SBML2LaTeX: Conversion of SBML files into human-readable reports

Journal:	*Bioinformatics*
Manuscript ID:	draft
Category:	Applications Note
Date Submitted by the Author:	n/a
Complete List of Authors:	Dräger, Andreas; University of Tübingen, Center for Bioinformatics Tübingen (ZBIT) Planatscher, Hannes; University of Tübingen, Center for Bioinformatics Tübingen (ZBIT) Wouamba, Dieudonné; University of Tübingen, Center for Bioinformatics Tübingen (ZBIT) Schröder, Adrian; University of Tübingen, Center for Bioinformatics Tübingen (ZBIT) Hucka, Michael; California Institute of Technology, Beckman Institute 139-74 Endler, Lukas; European Bioinformatics Institute, Wellcome Trust Genome Campus Golebiewski, Martin; EML Research gGmbH, Scientific Databases and Visualization Group Müller, Wolfgang; EML Research gGmbH, Scientific Databases and Visualization Group Zell, Andreas; University of Tübingen, Center for Bioinformatics Tübingen (ZBIT)
Keywords:	Knowledge representation, Ontology, Pathways, Systems biology, Visualization, Web services

BIOINFORMATICS APPLICATIONS NOTE

Vol. 00 no. 00 2008
Pages 1–2

SBML2LaTeX: Conversion of SBML files into human-readable reports

Andreas Dräger [1,*], Hannes Planatscher [1], Dieudonné Motsou Wouamba [1], Adrian Schröder [1], Michael Hucka [2], Lukas Endler [3], Martin Golebiewski [4], Wolfgang Müller [4], Andreas Zell [1]

[1]Center for Bioinformatics Tübingen (ZBIT), University of Tübingen, Sand 1, 72076 Tübingen, Germany
[2]Beckman Institute BNMC, California Institute of Technology, Pasadena, United States
[3]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom
[4]Scientific Databases and Visualization Group, EML Research gGmbH, Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Summary: The XML-based Systems Biology Markup Language (SBML) has emerged as a standard for storage, communication and interchange of models in systems biology. As a machine-readable format XML is difficult for humans to read and understand. Many tools are available for visualizing the reaction pathways stored in SBML files, but many components, e. g., unit declarations, complex kinetic equations or links to MIRIAM resources, are often not made 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 SBML2LaTeX, a Java-based stand-alone program to fill this gap. A convenient web service allows users to directly convert SBML to various formats, including DVI, LaTeX, and PDF, 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

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. Over one hundred software tools now support SBML, including many with intuitive graphical interfaces. Many tools 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 whether in Systems Biology Ontology (SBO) (Le Novère *et al.*, 2006a) or MIRIAM (Le Novère *et al.*, 2005) 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.

To address this problem, we have developed SBML2LaTeX, a tool that accepts SBML files as input and generates summaries of their contents as reports in LaTeX source code format. For convenience of usage an online web service directly produces various human-readable files in various formats. Several settings allow for customizing the output, such as extra title page or headline, the paper size, orientation (portrait or landscape), font sizes and font styles. SBML2LaTeX covers all constructs defined in the latest level and version of SBML (Level 2 Version 4) 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 LaTeX that mainly translated kinetic equations and did not focus on giving a complete report of the model.

2 TRANSLATION OF SBML

Besides the mandatory field "id" (short identifier), every SBML component may contain optional attributes for a detailed name, SBO term number, notes (XHTML formatted explanation to be displayed to humans) and annotation (machine readable extension for software tools). Most SBML components contain special additional fields specific to each component type, e. g., the unit of a parameter, species or compartment. SBML2LaTeX translates every optional field if it exists and writes this information in the description of the respective component. The URNs in MIRIAM annotations are translated to hyperlinks to the actual URLs. However, software-specific annotations (such as graph layout extensions) are not translated. The headline of the model report contains the model's name, or its "id" attribute value if the SBML file does not assign a name to the model. The first section presents a general overview of the model, including the number of SBML components within the model, SBML level and version, and the model's history. All five predefined SBML unit definitions are made explicit, which simplifies the error detection process. As SBML does not contain mandatory components, SBML2LaTeX

*to whom correspondence should be addressed

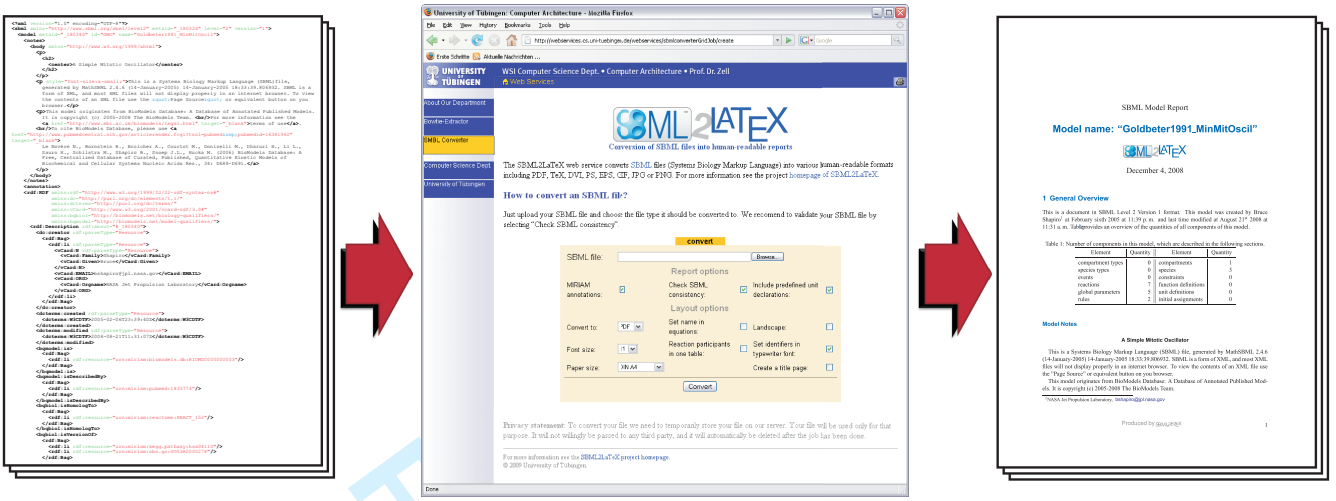

Fig. 1. Example workflow using the SBML2LaTeX web service and SBML file `BIOMD0000000003.xml` (available at http://www.biomodels.net). After uploading an SBML file, several options allow customizing the output: MIRIAM annotations, an SBML consistency check or predefined unit declarations can be excluded, the desired file format can be selected; the paper size can be set to the US formats letter, legal or executive as well as to the European formats DIN A0-9 and the paper orientation can be switched to landscape (especially important if the report contains fractions with very long denominators, where no automatic line break can be inserted). Several other options influence the layout of the report, e. g., identifiers can be displayed in typewriter or roman font or names can be used in equations instead of identifiers. When the user clicks on the "convert" button, the report file is generated and accessible for download.

displays the sections about the following components in this order only if they are declared in the model: compartment types, compartments, species types, species, global parameters, initial assignments, function definitions, rules, events, constraints, and reactions. Each one of these sections in the report 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 units 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 whose units cannot be verified to equal *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 presents the SBO numbers together with terms and definitions. Finally, a consistency report of the model is included at the end of the document. SBML2LaTeX is distributed under the GNU General Public License and completely written in Java™. It contains a modified version of HTML2LaTeX (http://htmltolatex.sourceforge.net) and depends on an installation of libSBML.

3 CONCLUSION

SBML2LaTeX facilitates the complicated and combersome model development process by providing a simple method to translate such models to human-readable reports. These reports support scientific writing because sophisticated formulas can be directly adopted and ease the error detection and model communication. The web service version provides a convenient way to create such reports in various formats and offers 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. Further, such automatic conversion could be used by databases of models, for instance the BioModels Database (Le Novère *et al.*, 2006b).

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 National Genome Research Network (NGFN) under Project Number 0313323 and HepatoSys under Project Number 0313080 L and federal state Baden-Württemberg in the Tübinger Bioinformatik-Grid under Project Number 23-7532.24-4-18/1. *Conflict of Interest:* none declared.

REFERENCES

Bornstein, B. J. *et al.* (2008). LibSBML: an API Library for SBML. *Bioinformatics*, **24**(6), 880–881.

Dräger, A. *et al.* (2008). SBMLsqueezer: A CellDesigner plug-in to generate kinetic rate equations for biochemical networks. *BMC Systems Biology*, **2**(1), 39.

Funahashi, A. *et al.* (2003). CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *BioSilico*, **1**(5), 159–162.

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

Le Novère, N. *et al.* (2005). Minimum information requested in the annotation of biochemical models (MIRIAM). *Nature Biotechnology*, **23**(12), 1509–1515.

Le Novère, N. *et al.* (2006a). Adding semantics in kinetics models of biochemical pathways. In C. Kettner and M. G. Hicks, editors, *2nd International ESCEC Workshop on Experimental Standard Conditions on Enzyme Characterizations. Beilstein Institut, Rüdesheim, Germany*, Rüdessheim/Rhein, Germany. ESEC.

Le Novère, N. *et al.* (2006b). BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research*, **34**, D689–D691.

Rojas, I. *et al.* (2007). Storing and Annotating of Kinetic Data. *In Silico Biology*, **7**, 37–44.