

# SBMLsqueezer 2: a context-sensitive rate law generator for biochemical networks with access to SABIO-RK

Andreas Dräger<sup>1,2,\*</sup>, Roland Keller<sup>2</sup>, Matthias Rall<sup>2</sup>, Johannes Eichner<sup>2</sup>, Bernhard Ø. Palsson<sup>1</sup>, Andreas Zell<sup>2</sup>

<sup>1</sup>Systems Biology Research Group, University of California, San Diego, La Jolla, CA, United States

<sup>2</sup>Center for Bioinformatics Tuebingen (ZBIT), University of Tuebingen, Tübingen, Germany

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXXX

## ABSTRACT

**Summary:** Modeling metabolic networks belongs to the most laborious and error-prone tasks in systems biology. Size and complexity of published reconstructions are steadily increasing. In order to simulate the dynamics of quantitative models, kinetic equations need to be derived for each reaction. Parameters and units also need to be specified. The manual assignment of these equations is not practicable for large numbers of reactions. Complex test-and-evaluation cycles require automated methods for rate law assignment. The program SBMLsqueezer 2 is a generator for kinetic equations, parameters, and units. It distinguishes between multiple types of reactions and selects only suitable rate laws. The user can influence all choices made by the program in order to assign the desired type of rate law to each reaction. Experimentally derived rate laws can be obtained through a connection to the kinetics database SABIO-RK. This platform-independent program can be used in a large variety of ways, enabling flexible solutions and use-case scenarios.

**Availability:** Program, source code, and documentation can be obtained under the terms of the GPL version 3 from the website <http://www.cogsys.cs.uni-tuebingen.de/software/SBMLsqueezer/>.

**Contact:** sbmlsqueezer@googlegroups.com

## 1 INTRODUCTION

The reconstruction of genome-scale networks has been recognized as a highly laborious long-term effort, which requires several iterations of curation (Thiele and Palsson, 2010). However, for the creation of dynamic network simulations, the formulation of such a structural network is just the first step. For each reaction within the network, a specific kinetic equation, a so-called rate law, needs to be derived. These rate laws typically contain parameters, such as Michaelis constants, or even plain numbers, which are not always intended to be dimensionless quantities. Reactive species in these models may cross compartments through transport reactions. Hence, their reaction space may change and consequently their molarity with the volume differences of each compartment. Ensuring unit consistency for the entire network requires careful consideration. When a model is finally used as the basis for technical applications, inconsistent units could compromise therapeutic procedures or even endanger health and safety of patients.

Manually deriving both, kinetic equations and all units, brings several problems with it, because it a) is highly error-prone, b) very time-consuming, and c) cannot be applied in large-scale or automated approaches. For these reasons, automatic procedures are required for the assembly of rate laws. Programs, such as COPASI (Hoops *et al.*, 2006), CellDesigner (Funahashi *et al.*, 2007), the MASS-Toolbox (<http://opencobra.github.io/MASS-Toolbox/>), and Cellerator (Shapiro *et al.*, 2002), provide pre-defined lists of kinetic equations and also allow the user to modify these rate laws or to even create customized equations. CellDesigner 4.3 provides a dialog that assists the user to obtain rate laws from the kinetics database SABIO-RK (Wittig *et al.*, 2012). The MASS-Toolbox focuses on the creation of elementary rate laws and automatically derives pseudo-elementary rate constants with their units. Inference programs, such as NetGenerator (Weber *et al.*, 2013), estimate a topology and generate specific rate laws for gene-regulatory processes. Odefy (Krumsiek *et al.*, 2010) converts discrete Boolean networks into quantitative differential equation systems by applying Hill-type rate laws to each transition.

In contrast, SBMLsqueezer 2 applies several criteria to automatically select appropriate equations for each reaction. The user can influence these criteria and choose which rate law to apply. The aims of this approach are a) to ensure that only applicable rate laws can be selected and thus the consistency of the model, and b) to reduce the required human interaction to a minimum. SBMLsqueezer 2 is intended to be useful for modeling not only metabolic networks but also signal transduction processes and gene-regulatory mechanisms.

## 2 RESULTS

Originally, SBMLsqueezer has been developed as a plug-in for CellDesigner (Dräger *et al.*, 2008). It was then extended to a stand-alone tool (Dräger *et al.*, 2010). SBMLsqueezer 2 can be used as a) on-line program, b) stand-alone tool via graphical user interface or command-line, c) plug-in for CellDesigner, d) Garuda gadget, and e) through its API in complex workflows and algorithms.

The vast majority of biochemical reactions can be categorized into a limited number of classes. SBMLsqueezer 2 takes several features of the reaction into account in order to discriminate these classes. For each class it either determines all kinds of principally applicable rate laws, or the most suitable rate law. The user can influence how rate laws are picked. An equation preview assists the

\*to whom correspondence should be addressed

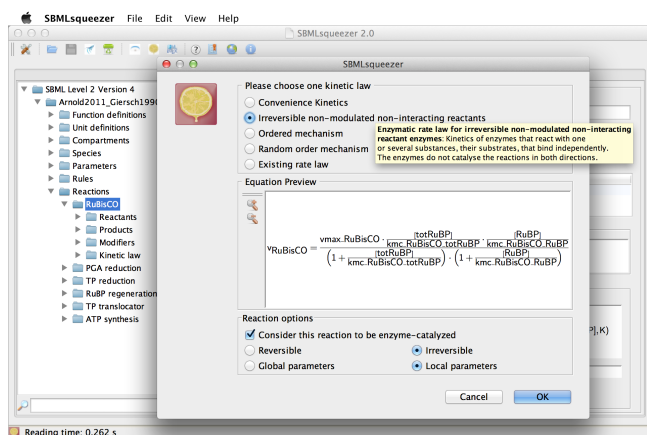


Figure 1: Reaction context dialog of SBMLsqueezer 2.

user to make this decision. The program equips all newly generated parameters with units in order to ensure consistency. It should be noted that for some levels and versions of SBML numbers cannot be associated with units and that some rate laws can under certain conditions not be evaluated to reaction extend per time units.

The dialog for access to SABIO-RK is similar to the on-line database service and uses the annotation of the reaction and its components to identify the best match in SABIO-RK. When successful, rate laws, parameters, units, and annotations will be transferred from SABIO-RK to the local model. Rate law generation and extraction from SABIO-RK can be performed for individual reactions (fig. 1) or for the entire model in a single step. Optionally, SBMLsqueezer 2 can remove unused variables and units from the model (cleaning) and update annotations where required. The content of models can be summarized in a comprehensive human-readable PDF report.

### 3 IMPLEMENTATION

SBMLsqueezer 2 is based on the data format SBML (Hucka et al., 2004). It is entirely implemented in Java™ and runs on every platform, for which a JVM is available. Reading and writing of SBML files is done with JSBML (Dräger et al., 2011), which also acts as the internal data structure. SBMLsqueezer 2 can also be launched using a libSBML (Bornstein et al., 2008) back-end. The on-line program version is based on the command-line interface of the stand-alone tool, which is wrapped in a Galaxy (Goecks et al., 2010) framework. For writing model reports, SBMLsqueezer 2 contains a development release of SBML2L<sup>A</sup>T<sub>E</sub>X (Dräger et al., 2009). The Garuda gadget (Ghosh et al., 2011) is implemented based on the back-end API for Java™. The CellDesigner plug-in uses the communication interface between CellDesigner's plug-in API and JSBML. Changes made by SBMLsqueezer 2 are synchronized with CellDesigner through a change listener interface. SBMLsqueezer 2 determines the type of reaction by interpreting SBO and MIRIAM annotations (Courtot et al., 2011) of all components as well as the number and kind of reaction participants. Access to SABIO-RK (Wittig et al., 2012) requires an active Internet connection and uses the RESTFUL API provided by SABIO-RK through a Java™ URL connection.

### 4 CONCLUSION

With SBMLsqueezer 2 a mature and stable application has been released, which can be applied in diverse ways. It can easily be integrated into versatile workflows and complex procedures. The Users'

Guide at the project web-site explains in detail how to exploit all program functions and provides several sample use-cases.

The path2models project (Büchel et al., 2013) has demonstrated the usefulness of automated rate law assignment. Based on SBMLsqueezer 2 complex try-and-evaluate cycles, in which the most suitable rate law for a certain reaction needs to be identified in repeated simulation runs (Dräger, 2011) now become possible.

### ACKNOWLEDGMENTS

The authors are grateful to Meike Aichele, Hannes Borch, Alexander Dörr, Martin Golebiewski, Nadine Hassis, Marcel Kronfeld, Oliver Kohlbacher, Wolfgang Müller, Sarah R. Müller vom Hagen, Sebastian Nagel, Leif J. Pallesen, Alexander Peltzer, Julianus Pfeuffer, Sandra Saliger, Simon Schäfer, Adrian Schröder, Jochen Supper, Dieudonné M. Wouamba, Shaowu Yang, Michael J. Ziller.

**Funding:** Thanks to the Federal Ministry of Education and Research (BMBF, Germany) for funding the Virtual Liver Network (grant number 0315756) and to the 7<sup>th</sup> EU Framework Program for Research and Technological Development for AD's Marie Curie International Outgoing Fellowship (project AMBiCon, 332020).

**Conflict of Interest:** none declared.

### REFERENCES

- Bornstein et al. (2008). LibSBML: an API library for SBML. *Bioinformatics*, **24**(6), 880–881.
- Büchel et al. (2013). Large-scale generation of computational models from biochemical pathway maps. *BMC Syst Biol*, **7**(1), 116.
- Courtot et al. (2011). Controlled vocabularies and semantics in systems biology. *Mol Syst Biol*, **7**, 543.
- Dräger, A. (2011). *Computational Modeling of Biochemical Networks*. Ph.D. thesis, University of Tuebingen.
- Dräger, et al. (2008). SBMLsqueezer: a CellDesigner plug-in to generate kinetic rate equations for biochemical networks. *BMC Syst Biol*, **2**(1), 39.
- Dräger et al. (2009). SBML2L<sup>A</sup>T<sub>E</sub>X: Conversion of SBML files into human-readable reports. *Bioinformatics*, **25**(11), 1455–1456.
- Dräger et al. (2010). *Systems Biology for Signaling Networks*, volume 2, chap. Automating mathematical modeling of biochemical reaction networks. Springer.
- Dräger et al. (2011). JSBML: a flexible Java library for working with SBML. *Bioinformatics*, **27**(15), 2167–2168.
- Funahashi et al. (2007). Integration of CellDesigner and SABIO-RK. *In Silico Biology*, **7**(2 Suppl), S81–S90.
- Ghosh et al. (2011). Software for systems biology: from tools to integrated platforms. *Nat Rev Genet*, **12**(12), 821–832.
- Goecks et al. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*, **11**(8), R86.
- Hoops et al. (2006). COPASI—a Complex Pathway Simulator. *Bioinformatics*, **22**(24), 3067–3074.
- Hucka et al. (2004). Evolving a *lingua franca* and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project. *Systems Biology, IEE*, **1**(1), 41–53.
- Krumsiek et al. (2010). Odepy—from discrete to continuous models. *BMC Bioinformatics*, **11**, 233.
- Shapiro et al. (2002). Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics*, **19**(5), 677–678.
- Thiele, I. and Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*, **5**(1), 93–121.
- Weber et al. (2013). Inference of dynamical gene-regulatory networks based on time-resolved multi-stimuli multi-experiment data applying NetGenerator v2.0. *BMC Syst Biol*, **7**(1), 1.
- Wittig et al. (2012). SABIO-RK—database for biochemical reaction kinetics. *Nucleic Acids Res*, **40**(Database issue), D790–D796.