

SBML2L^AT_EX: Conversion of SBML files into human-readable reports

Andreas Dräger^{1,*}, Hannes Planatscher¹, Dieudonne Motsou Wouamba¹,
Adrian Schröder¹, Michael Hucka², Lukas Endler³, Martin Golebiewski⁴,
Andreas Zell¹

¹Center for Bioinformatics Tübingen (ZBIT), University of Tübingen, Sand 1, 72076 Tübingen, Germany

²California Institute of Technology, Beckman Institute BNMC, Pasadena, United States

³European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom

⁴Scientific Databases and Visualization Group, EML Research gGmbH, Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXXX

ABSTRACT

Summary: The XML-based language SBML has emerged as the standard for storage, communication and interchange of models in systems biology. As the machine-readable standard XML can hardly be understood by humans, many tools are available to visualize the reaction pathways stored in SBML files, but many components, e.g., unit declarations or complex kinetic equations, are often not visible in these diagrams. For a broader understanding of the models, support in scientific writing and error detection, a human-readable report of the complete model is needed. We present SBML2L^AT_EX, a Java-based tool to fill this gap. SBML2L^AT_EX can be downloaded as a stand-alone tool from the project homepage. A convenient web-service allows the user to convert SBML to various formats, including PDF, PS, DVI, PNG and L^AT_EX, and provides many settings for customization.

Availability: Source code, documentation, and a web-service are freely available at <http://webservices.cs.uni-tuebingen.de/webservices>.

Contact: andreas.draeger@uni-tuebingen.de

1 INTRODUCTION

The XML-based Systems Biology Markup Language (SBML) (Hucka *et al.*, 2003) has become the *de facto* standard format for storing models of biochemical systems. SBML allows for defining complex models of molecular interactions and cellular processes. It is not necessary to type XML manually; over one hundred software tools now support SBML, including many with intuitive graphical interfaces. Much effort was put into the visualization of the reaction pathways contained in SBML models (Gauges *et al.*, 2006). Many tools also support visualizing and saving molecular interaction graphs (Funahashi *et al.*, 2003), but important details such as unit definitions, kinetic rate equations, user-defined functions, events, model notes, or annotations given in the Systems Biology Ontology (SBO) are usually not made explicit in the graphical presentations. To detect potential errors or to gain an overview of the model as

a whole, it is necessary to examine the full content of the SBML file, but the unfriendliness of XML to human readers makes this an inconvenient and difficult task.

We therefore present SBML2L^AT_EX, a converter for SBML files, which produces L^AT_EX source code. For convenience of usage, an on-line web-service has been implemented, which directly produces human-readable files in various formats (currently L^AT_EX, DVI, PS, PDF, or a ZIP file containing JPG, PNG, GIF or EPS files). Several settings allow for customization of the output, such as extra title page or headline, the paper size, orientation (portrait or landscape), font sizes and font styles. SBML2L^AT_EX covers all constructs defined in the latest level and version of SBML and is able to typeset complex kinetic formulas. It computes the derived units for all SBML elements using libSBML (Bornstein *et al.*, 2008) and shows warnings if kinetic equations cannot be evaluated to the correct units. All information is presented in clearly arranged tables, reaction equations and plain text, simplifying the task of understanding and communicating the model as well as detecting and correcting errors. This work extends an earlier approach (Dräger *et al.*, 2008) to translate SBML to L^AT_EX that did not focus on giving a complete model report and mainly translated the kinetic equations.

2 TRANSLATION OF SBML

The following paragraphs describe how SBML2L^AT_EX structures its model reports. Besides the mandatory field “id” (short identifier) every SBML component contains the optional attributes name (a detailed identifier), SBO term (number of the corresponding term), notes (XHTML formatted explanation to be displayed to humans) and annotation (machine readable extension for software tools). Most SBML components contain special additional fields for the certain type, e.g., the unit of a parameter, species or compartment. SBML2L^AT_EX translates every optional field if it exists and writes this information to the description of the respective SBML component. MIRIAM annotations are translated as well. Software-specific additional annotations, however, e.g., graph layout extensions, remain untranslated.

The headline of the model report favors to contain the name of the model or its id if no name is declared. The first section presents a general overview of

*to whom correspondence should be addressed

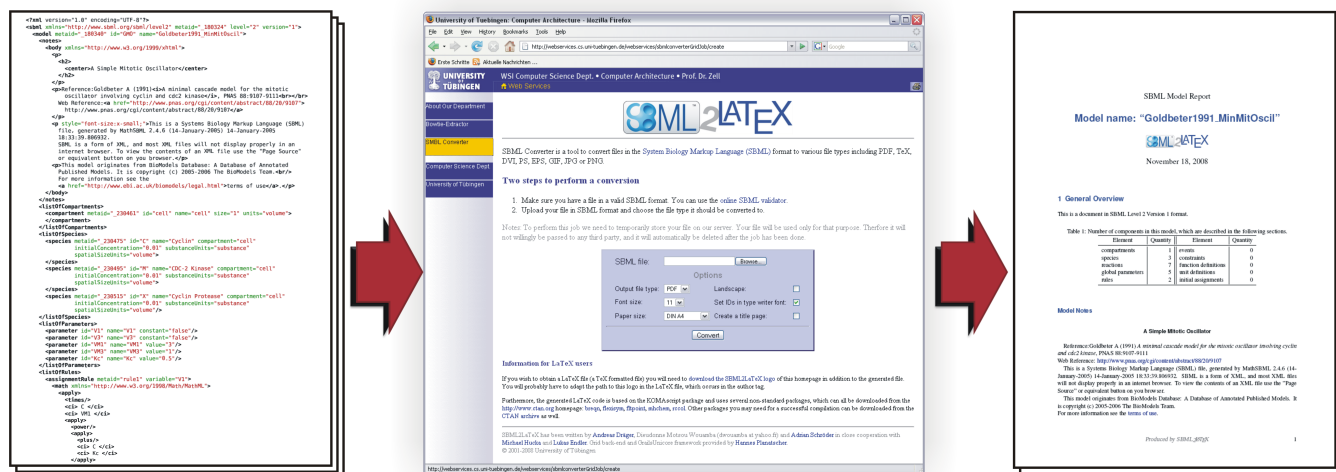


Fig. 1. Workflow of SBML2LATEX using the web interface for the SBML file BIOMD0000000003.xml (available at <http://www.biomodels.net>). After upload of an SBML file several options allow for customization of the output: The desired file type can be selected, the paper orientation can be switched to landscape, which is especially important if the report contains fractions with very long denominators, where no automatic line break can be inserted. The font size can be varied, text can be written in typewriter or roman font. The paper size can be set to the US formats letter, legal or executive as well as to the European formats DIN A0-9. An extra title page can be generated. By clicking on the “convert” button the report file is generated and accessible for download.

the model, including the number of SBML components within the model as well as SBML level and version. If not overridden by the user, SBML implicitly declares unit definitions for substance, volume, area, length and time, which are all made explicit in the report because these units, even if undeclared, are valid in the model. Making all unit declarations explicit simplifies the error detection process. As SBML does not contain mandatory components, SBML2LATEX displays the sections about the following components in this order only if declared: compartment types, compartments, species types, species, global parameters, initial assignments, function definitions, rules, events, constraints, and reactions. Each one of these sections starts with a sentence giving the number of components to be described and displays all available information about the respective component. For instance, section “Reactions” contains a table with all reaction equations and one subsection for each single reaction. For each reaction its reactants, products and modifiers are displayed in a table, followed by the formula of the kinetic law, its derived unit and a table of local parameters. For events the trigger condition, the delay function, if one exists, and all assignments are given. If the model contains any species, the last section shows the derived rate equations for the temporal changes of their amount. SBML2LATEX highlights kinetic equations, which do not return the unit substance per time. Hyperlinks allow the user to jump to each referenced kinetic equation, event or rule a species is involved in. If the model contains any SBO annotations, a glossary gives the SBO numbers together with terms and definitions at the end of the document.

SBML2LATEX is distributed under the GNU General Public License and completely written in Java™ (platform independent). It contains a modified version of HTML2LATEX (<http://html2latex.sourceforge.net>) for the translation of XHTML elements and depends on an installation of libSBML.

3 CONCLUSION

SBML2LATEX facilitates the complicated and cumbersome model building process by providing a simple method to translate such models to human-readable reports. These reports ease the complex process of error detection and model communication. Furthermore, it supports scientific writing as sophisticated formulas can be directly

adopted. The web-service version provides a convenient way to create such reports in various formats and offering several options. If further customization becomes necessary, the source code and the binaries can be downloaded and used locally. SBML2LATEX has been integrated into the SABIO-RK database (Rojas *et al.*, 2007) and can directly be accessed from the SBML homepage <http://sbml.org>. Furthermore, it is planned to equip each model in the Biomodels.net database with one model report generated by SBML2LATEX.

ACKNOWLEDGMENTS

The authors are grateful to Henning Schmidt, Detlev Bannasch and Jochen Supper. This work was funded by the German Federal Ministry of Education and Research (BMBF) in the projects National Genome Research Network (NGFN) under project number 0313323 and HepatoSys under Project Number 0313080 L.

REFERENCES

- Bornstein, B. J., Keating, S. M., Jouraku, A., and Hucka, M. (2008). LibSBML: an API Library for SBML. *Bioinformatics*, **24**(6), 880–881.
- Dräger, A., Hassis, N., Supper, J., Schröder, A., and Zell, A. (2008). SBMLsqueezer: A CellDesigner plug-in to generate kinetic rate equations for biochemical networks. *BMC Systems Biology*, **2**(1), 39.
- Funahashi, A., Tanimura, N., Morohashi, M., and Kitano, H. (2003). CellDesigner, a process diagram editor for gene-regulatory and biochemical networks. *BioSilico*, **1**(5), 159–162.
- Gauges, R., Rost, U., Sahle, S., and Wegner, K. (2006). A model diagram layout extension for SBML. *Bioinformatics*, **22**(15), 1879–1885.
- Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Arkin, A. P., Bornstein, B. J., Bray, D., Cornish-Bowden *et al.* (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**(4), 524–531.
- Rojas, I., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., and Wittig, U. (2007). Storing and Annotating of Kinetic Data. *In Silico Biology*, **7**, 37–44.