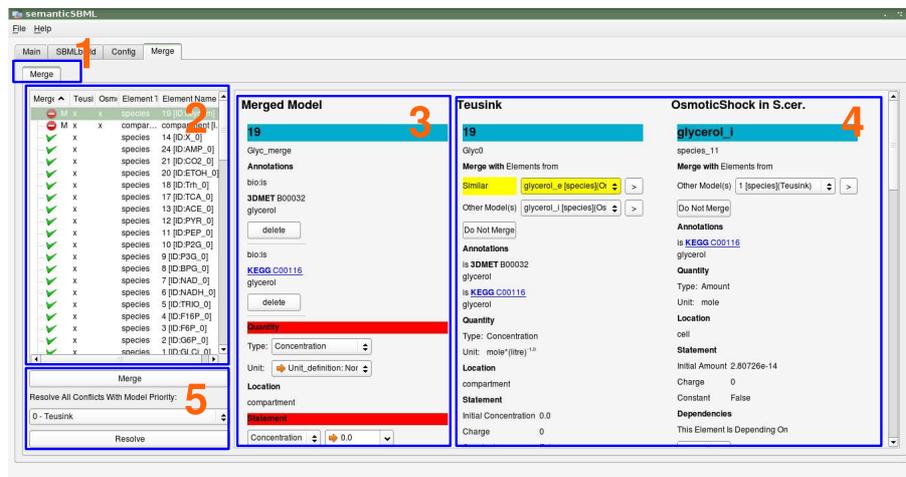
Figure 2: Model merging window of semanticSBML. The framed area 1 contains tabs for the current merging processes. For one of them, area 2 shows an overview of the merged model. Each row corresponds to an element of the output model (according to the current matching between model elements). The first column of the overview indicates whether an element represents one more matching elements. The remaining columns show the origin, the element type and the identifier of the matched elements from the original models. After selecting a row in the overview, the merge menu for the selected element is shown in areas 3 and 4. Area 3 shows the current state of the merged element, highlights conflicts, and provides options for resolving them. Area 4 displays the original model elements and allows the user to modify the matching between them and navigate to other related elements. Conflicts between elements (e.g., different properties attributed to them) can also be resolved automatically using a push-button in area 5. In this case, conflicts are resolved according to a priority list for the models, which can be set by the user. Another push-button allows for the creation of the merged model.

recognised in the automatic model check. In addition, semanticSBML issued a warning because atom numbers are not conserved in some of the reactions. However, if we do not require correct atom balances, this does not pose any problems for model merging.

2. **Glycerol model** In the glycolysis model, glycerol production is de-

5

scribed as a single effective reaction leading from the lumped substance `Triose phosphates` to glycerol. In reality, glycerol is produced from DHAP *via* the intermediate glycerol-3-phosphate. The individual reaction steps have been described in the glycerol pathway model by Cronwright et al. [9]. Four issues arise in merging these two models:

(a) The elements representing glycerol have to be matched.

(b) The glycerol-producing reaction from the glycolysis model (annotated by `isVersionOf` the EC number 1.1.1.8) has to be replaced by the two reactions in the glycerol model. In order to allow semanticSBML to detect the redundancy between the reactions, we annotated the glycerol production reaction in the glycolysis model with the EC number 1.1.1.94 and the `hasPart` qualifier.

(c) The compartment in the glycerol model was annotated with the GO term for cytoplasm, in contrast to the glycolysis model, which is situated in the cytosol. By definition, the cytosol does not comprise the organelles, while the cytoplasm does: in a model, this can make a difference for metabolites that are inhomogeneously distributed in the cell. The reason is possibly an annotation mistake because in the original publication, concentrations refer to the cytosolic volume. We could have changed this annotation in advance, but we decided to keep it and and to match the compartments manually during merging.

(d) In the glycerol model, the concentrations of the cofactors ATP, ADP, NAD, and NADH appear in the reaction kinetics as parameters, and the substances are not declared as modifiers - which poses a problem for merging.

3. **Glyoxylase pathway.** The glyoxylase model of Martins et al. [10] describes the degradation of the toxic compound methylglyoxal, which is formed non-enzymatically from both DHAP and GAP. The model represents a side branch that has not been accounted for in the glycolysis model. To merge it with the glycolysis model, Snoep et al. added two reactions (with mass action kinetics) describing the formation of methylglyoxal from DHAP and GAP, respectively. In the model, they are described by two parallel reactions degrading the lumped substance `triose phosphate`. We extracted these two reactions and all related

Entity	Type	Name	Qualifier	Database	ID
Glyoxylase model	`model`	model	`is`	pubmed	11453985
Cytosol	`compartment`	compartment	`is`	GO	0005829
Methylglyoxal	`species`	MG	`is`	KEGG	C00546
(R)-Lactate	`species`	Lac	`is`	KEGG	C01432
(R)-S-Lactoylglutathione	`species`	SLG	`is`	KEGG	C03451
Hemithioacetal	`species`	HTA	`is`		
Glutathione	`species`	GSH	`is`	KEGG	C00051
Glyoxalase I	`reaction`	GlxI	`is`	EC	4.4.1.5
Glyoxalase II	`reaction`	GlxII	`is`	EC	3.1.2.6
	`reaction`	neHTA			

Table 1: Annotations inserted into the glyoxylase model [10] using the annotation function of semanticSBML. Following Snoep et al., we introduced additional non-enzymatic reactions DHAP → MG and GAP → MG for merging the model with the glycolysis model.

SBML elements from the joined model and put them into a small SBML model to be merged with the three pathway models.

5 Merging the models

We used semanticSBML to merge the three models and the two linking reactions:

1. We manually matched the compartment "cytoplasm" (from the glycerol model) with the the compartment "cytosol" (from the other two models) and removed the annotation "cytoplasm" in the resulting merged compartment element.

2. SemanticSBML matched the elements `Triose phosphates` (glycolysis model) and `DHAP` (glycerol model) and detected a conflict between them. As the two elements are semantically dependent [7], the only safe way to merge the models would be to modify the glycolysis model, replacing the `Triose phosphates` element with separate elements for glyceraldehyde-3-phosphate and DHAP. For simplicity, we ignored this problem and assumed that both triose phosphates act as substrates for the glycerol branch.

3. The conflict between the two versions of the glycerol branch (one or two reactions, respectively) was detected owing to our additional annotation.

4. Glycerol has a concentration of 15.1 mM in the glycerol model and 0.15 mM in the glycolysis model; we chose the value of 0.15 mM.

Another kind of conflict was missed by the program: ATP, ADP, NADH and NAD are described by `species` elements in the glycolysis model, while in the glycerol model, they appear as local parameters. As these parameters were not annotated (and in fact, the current version of semanticSBML does not support annotations for local parameters), this conflict could not be detected. Currently, the best way to fix this problem would be to modify the glycerol model by replacing the four parameters with either `species` elements or with global parameters.

6   Discussion

To give modellers full control over the merging process, we developed semanticSBML as an interactive tool: on the one hand, it serves as an editor for SBML files with MIRIAM/RDF annotations; on the other hand, it guides the user in model merging by suggesting an initial matching between elements, by highlighting potential conflicts, and by providing solutions for them.

The example shows that reliable annotations are a key to automatic merging: the conflicts between cytosol/cytoplasm, triose phosphates/DHAP and the match for glycerol could only be detected because of the annotations, while the match for ATP, ADP, NAD, and NADH could not be detected without the appropriate annotations. In the case of triose phosphates, it was helpful (and also necessary) that the curator listed explicitly the subspecies glyceraldehyde-3-phosphate and DHAP in the annotations because semanticSBML does not yet have a database for inferring relationships between metabolite entities. For comparing the two compartments cytosol and cytoplasm, biological information was necessary to find the overlap between the compartments. Thus, annotations in SBML models should be as explicit as possible; in particular, lumped metabolites and reactions must be annotated carefully.

Problems in model merging can be roughly assigned to the categories "syntax", "semantics", "statements", and "context" [7]. In our test example

for semanticSBML, they were addressed to different extents: (i) syntactical problems (like some missing `listOfModifiers` tags in the glyoxylase model) are already detected by the underlying libSBML library. (ii) Semantic conflicts appeared in several cases: between the compartments, between the triose phosphates and DHAP, and between the different descriptions of the glycerol branch. (iii) Conflicting statements were made about the initial concentration of glycerol. (iv) As "context", we regard the purposes, assumptions, and the biological scope of a model. In this case, the aim was simply to show that semanticSBML supports model merging as shown in Snoep et al. [6].

Acknowledgements

References

[1] N. LeNovère, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, J.L. Snoep, and M. Hucka. BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res*, 34(Database issue):D689–D691, Jan 2006.

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

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

[4] U.S. Bhalla and R. Iyengar. Emergent properties of networks of biological signaling pathways. *Science*, 283(5400):381–7, 1999.

[5] M. Schulz, J. Uhlendorf, E. Klipp, and W. Liebermeister. SBMLmerge, a system for combining biochemical network models. *Genome Informatics Series*, 17(1), 2006.

[6] J.L. Snoep, F. Bruggeman, B.G. Olivier, and H.V. Westerhoff. Towards building the silicon cell: A modular approach. *Biosystems*, 83:207–216, 2006.

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

[8] B. Teusink, J. Passarge, C. A. Reijenga, E. Esgalhado, C. C. van der Weijden, M. Schepper, M. C. Walsh, B. M. Bakker, K. van Dam, H. V. Westerhoff, and J. L. Snoep. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? testing biochemistry. *Eur J Biochem*, 267(17):5313–5329, Sep 2000.

[9] G.R. Cronwright, J.M. Rohwer, and B.A. Prior. Metabolic control analysis of glycerol synthesis in Saccharomyces cerevisiae. *Appl Environ Microbiol*, 68(9):4448–4456, 2002.

[10] A.M. Martins, P. Mendes, C. Cordeiro, and A.P. Freire. In situ kinetic analysis of glyoxalase I and glyoxalase II in Saccharomyces cerevisiae. *European Journal of Biochemistry*, 268(14):3930–3936, 2001.