

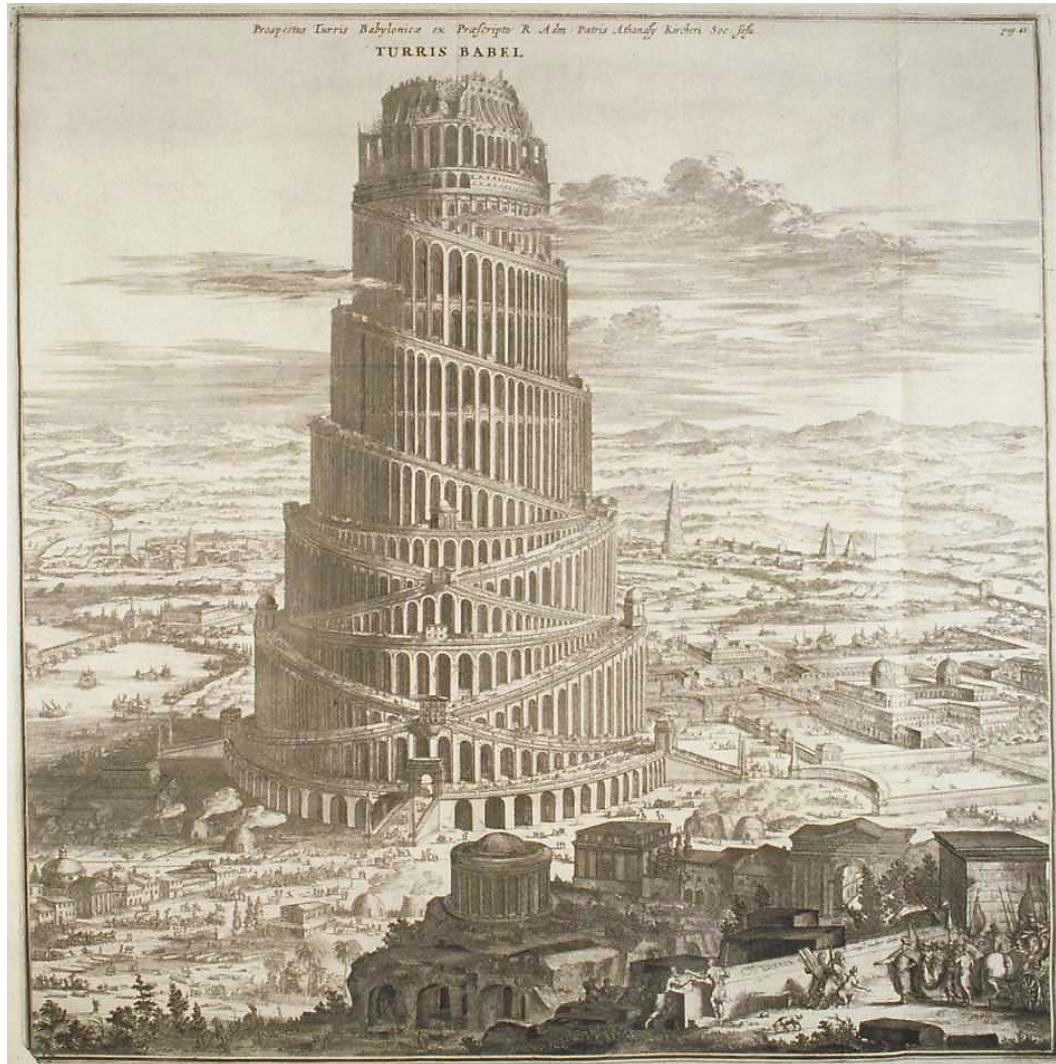
Fighting individualism -

Standard formats for Systems Biology

Wolfram Liebermeister
Theoretische Biophysik

Humboldt-Universität zu Berlin

Big project of the past



Big project of the past



Yahoo! Babel Fish - Text Translation and Web Page

File Edit View History Bookmarks Tools Help

← → ↻ 🏠 📄 h ☆ Reflect G ba

SPIEGEL ONLINE - Nachri... Yahoo! Babel Fish - Text ...

New User? Sign Up | Sign In | Help

Yahoo! Babel Fish - Text Transl...

YAHOO! BABEL FISH

Auf deutsch

Lieber Joerg und Uwe, Die letzten CcpA Span-Span Daten wurden von NimbleGen am Montag empfangen,

Translate again 🐟 (Enter up to 150 words)

Dear Joerg and Uwe,

The last CcpA CHIP-chip data was received from NimbleGen on Monday,

English to German

Find:

Done


Big project of the past



Big projects today

“Most of the published quantitative models in biology are lost for the community because they are either not made available or they are insufficiently characterized to allow them to be reused.”

Le Novère *et al*, (2005)



computational
BIOLOGY

PERSPECTIVE

Minimum information requested in the annotation of biochemical models (MIRIAM)

Nicolas Le Novère^{1,15}, Andrew Finney^{2,15}, Michael Hucka³, Upinder S Bhalla⁴, Fabien Campagne⁵, Julio Collado-Vides⁶, Edmund J Crampin⁷, Matt Halstead⁷, Edda Klipp⁸, Pedro Mendes⁹, Poul Nielsen⁷, Herbert Sauro¹⁰, Bruce Shapiro¹¹, Jacky L Snoep¹², Hugh D Spence¹³ & Barry L Wanner¹⁴

Most of the published quantitative models in biology are lost for the community because they are either not made available or they are insufficiently characterized to allow them to be reused. The lack of a standard description format, lack of stringent reviewing and authors' carelessness are the main causes for incomplete model descriptions. With today's increased interest in detailed biochemical models, it is necessary to define a minimum quality standard for the encoding of those models. We propose a set of rules for curating quantitative models of biological systems. These rules define procedures for encoding and annotating models represented in machine-readable form. We believe their application will enable users to (i) have confidence that curated models are an accurate reflection of their associated reference descriptions, (ii) search collections of curated models with precision, (iii) quickly identify the biological phenomena that a given curated model or model constituent represents and (iv) facilitate model reuse and composition into large subcellular models.

During the genomic era we have witnessed a vast increase in availability of large amounts of quantitative data. This is motivating a shift in the focus of molecular and cellular research from qualitative descriptions of biochemical interactions towards the quantification of such interactions and their dynamics. One of the tenets of systems biology is the use of quantitative models (see **Box 1** for definitions) as a mechanism for capturing precise hypotheses and making predictions^{1,2}. Many specialized models exist that attempt to explain aspects of the cellular machinery. However, as has happened with other types of biological information, such as sequences, macromolecular structures or

Box 1 Glossary

Some terms are used in a very specific way throughout the article. We provide here a precise definition of each one.

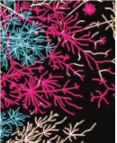
Quantitative biochemical model. A formal model of a biological system, based on the mathematical description of its molecular and cellular components, and the interactions between those components.

© 2005 Nature Publishing Group <http://www.nature.com/naturebiotechnology>

Model publishing: the MIRIAM rules

computational BIOLOGY

PERSPECTIVE



Minimum information requested in the annotation of biochemical models (MIRIAM)

Nicolas Le Novère^{1,15}, Andrew Finney^{2,15}, Michael Hucka³, Upinder S Bhalla⁴, Fabien Campagne⁵, Julio Collado-Vides⁶, Edmund J Crampin⁷, Matt Halstead⁷, Edda Klipp⁸, Pedro Mendes⁹, Poul Nielsen⁷, Herbert Sauro¹⁰, Bruce Shapiro¹¹, Jacky L Snoep¹², Hugh D Spence¹³ & Barry L Wanner¹⁴

Most of the published quantitative models in biology are lost for the community because they are either not made available or they are insufficiently characterized to allow them to be reused. The lack of a standard description format, lack of stringent reviewing and authors' carelessness are the main causes for incomplete model descriptions. With today's increased interest in detailed biochemical models, it is necessary to define a minimum quality standard for the encoding of those models. We propose a set of rules for curating quantitative models of biological systems. These rules define procedures for encoding and annotating models represented in machine-readable form. We believe their application will enable users to (i) have confidence that curated models are an accurate reflection of their associated reference descriptions, (ii) search collections of curated models with precision, (iii) quickly identify the biological phenomena that a given curated model or model constituent represents and (iv) facilitate model reuse and composition into large subcellular models.

During the genomic era we have witnessed a vast increase in availability of large amounts of quantitative data. This is motivating a shift in the focus of molecular and cellular research from qualitative descriptions of biochemical interactions towards the quantification of such interactions and their dynamics. One of the tenets of systems biology is the use of quantitative models (see Box 1 for definitions) as a mechanism for capturing precise hypotheses and making predictions^{1,2}. Many specialized models exist that attempt to explain aspects of the cellular machinery. However, as has happened with other types of biological information, such as sequences, macromolecular structures or

Box 1 Glossary

Some terms are used in a very specific way throughout the article. We provide here a precise definition of each one.

Quantitative biochemical model. A formal model of a biological system, based on the mathematical description of its molecular and cellular components, and the interactions between those components.

© 2005 Nature Publishing Group <http://www.nature.com/naturebiotechnology>

computational BIOLOGY

COMMENTARY

Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project

Chris F Taylor^{1,2}, Dawn Field^{2,3}, Susanna-Assunta Sansone^{1,2}, Jan Aerts⁴, Rolf Apweiler¹, Michael Ashburner⁵, Catherine A Ball⁶, Pierre-Alain Binz^{7,8}, Molly Bogue⁹, Tim Booth², Alvis Brazma¹, Ryan R Brinkman¹⁰, Adam Michael Clark¹¹, Eric W Deutsch¹², Oliver Fiehn¹³, Jennifer Fostel¹⁴, Peter Ghazal¹⁵, Frank Gibson¹⁶, Tanya Gray^{2,3}, Graeme Grimes¹⁵, John M Hancock¹⁷, Nigel W Hardy¹⁸, Henning Hermjakob¹, Randall K Julian Jr¹⁹, Matthew Kane²⁰, Carsten Kettner²¹, Christopher Kinsinger²², Eugene Kolker^{23,24}, Martin Kuiper²⁵, Nicolas Le Novère¹, Jim Leebens-Mack²⁶, Suzanna E Lewis²⁷, Phillip Lord¹⁶, Ann-Marie Mallon¹⁷, Nishanth Marthandan²⁸, Hiroshi Masuya²⁹, Ruth McNally³⁰, Alexander Mehrle³¹, Norman Morrison^{2,32}, Sandra Orchard¹, John Quackenbush³³, James M Reecy³⁴, Donald G Robertson³⁵, Philippe Rocca-Serra^{1,36}, Henry Rodriguez²², Heiko Rosenfelder³¹, Javier Santoyo-Lopez¹⁵, Richard H Scheuermann²⁸, Daniel Schober¹, Barry Smith³⁷, Jason Snape³⁸, Christian J Stockert Jr³⁹, Keith Tipton⁴⁰, Peter Sterk¹, Andreas Untergasser⁴¹, Jo Vandesompele⁴² & Stefan Wiemann³¹

The Minimum Information for Biological and Biomedical Investigations (MIBBI) project provides a resource for those exploring the range of extant minimum information checklists and fosters coordinated development of such checklists.

To fully understand the context, methods, data and conclusions that pertain to an experiment, one must have access to a range of background information. However, the current guidelines for reporting proteomics experiments and describing systems biology models are gaining broader support in their respective database communities⁴³; and progress is overlaps in scope and arbitrary decisions on wording and substructuring inhibit their use in combination. These issues present difficulties for checklist users, especially those who

© 2005 Nature Publishing Group <http://www.nature.com/naturebiotechnology>

MIRIAM resources <http://www.ebi.ac.uk/miriam/>

<http://mibbi.org/>

Quality aspects

- Quality of documentation (=paper)
- Correspondence model / documentation
- Accuracy and extent of annotations
- Encoding in a machine-readable form

Annotation schemes

- Attribution:
 - Reference description, authors, creators
- External data resources:
 - {data type, identifier, qualifier}

Model publishing: the MIRIAM rules

Rules for reference correspondence

- Encoding in a public, machine-readable format
- Comply with the standard in which model is encoded.
- Clear relation to a single reference description
- Model structure must reflect biological processes listed in the reference description
- Must be instantiated in a simulation.
- Reproduce all relevant results from reference description

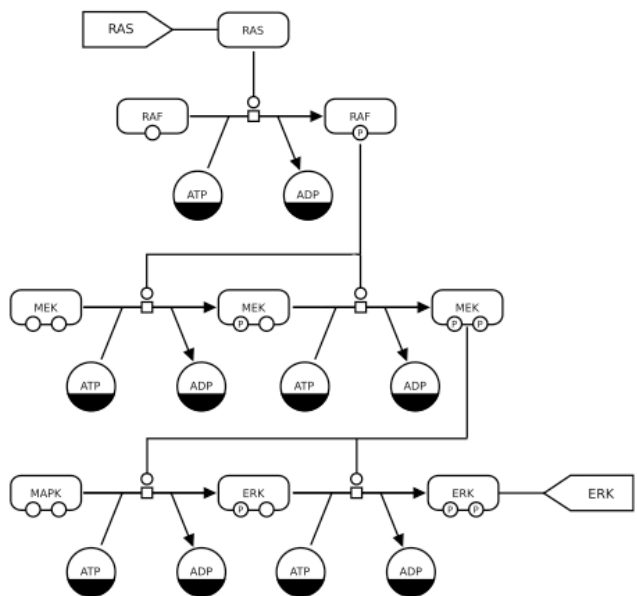
Mandatory annotations

- Preferred model name
- Citation of reference description
- Name and contact information for model creators (who *encoded* the model)
- Date and time of creation + last modification
- Precise statement about terms of distribution

“Esperanto Formats”
for Systems Biology Models

SBGN, the Systems Biology Graphical Notation

“Process description”



SYSTEMS BIOLOGY GRAPHICAL NOTATION REFERENCE CARD

<p>Entity Pool Nodes</p> <ul style="list-style-type: none"> LABEL unspecified entity LABEL simple chemical LABEL macromolecule LABEL nucleic acid feature LABEL multimer LABEL source/sink LABEL tag LABEL observable LABEL perturbation 	<p>Auxiliary units</p> <ul style="list-style-type: none"> pre:label unit of information val@var state variable LABEL marker LABEL clone marker 	<p>Process Nodes</p> <ul style="list-style-type: none"> transition omitted process uncertain process association dissociation 	<p>Connecting Arcs</p> <ul style="list-style-type: none"> consumption production modulation stimulation catalysis inhibition trigger logic arc equivalence arc
<p>Container Nodes</p> <ul style="list-style-type: none"> complex compartment submap 		<p>Logical Operators</p> <ul style="list-style-type: none"> AND and operator OR or operator NOT not operator 	

SBML, the Systems Biology Markup Language

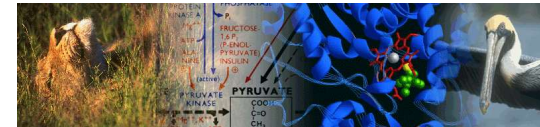


One exchange format - about 170 tools that understand each other

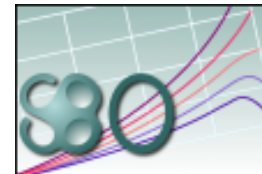
```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level2/version3" level="2" version="3">
  <model id="model" name="model">
    <listOfCompartments>
      <compartment id="c" name="c" size="1"/>
      <compartment id="ext" name="ext" size="1"/>
    </listOfCompartments>
    <listOfSpecies>
      <species id="C00022_c" name="Pyruvate" compartment="c"> </species>
      ...
      ...
      ...
    <reaction id="reaction_8">
      <listOfReactants>
        <speciesReference species="C00022_c" stoichiometry="0.03"/>
        ...
        <speciesReference species="O2_c" stoichiometry="0.01"/>
      </listOfReactants>
      <listOfProducts>
        <speciesReference species="C00008_c" stoichiometry="0.81"/>
        ...
      </listOfProducts>
      <listOfModifiers>
        <modifierSpeciesReference species="enzyme_reaction_8_c"/>
      </listOfModifiers>
    </reaction>
  </listOfReactions>
</model>
</sbml>
```



Database of curated annotated models
<http://biomodels.org/>



JWS online: database of curated models
<http://jjj.biochem.sun.ac.za/>



Systems biology ontology
<http://www.ebi.ac.uk/sbo/>

SBML main site <http://sbml.org/>



Systems Biology Operational Software

Ubuntu Linux 8.10 booting directly from DVD

Includes

- Preinstalled software tools for SBML models
- Models from the BioModels.net database
- Documentation and video tutorials

Allows for

model building, layout, simulation, fitting, annotation, merging

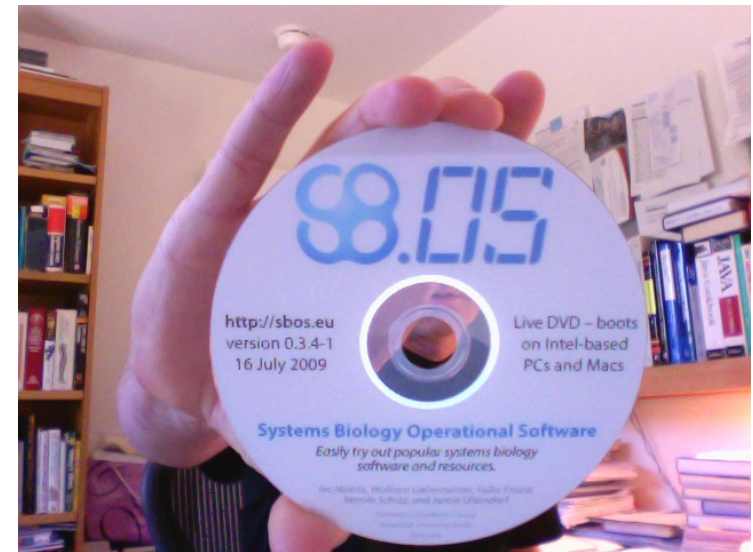
Licenses:

All software free for non-commercial use

For commercial use, licenses depend on the tools

Download of DVD image file

and further information at <http://www.sbos.eu/>



SB.OS / Systems Biology Operational Software

home download contribute internal edit

Welcome to SB.OS - Systems Biology Operational Software

SB.OS is a live DVD based on Ubuntu Linux that comes with a comprehensive list of Systems Biology Software. Text and video documentation material, as well as an offline copy of the BioModels.net database, are included.

Screenshots

You will find a screenshot here.

Download

You can download an image file of the current SB.OS DVD here.

Creating a Bootable DVD / USB Stick

You can run SB.OS on your computer without having to install anything. To create a bootable DVD just burn the disk image with your favourite burning software to a DVD. To create a bootable USB stick use [UNetbootin](#) together with our disk image. All you have to do now is to change the boot device of your computer to first boot from your DVD drive (this is the default on most computers) or to USB. Insert the DVD/USB stick and boot your computer.

start1

Software on SB.OS

- Included (version 3.3)
- Pending
- Excluded

About SB.OS

- SB.OS Documentation
- Team

Links

- BiLinux
- Theoretical Biophysics
- HU Berlin

Wiki Editing

- Syntax
- Edit

Annotating the model elements

Biological annotations in SBML

SBO term

www.ebi.ac.uk/sbo/

Species called "enzyme_R00001" represents an enzyme

```
<species id="enzyme_R00001" sboTerm="SBO:0000014"/>
```

MIRIAM annotation

<http://www.ebi.ac.uk/miriam/>

Species called "ATP" represents KEGG C06262 (ATP)

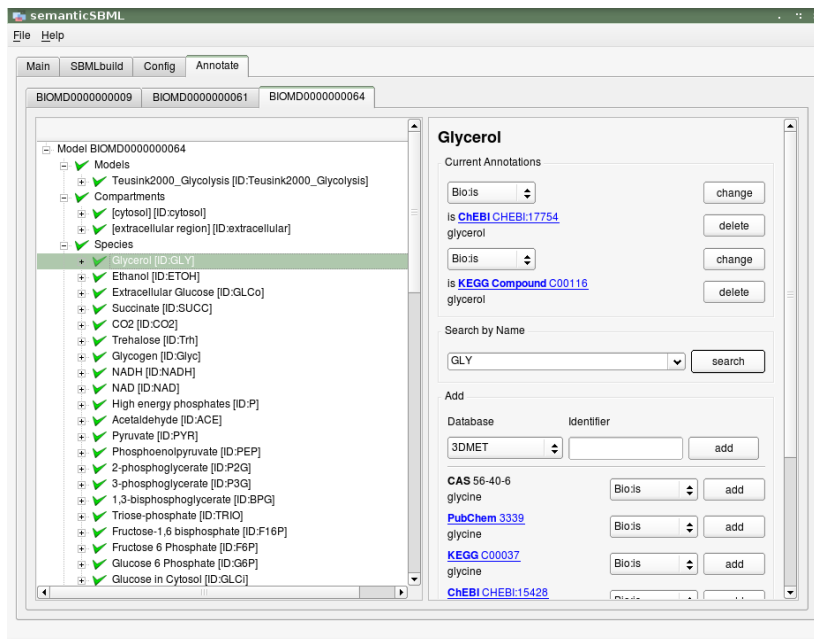
```
<species metaid=".." id="ATP" name="ATP concentration" compartment="cytosol">
  <annotation>
    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      xmlns:bqbiol="http://biomodels.net/biology-qualifiers/"
      xmlns:bqmodel="http://biomodels.net/model-qualifiers/">
      <rdf:Description rdf:about="#metaid_0000076">
        <bqbiol:is>
          <rdf:Bag>
            <rdf:li rdf:resource="urn:miriam:obo.chebi:CHEBI%3A15422"/>
            <rdf:li rdf:resource="urn:miriam:kegg.compound:C00002"/>
          </rdf:Bag>
        </bqbiol:is>
      </rdf:Description>
    </rdf:RDF>
  </annotation>
  ...
</species>
```

Species Qualifier Database Identifier

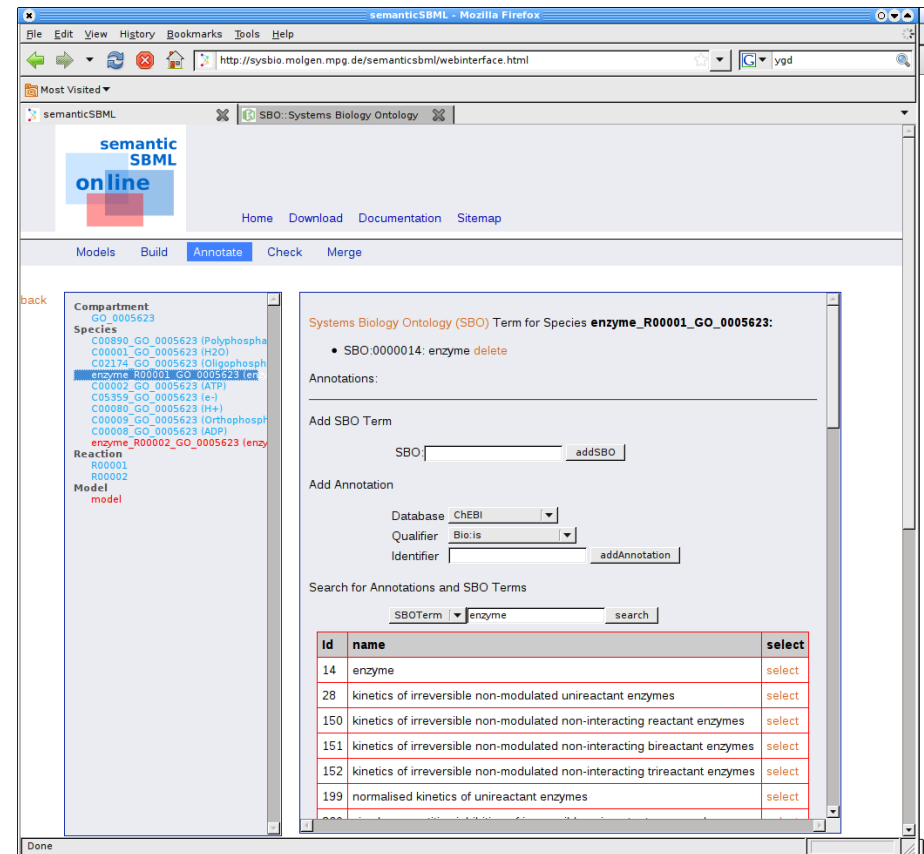
"A simple scheme for annotating SBML with references to controlled vocabularies and database entries" Le Novère and Finney, 2005

SemanticSBML: Annotation and merging of SBML models

Stand-alone version



Online version



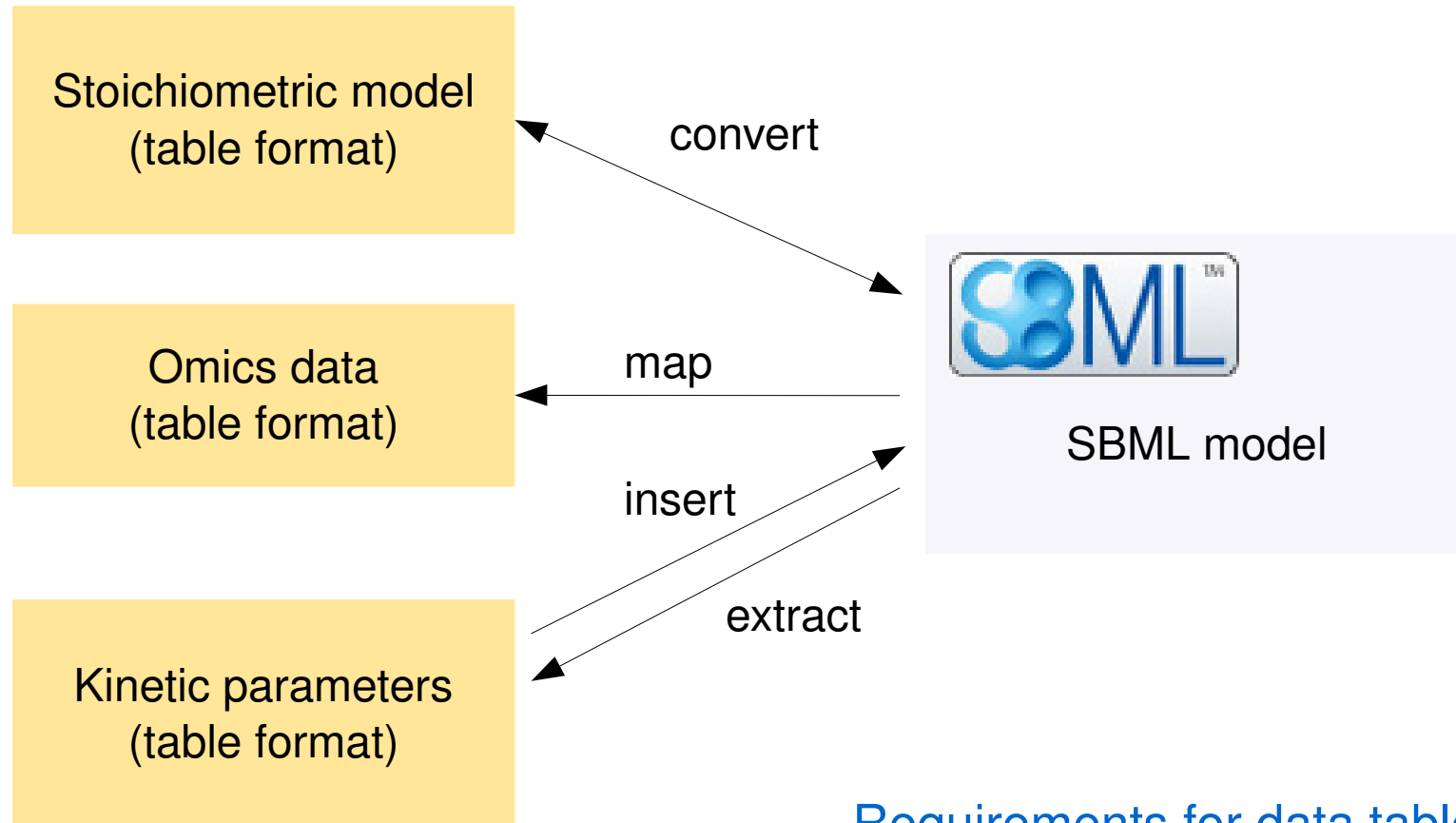
www.semanticsbml.org

... included in



Own efforts

Data conversion in model building: table formats



Requirements for data table format

- Completeness
- Annotations
- Defined structure
- Compatibility
- Intuitive format

SBtab: our proposal for standard table formats

Reaction*	ReactionFormula	ID:kegg.reaction	GeneName	EnzymeRegulation
R1	ATP + F6P <=> ADP + F16P	R00658	pfk	
R2	F16P + H2O <=> F6P + Pi	R01015	fbp	+ PEP – AMP

Compound*	Name	ID:kegg.compound
F6P	Fructose-6-phosphate	C05345
ATP	ATP	C00002
ADP	ADP	C00008
F16P	Fructose-1,6-bisphosphate	C00354
H2O	Water	C00001
Pi	Inorganic phosphate	C00009
PEP	Phosphoenolpyruvate	C00074
AMP	AMP	C00020

Compound*	CompoundID:obo.chebi	's1 Mean'	's1 Std'	's2 Mean'	's2 Std'
#Quantity		s1	s1	s2	s2
#MathDescriptor		Mean	StdDev	Mean	StdDev
4abut	CHEBI:16865	27.5	0.1	28.2	0.1
fum	CHEBI:18012	0.13	0.1	0.15	0.1
succ	CHEBI:15741	0.17	0.1	0.19	0.1

SBtab: our proposal for standard table formats

Defined column names Defined syntax for reactions and regulation MIRIAM-compliant annotations

Reaction*	ReactionFormula	ID:kegg.reaction	GeneName	EnzymeRegulation
R1	ATP + F6P <=> ADP + F16P	R00658	pfk	
R2	F16P + H2O <=> F6P + Pi	R01015	fbp	+ PEP - AMP

Compound*	Name	ID:kegg.compound
F6P	Fructose-6-phosphate	C05345
ATP	ATP	C00002
ADP	ADP	C00008
F16P	Fructose-1,6-bisphosphate	C00354
H2O	Water	C00001
Pi	Inorganic phosphate	C00009
PEP	Phosphoenolpyruvate	C00074
AMP	AMP	C00020

Similar to existing formats

Structured: easy to parse

Compound*	CompoundID:obo.chebi	's1 Mean'	's1 Std'	's2 Mean'	's2 Std'
#Quantity		s1	s1	s2	s2
#MathDescriptor		Mean	StdDev	Mean	StdDev
4abut	CHEBI:16865	27.5	0.1	28.2	0.1
fum	CHEBI:18012	0.13	0.1	0.15	0.1
succ	CHEBI:15741	0.17	0.1	0.19	0.1

Draft specification (formerly called “biotables”) is on the BaSysBio wiki

Online validator is still under construction ...

We are still working on the format, suggestions are appreciated!!

Standardised rate laws

Reaction



Mass-action $v = k^+ a b - k^- c$

Linlog $\frac{v}{v_0} = \frac{u}{u_0} \left(1 + E_A \ln \frac{a}{a_0} + E_B \ln \frac{b}{b_0} + E_C \ln \frac{c}{c_0} \right)$

TKM $v = R^{-1} (\xi_A \xi_B - \xi_C) \quad \xi_i = e^{\mu_i/RT}$

Convenience kinetics $v = u \frac{k^+ (a/k_A^M) (b/k_B^M) - k^- (c/k_C^M)}{(1 + a/k_A^M)(1 + b/k_B^M) + c/k_C^M}$

Possible requirements

- Simple and unique
- Cover various stoichiometries, regulation, cooperativity
- Saturable ?
- Correct thermodynamics ?
- Mechanistic interpretation ?
- Agree well with SBML, SBGN, ... ?

Our standardised rate laws

3 thermodynamic versions (numerator)

Standard chemical potentials (satisfy Wegscheider cond.)
Equilibrium constants (satisfy Haldane relationships)
Catalytic rate constants

Cooperativity (exponents)

Thermodynamically correct formulas with Hill-like exponents

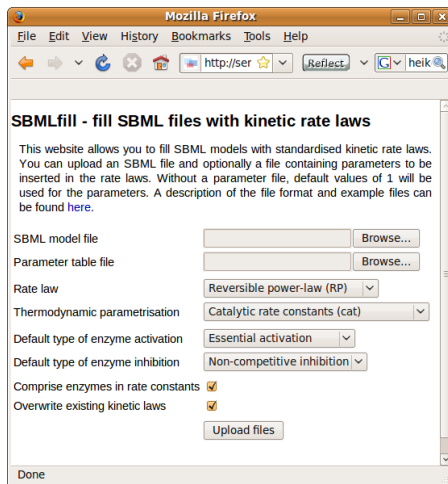
$$v(a, b, c, u) = u \frac{k^+ (a/k_A^M) (b/k_B^M) - k^- (c/k_C^M)}{(1 + a/k_A^M)(1 + b/k_B^M) + (1 + c/k_C^M) - 1}$$

4 types of regulation

Inhibition (non-competitive)
Inhibition (competitive)
Activation (essential)
Activation (non-essential)

5 rate laws (denominator)

'Reversible power-law': mass-action, power law
'Common saturable': similar to convenience kinetics, Michaelis-Menten
'Direct saturable': simplified version of common saturable
'Multiplicative saturable': simplified version of common saturable
'Force-dependent': nice thermodynamic properties

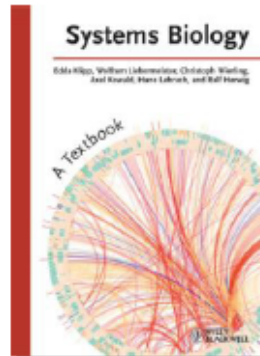


Web interface for SBML models
under development

Text book “Systems Biology”

KEY TITLE

Molecular Life Sciences



**FINALLY
AVAILABLE AS
TEXTBOOK!**

This course book in systems biology is tailored to the needs of advanced students of biology, engineering, and computer science.

It has a companion website with solutions to questions in the book, additional chapters, and computer implementations of systems biology models.

The book is a follow-up of the very successful *Systems Biology in Practice* and incorporates the feedback and suggestions of many lecturers worldwide.

The interdisciplinary team of acclaimed authors have worked closely together to ensure a comprehensive coverage of the topic in a fluent and compelling style.

Further material is available on www.wiley-vch.de/home/systemsbiology

Systems Biology A Textbook

Authors:

EDDA KLIPP

WOLFRAM LIEBERMEISTER

CHRISTOPH WIERLING

HANS LEHRACH

RALF HERWIG

all of MPI for Molecular Genetics, Berlin,
Germany

AXEL KOWALD

University of Bochum, Germany

Paperback · 592 pages

June 2009

ISBN: 978-3-527-31874-2

~~£60.00 / \$100.00 / €69.00~~

£48.00 / \$80.00 / €55.20*

And don't worry

Individualism will never die ...



Thank you !!!





Model documentation in UNICELLSYS

Model source

Unique model identifier containing short pathway identifier and version number.

Model author(s) with contact information (via vcard annotation).

Date of creation, date of last modification (or even the modification history).

A reference to a publication or report containing further information on the model.

Statement of terms of distribution (to be distributed only within the consortium??).

Model

Model assumptions, boundaries, and simplifications

Verbal description and phenomena covered

Simulation results

Point to UNICELLSYS experimental data sets used for model calibration and testing.

Taxonomy information; *S. cerevisiae* has NCBI taxonomy identifier 4932



Model element annotations in UNICELLSYS

Mandatory

General: ODE framework, SBML file format, global model information and documentation (MIRIAM!)

Model: complete and syntactically correct

Compartments: Naming scheme and GO annotations

Species (proteins, genes, mRNA, metabolites): Naming and annotations (SGD and ChEBI)

Post-translational protein modifications: Naming scheme and annotation

Units and unit prefixes: use of SBML guidelines

Recommended

Parameters: Coarse- or fine-grained annotation with SBO terms

Reactions: Annotation with SBO terms

Species: Use only amounts instead of concentrations (easier modeling of volume changes)

'Global module' model

including some variables with initial values