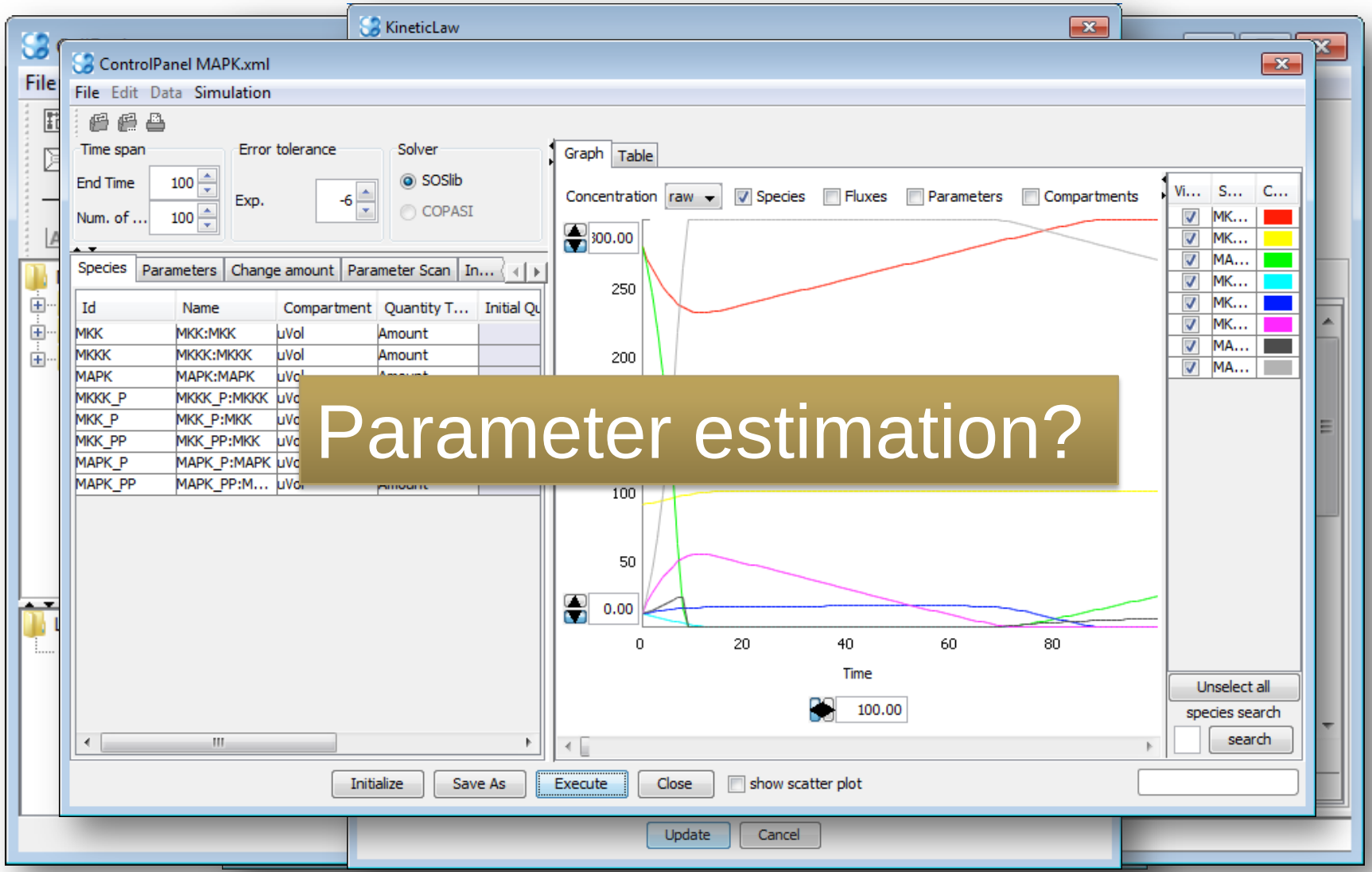
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

MATHEMATISCH-
NATURWISSENSCHAFTLICHE FAKULTÄT

From KEGG to dynamic pathway models: a collection of tools to facilitate the modeling of biochemical networks

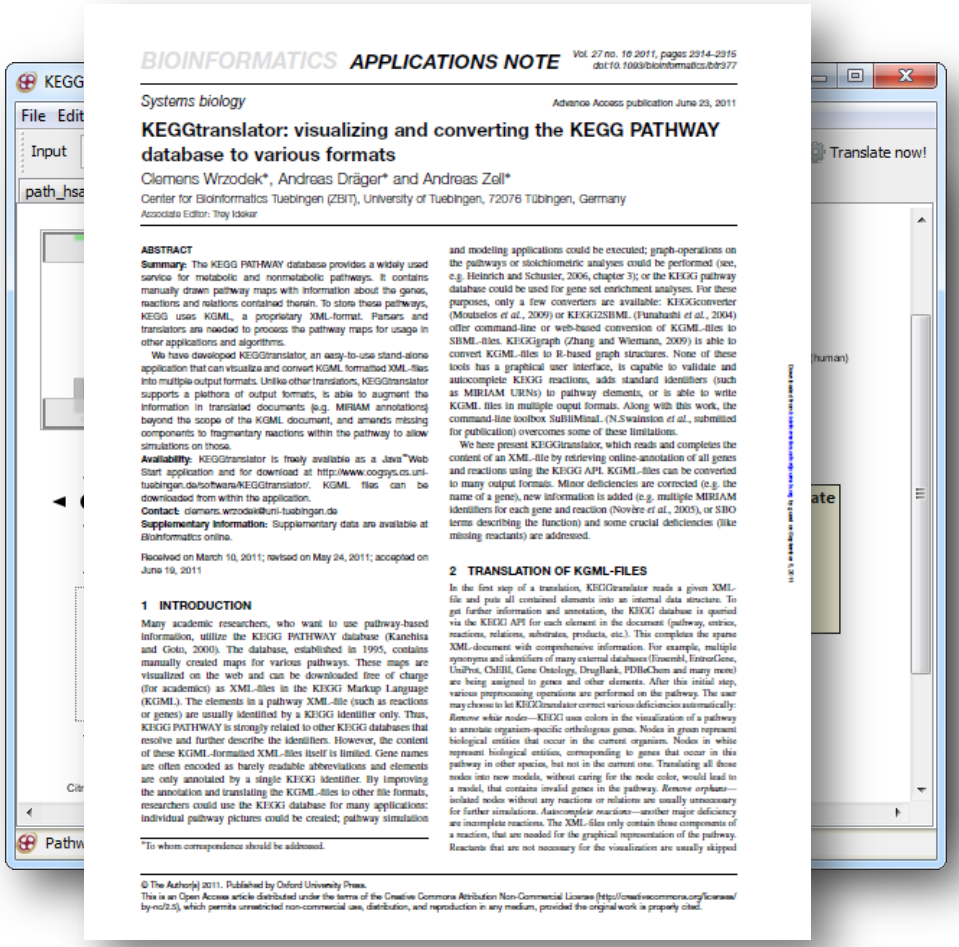
Andreas Dräger, Roland Keller, Clemens Wrzodek,
Alexander Dörr, and Andreas Zell

Common workflow



KEGGtranslator

- Translating KEGG pathways to
  - GraphML
  - SBML
- Improve KEGG annotations
  - Automated modeling
  - Easy linkage of analysis results to KEGG pathways



BIOINFORMATICS APPLICATIONS NOTE Vol. 27 no. 10 2011, pages 2914–2915
doi:10.1093/bioinformatics/btr377

Systems biology

Advance Access publication June 23, 2011

KEGGtranslator: visualizing and converting the KEGG PATHWAY database to various formats

Clemens Wrzodek*, Andreas Dräger* and Andreas Zell*

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

Associate Editor: Thay Ideker

ABSTRACT

Summary: The KEGG PATHWAY database provides a widely used service for metabolic and nonmetabolic pathways. It contains manually drawn pathway maps with information about the genes, reactions and relations contained therein. To store these pathways, KEGG uses KGML, a proprietary XML-format. Parsers and translators are needed to process the pathway maps for usage in other applications and algorithms.

We have developed KEGGtranslator, an easy-to-use stand-alone application that can visualize and convert KGML formatted XML-files into multiple output formats. Unlike other translators, KEGGtranslator supports a plethora of output formats, is able to augment the information in translated documents (e.g. MIRIAM annotations) beyond the scope of the KGML document, and amends missing components to fragmentary reactions within the pathway to allow simulations on those.

Availability: KEGGtranslator is freely available as a Java™ Web Start application and for download at http://www.cogsys.cs.uni-tuebingen.de/software/KEGGtranslator/. KGML files can be downloaded from within the application.

Contact: clemens.wrzodek@uni-tuebingen.de

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Received on March 10, 2011; revised on May 24, 2011; accepted on June 19, 2011

1 INTRODUCTION

Many academic researchers, who want to use pathway-based information, utilize the KEGG PATHWAY database (Kanehisa and Goto, 2000). The database, established in 1995, contains manually created maps for various pathways. These maps are visualized on the web and can be downloaded free of charge (for academics) as XML-files in the KEGG Markup Language (KGML). The elements in a pathway XML-file (such as reactions or genes) are usually identified by a KEGG identifier only. Thus, KEGG PATHWAY is strongly related to other KEGG databases that resolve and further describe the identifiers. However, the content of these KGML-formatted XML-files itself is limited. Gene names are often encoded as barely readable abbreviations and elements are only annotated by a single KEGG identifier. By improving the annotation and translating the KGML-files to other file formats, researchers could use the KEGG database for many applications: individual pathway pictures could be created; pathway simulation and modeling applications could be executed; graph-operations on the pathways or stoichiometric analyses could be performed (see, e.g. Heinrich and Schuster, 2006, chapter 3); or the KEGG pathway database could be used for gene set enrichment analyses. For these purposes, only a few converters are available: KEGGconverter (Moutselos *et al.*, 2009) or KEGG2SBML (Funahashi *et al.*, 2004) offer command-line or web-based conversion of KGML-files to SBML-files. KEGGgraph (Zhang and Wiemann, 2009) is able to convert KGML-files to R-based graph structures. None of these tools has a graphical user interface, is capable to validate and autocomplete KEGG reactions, adds standard identifiers (such as MIRIAM URNs) to pathway elements, or is able to write KGML files in multiple ouput formats. Along with this work, the command-line toolbox SuBliMinal. (N.Swainston *et al.*, submitted for publication) overcomes some of these limitations.

We here present KEGGtranslator, which reads and completes the content of an XML-file by retrieving online-annotation of all genes and reactions using the KEGG API. KGML-files can be converted to many output formats. Minor deficiencies are corrected (e.g. the name of a gene), new information is added (e.g. multiple MIRIAM identifiers for each gene and reaction (Novère *et al.*, 2005), or SBO terms describing the function) and some crucial deficiencies (like missing reactants) are addressed.

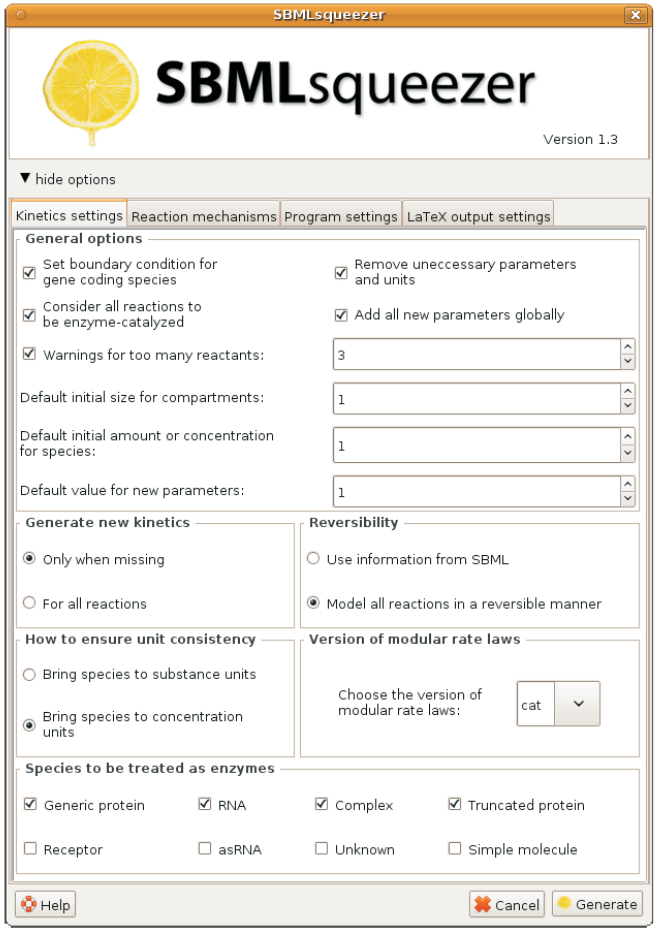
2 TRANSLATION OF KGML-FILES

In the first step of a translation, KEGGtranslator reads a given XML-file and puts all contained elements into an internal data structure. To get further information and annotation, the KEGG database is queried via the KEGG API for each element in the document (pathway, entries, reactions, relations, substrates, products, etc.). This completes the sparse XML-document with comprehensive information. For example, multiple synonyms and identifiers of many external databases (Ensembl, EnsemGene, UniProt, ChEBI, Gene Ontology, DrugBank, PDBeChem and many more) are being assigned to genes and other elements. After this initial step, various preprocessing operations are performed on the pathway. The user may choose to let KEGGtranslator correct various deficiencies automatically: *Remove white nodes*—KEGG uses colors in the visualization of a pathway to annotate organism-specific orthologous genes. Nodes in green represent biological entities that occur in the current organism. Nodes in white represent biological entities, corresponding to genes that occur in this pathway in other species, but not in the current one. Translating all those nodes into new models, without caring for the node color, would lead to a model, that contains invalid genes in the pathway. *Remove orphans*—isolated nodes without any reactions or relations are usually unnecessary for further simulations. *Autocomplete reactions*—another major deficiency are incomplete reactions. The XML-files only contain those components of a reaction, that are needed for the graphical representation of the pathway. Reactants that are not necessary for the visualization are usually skipped

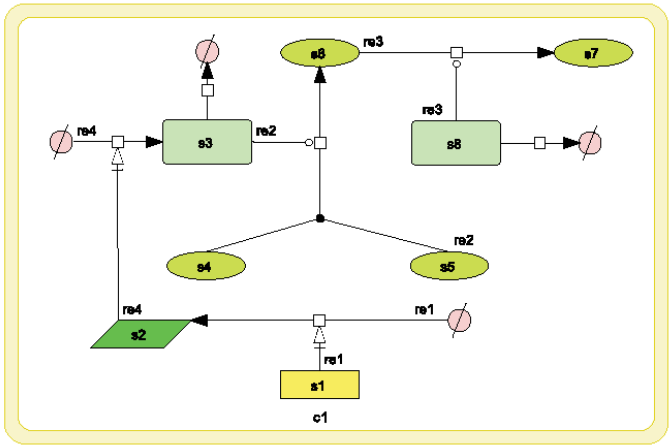
*To whom correspondence should be addressed.

http://www.cogsys.cs.uni-tuebingen.de/software/KEGGtranslator/

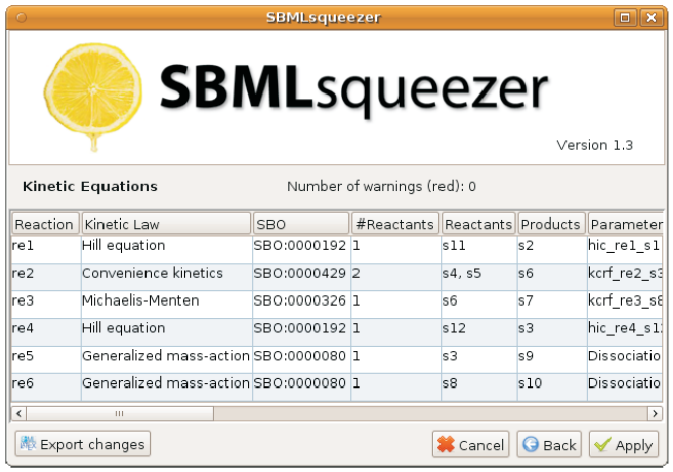
SBMLsqueezer 1.3 workflow



1



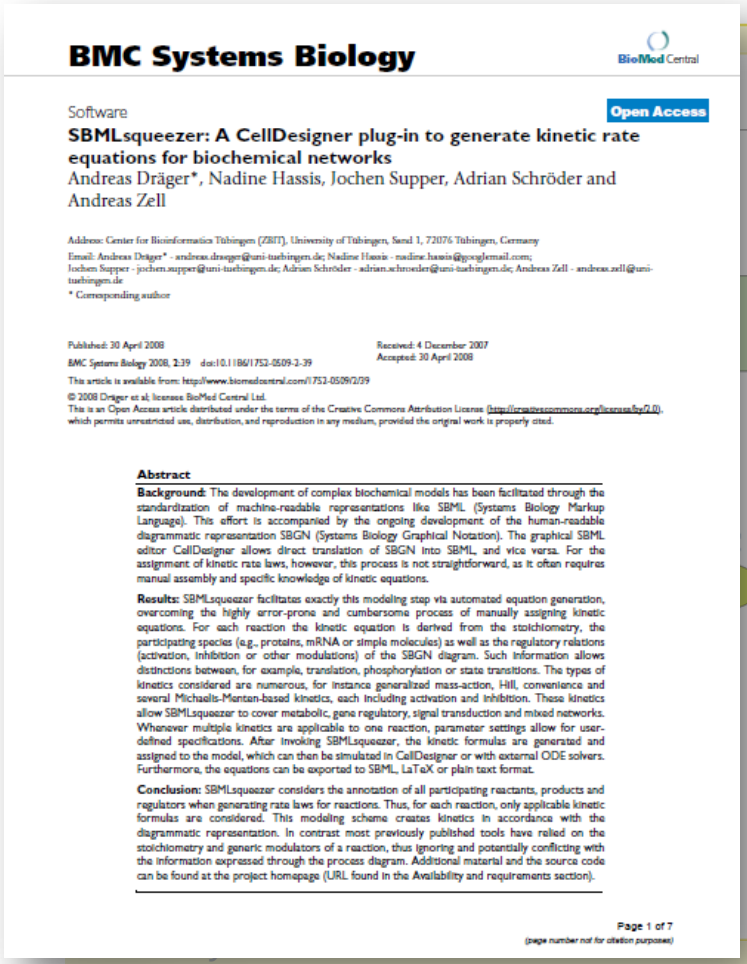
2



Powered by JSBML

http://www.cogsys.cs.uni-tuebingen.de/software/SBMLsqueezer/

SBMLsqueezer 1.3 context menu



BMC Systems Biology

BioMed Central

Software

Open Access

SBMLsqueezer: A CellDesigner plug-in to generate kinetic rate equations for biochemical networks

Andreas Dräger*, Nadine Hassis, Jochen Supper, Adrian Schröder and Andreas Zell

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

Email: Andreas Dräger* - andreas.drager@uni-tuebingen.de; Nadine Hassis - nadine.hassis@googlemail.com; Jochen Supper - jochen.supper@uni-tuebingen.de; Adrian Schröder - adrian.schroeder@uni-tuebingen.de; Andreas Zell - andreas.zell@uni-tuebingen.de

* Corresponding author

Published: 30 April 2008

BMC Systems Biology 2008, **2**:39 doi:10.1186/1752-0509-2-39

Received: 4 December 2007

Accepted: 30 April 2008

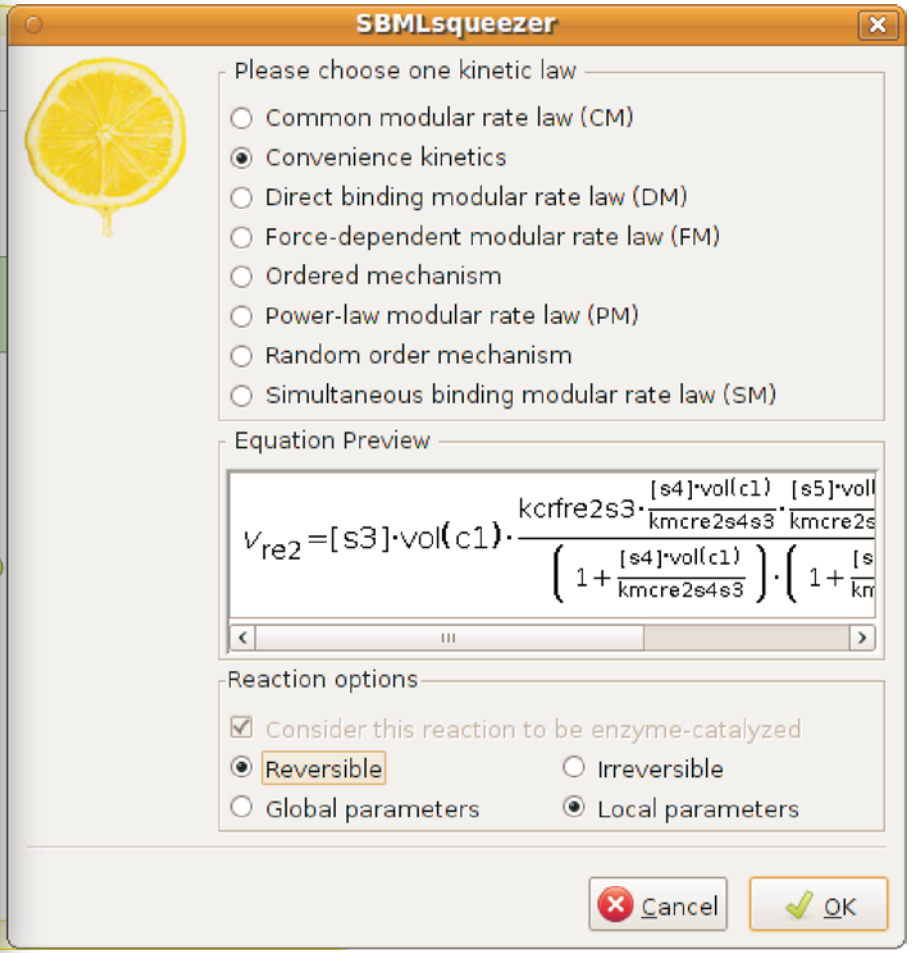
This article is available from: http://www.biomedcentral.com/1752-0509/2/39

Abstract

Background: The development of complex biochemical models has been facilitated through the standardization of machine-readable representations like SBML (Systems Biology Markup Language). This effort is accompanied by the ongoing development of the human-readable diagrammatic representation SBGN (Systems Biology Graphical Notation). The graphical SBML editor CellDesigner allows direct translation of SBGN into SBML, and vice versa. For the assignment of kinetic rate laws, however, this process is not straightforward, as it often requires manual assembly and specific knowledge of kinetic equations.

Results: SBMLsqueezer facilitates exactly this modeling step via automated equation generation, overcoming the highly error-prone and cumbersome process of manually assigning kinetic equations. For each reaction the kinetic equation is derived from the stoichiometry, the participating species (e.g., proteins, mRNA or simple molecules) as well as the regulatory relations (activation, inhibition or other modulations) of the SBGN diagram. Such information allows distinctions between, for example, translation, phosphorylation or state transitions. The types of kinetics considered are numerous, for instance generalized mass-action, Hill, convenience and several Michaelis-Menten-based kinetics, each including activation and inhibition. These kinetics allow SBMLsqueezer to cover metabolic, gene regulatory, signal transduction and mixed networks. Whenever multiple kinetics are applicable to one reaction, parameter settings allow for user-defined specifications. After invoking SBMLsqueezer, the kinetic formulas are generated and assigned to the model, which can then be simulated in CellDesigner or with external ODE solvers. Furthermore, the equations can be exported to SBML, LaTeX or plain text format.

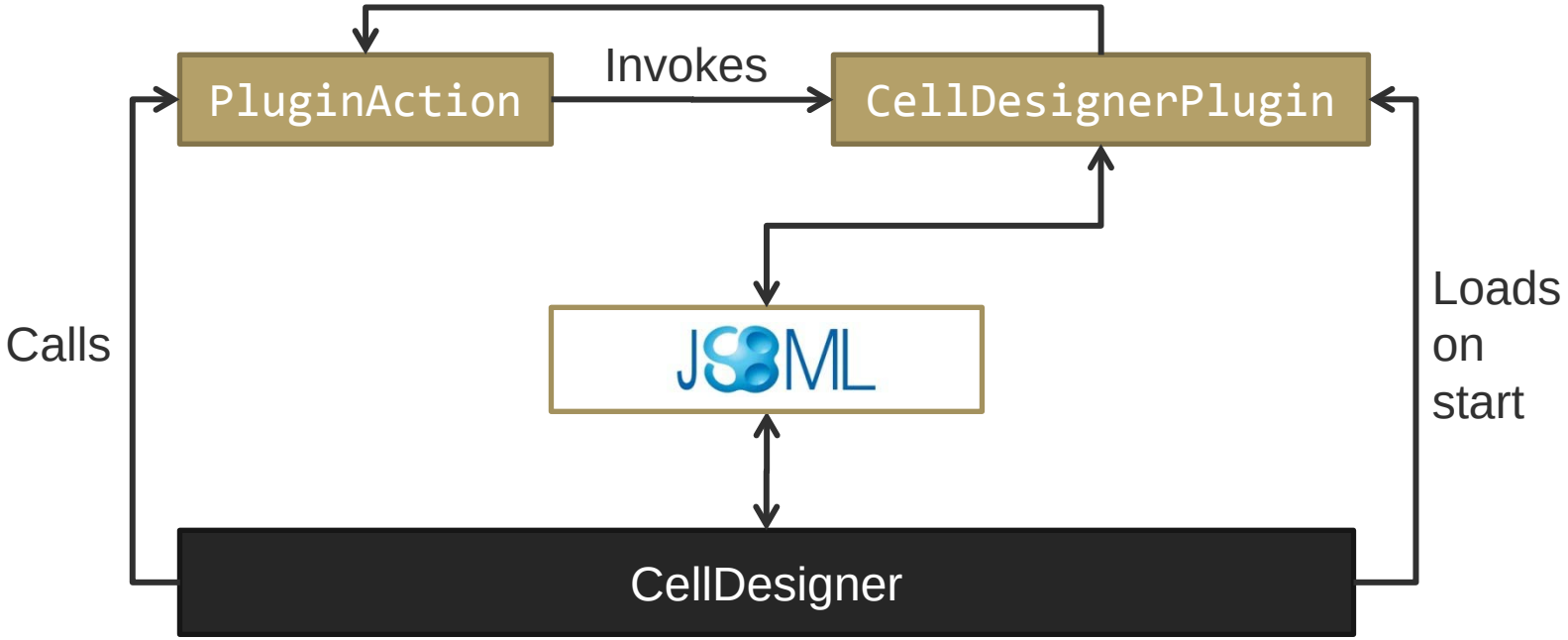
Conclusion: SBMLsqueezer considers the annotation of all participating reactants, products and regulators when generating rate laws for reactions. Thus, for each reaction, only applicable kinetic formulas are considered. This modeling scheme creates kinetics in accordance with the diagrammatic representation. In contrast most previously published tools have relied on the stoichiometry and generic modulators of a reaction, thus ignoring and potentially conflicting with the information expressed through the process diagram. Additional material and the source code can be found at the project homepage (URL found in the Availability and requirements section).



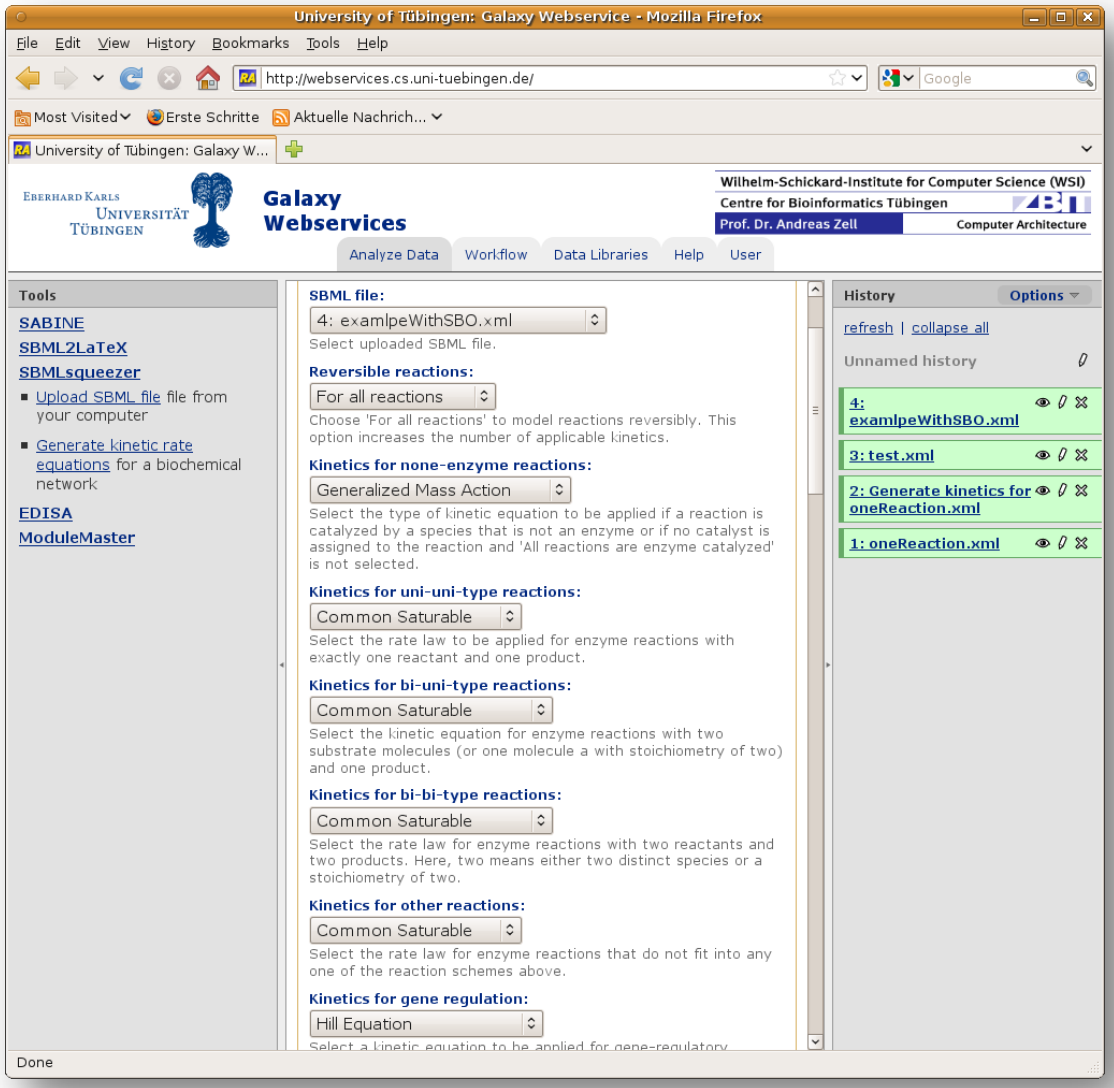
http://www.cogsys.cs.uni-tuebingen.de/software/SBMLsqueezer/

JSBML's CellDesigner module

- Turning an existing application into a plugin for CellDesigner
- Only implementation of two abstract classes required

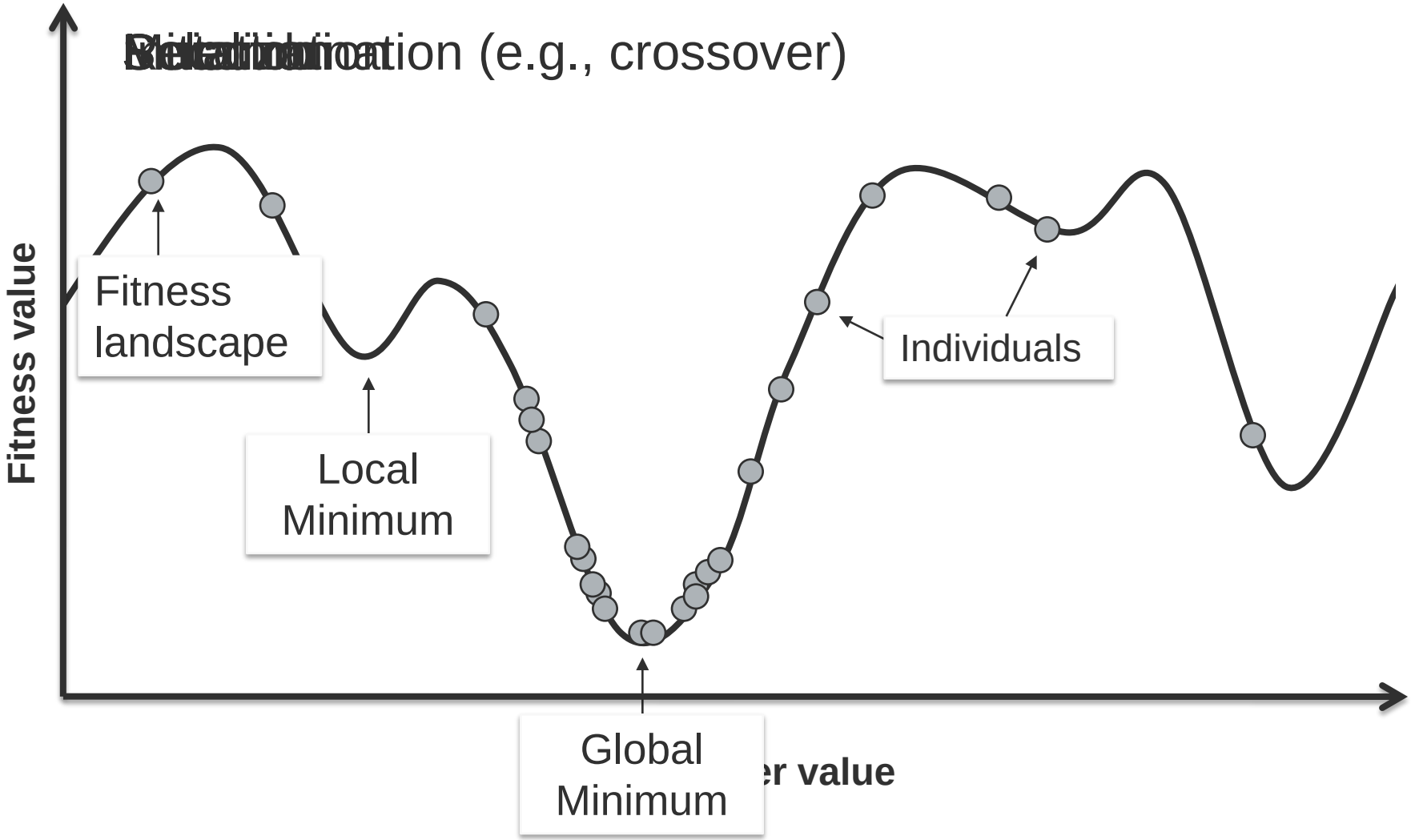


Online version of SBMLsqueezer

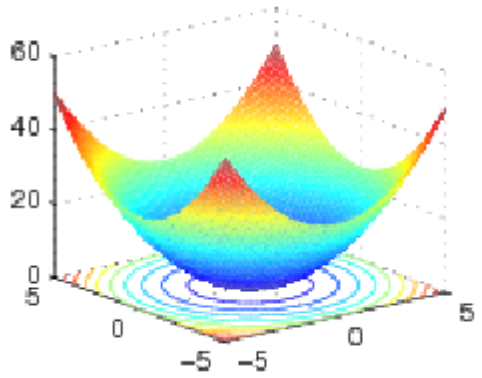


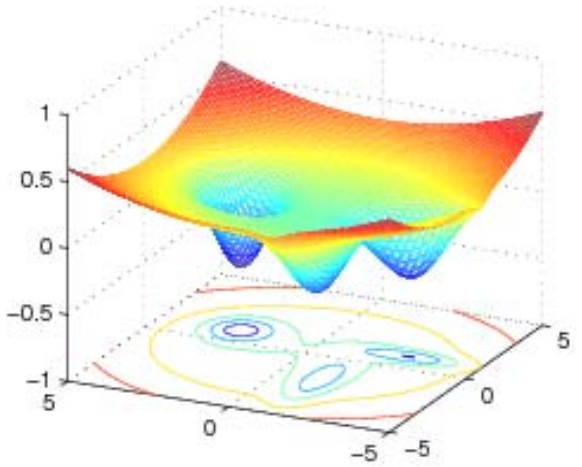
http://webservices.cs.uni-tuebingen.de

Principle of evolutionary algorithms



Multimodal optimization problems





Problem:

Often many local optima of similar quality

Therefore:

Often consideration of additional side constraints important

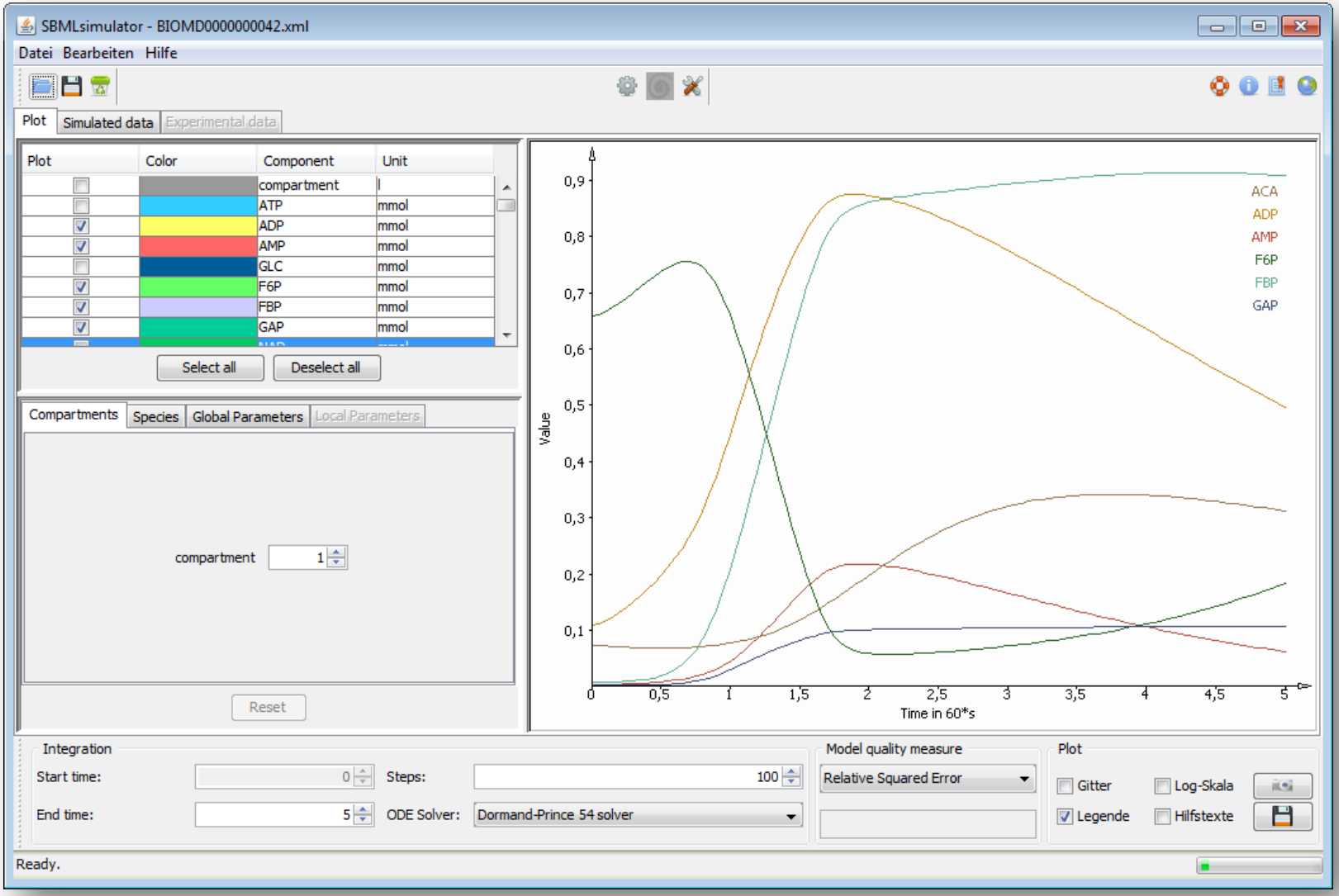
Parameter fitting with the EvA2

- Workbench for naturally-inspired heuristic optimization procedures
- Implemented in Java™
- EvA2 is a:
  - platform independent optimization
  - development platform for software
- Provides both
  - Graphical User Interface (GUI) an
  - Abstract Programming Interface (A

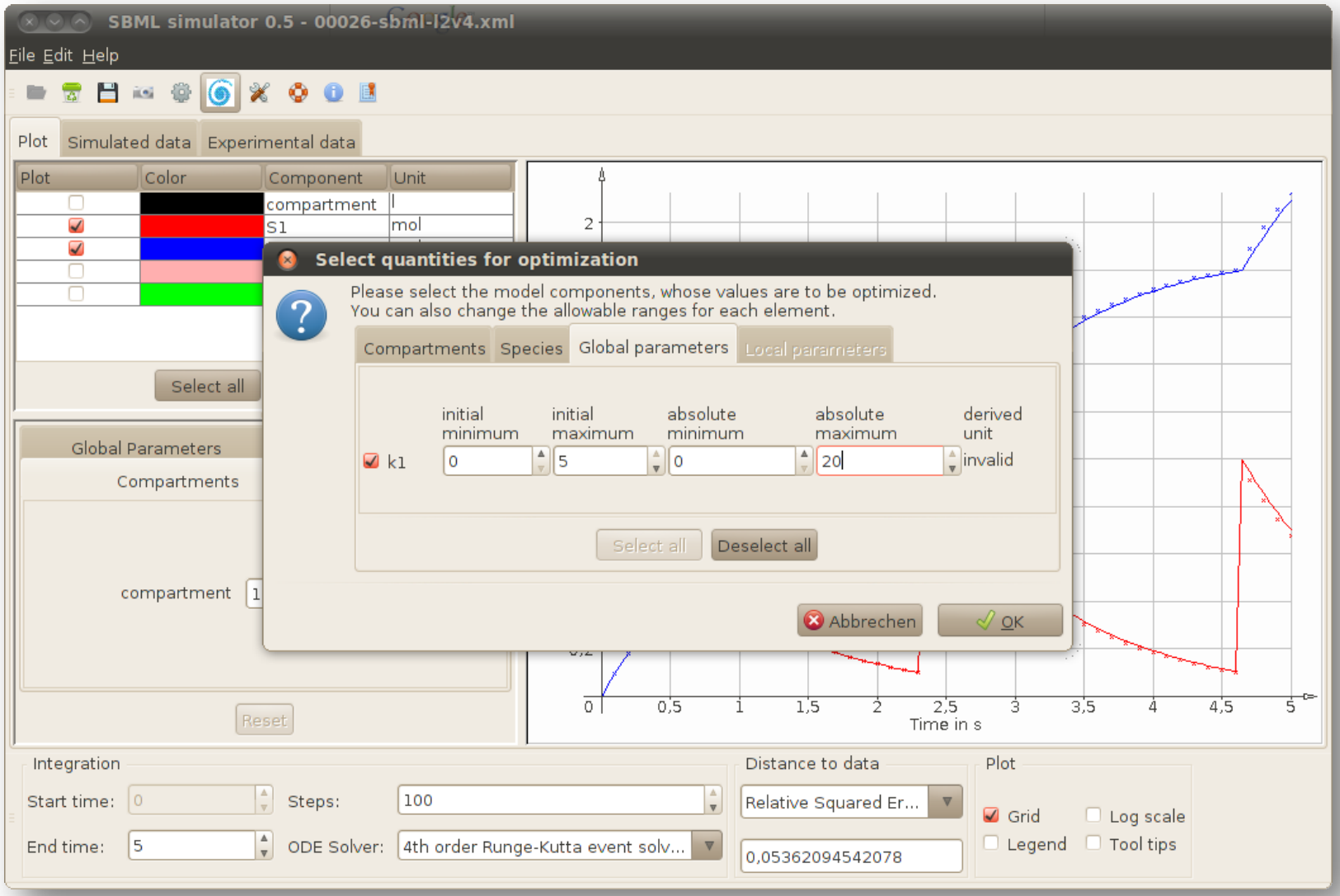
http://www.cogsys.cs.uni-tuebingen.de/software/EvA2

Bringing networks to life with SBMLsimulator

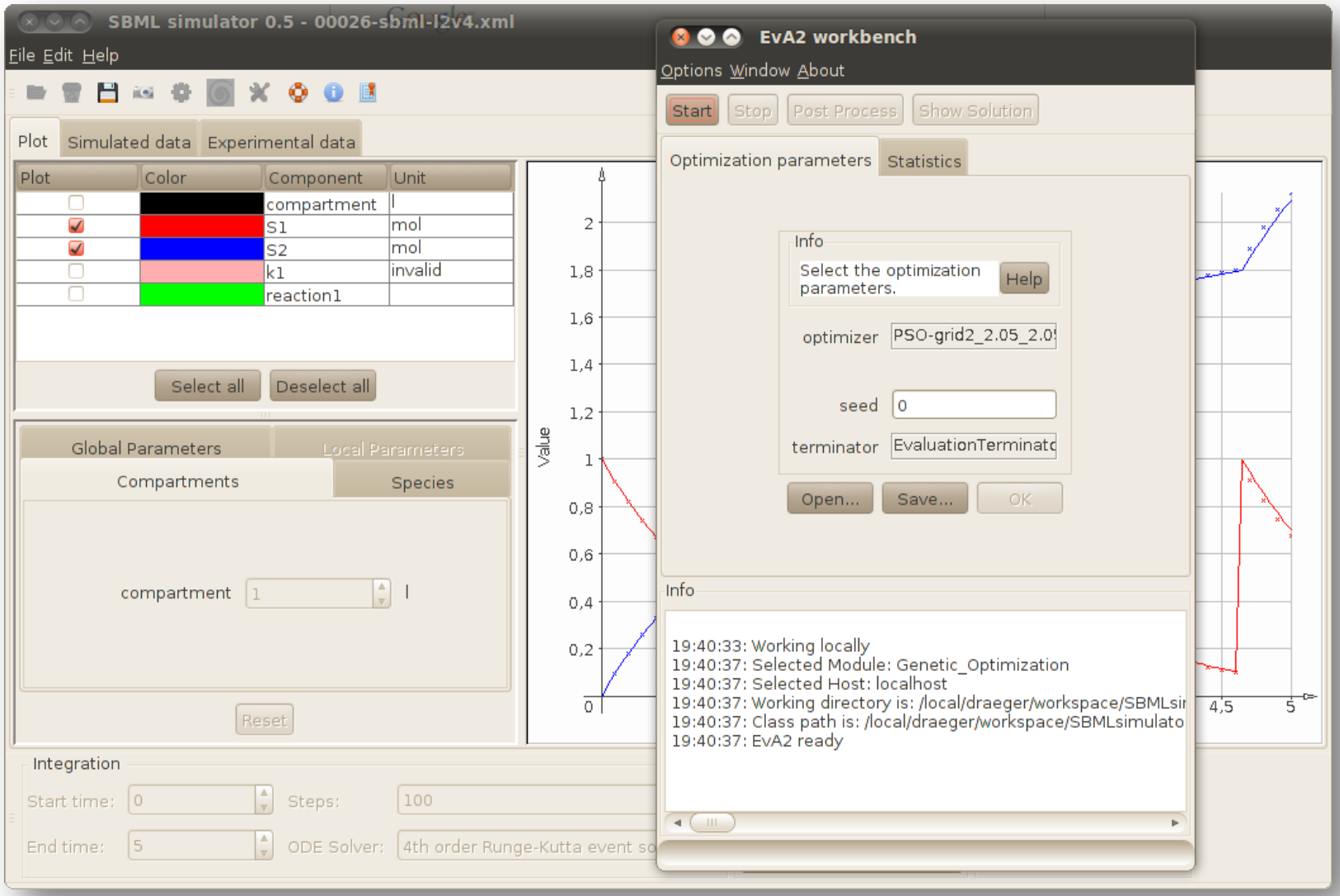
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Parameter estimation in SBMLsimulator



Using EvA2 from SBMLsimulator



SBMLsimulator

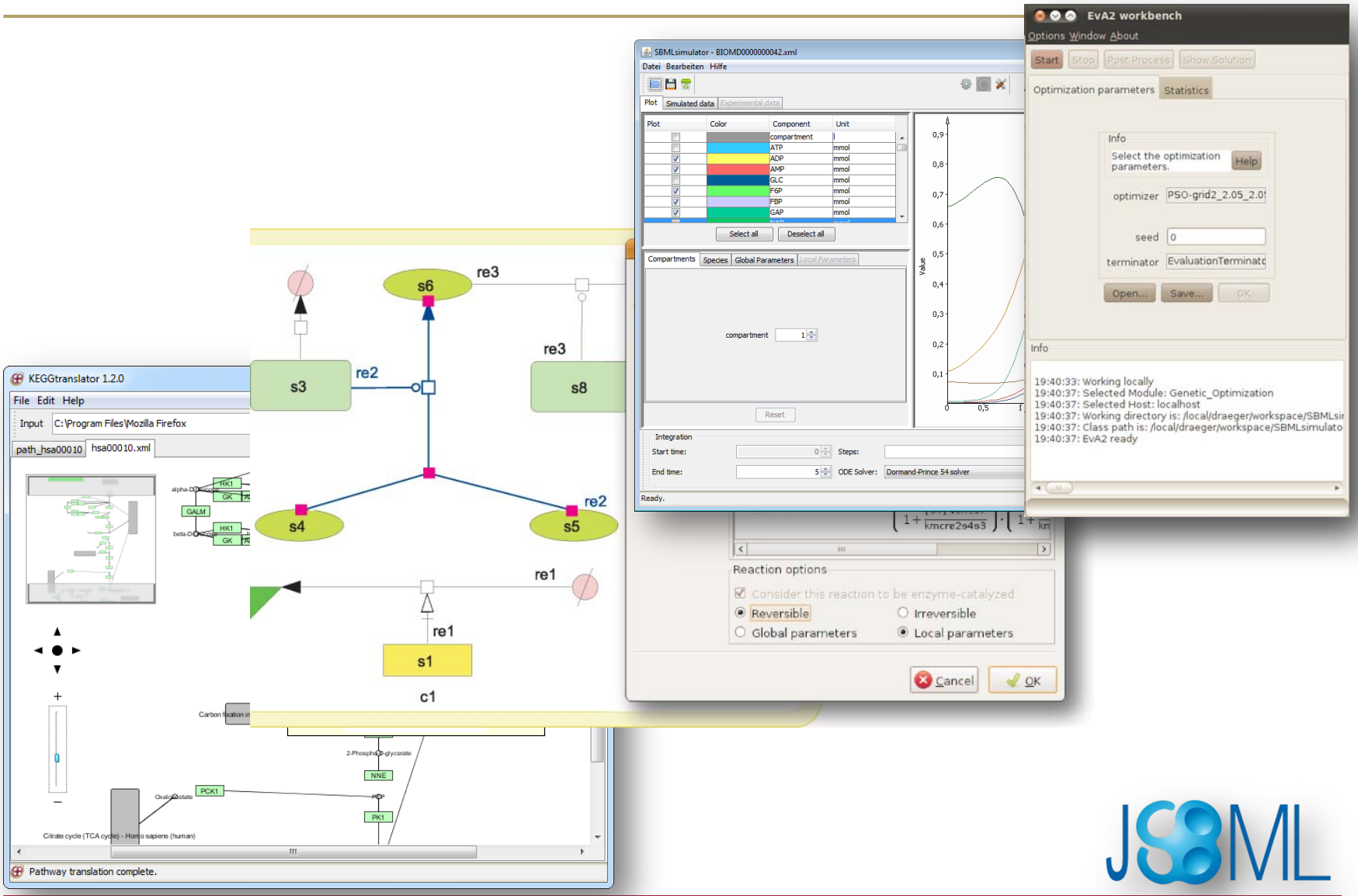
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

- Two parts: core and application
- Sourceforge.net
- Solvers:
  - Apache package (non stiff)
  - Rosenbrock solver for stiff equations
  - No implementation of LSODA available at the moment

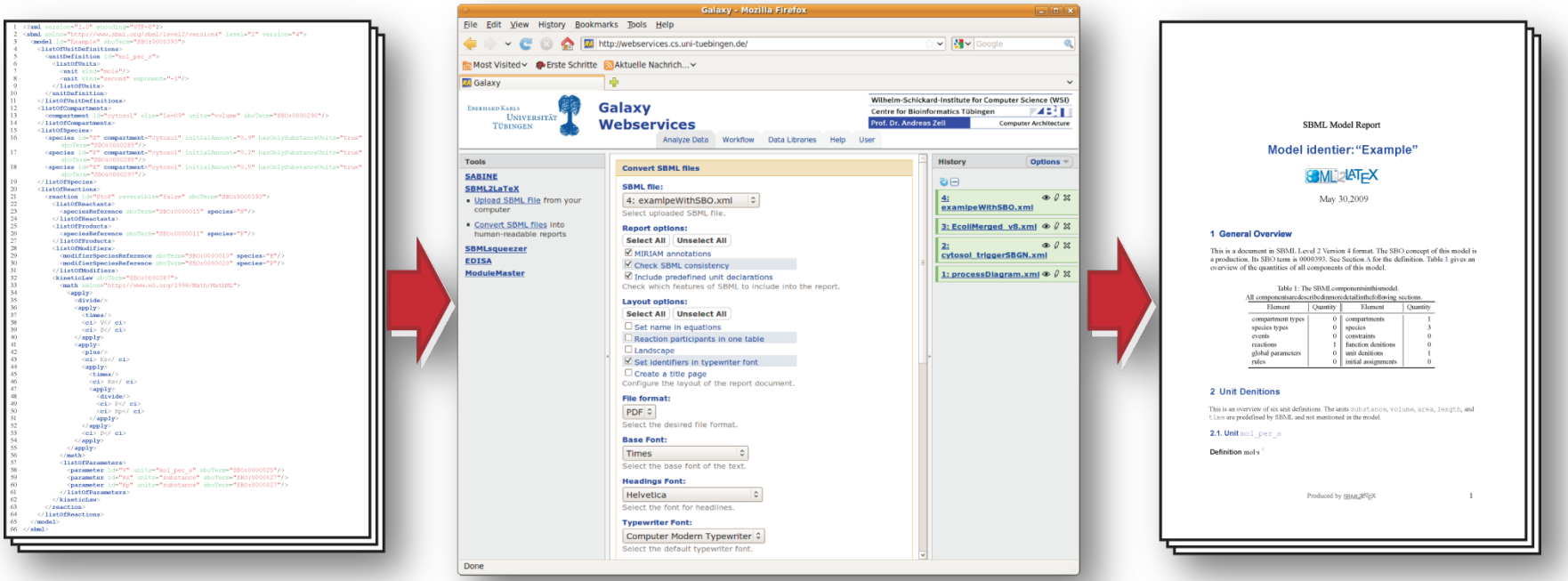
Current work

- Support for SED-ML
- Support for SBRML
- Running time improvement (mainly JSBML‘s ASTNode implementation)
- Inclusion of stochastic methods
- Aim: Support for all SBML test cases and all Biomodels.net

Model generation pipeline



SBML2LaTeX: Model documentation



http://webservices.cs.uni-tuebingen.de

Acknowledgments

- Roland Keller
- Clemens Wrzodek

- Students:
  - Alexander Dörr
  - Sarah Müller vom Hagen
  - Max Zwießele
  - Philip Stevens

SPONSORED BY THE
Federal Ministry of Education and Research

Baden-Württemberg

virtual liver network