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
Big project of the past
“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 Novere et al, (2005)
Model publishing: the MIRIAM rules

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

Nicoletta Leonardi1, Andrew Finney2, Michael Hautha3, Upinder S Bhalla4, Fabien Campagne5, Julio Collado-Vides6, Edmund J Cruickshank7, Matt Halstead7, Idda Klipp8, Pedro Mendes9, Poul Nielsen9, Herbert Saum10, Bruce Shapiro11, Jacqui L. Speechley12, Hugh D. Spence12 & Barry J. Wain12

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}

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

http://mibbi.org/

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.

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

Chris P. Taylor1,2,3, Dawn Field1,2, Susanna Assunção Santos1,2,3, Jan Arts1,2, Be0 Appel1,2, Michael Ashburner1,2, Catherine A. Ball1,2, Pierre-Alain Beur1,2, Molly Bogue1,2, Tim Boeck1,2, Alvis Brannas3, Ryan E. Britton3,4, Adam Michael Clark1,5, Eric W Deutsch1,2, Oliver Flen1,2, Jennifer Fogel1,2, Peter Glazoul1,2, Frank Gibson1,6, Tanya Grant1,2, Graeme Grimes1,2, John M. Hanks7,2, Nigel W. Hardy4,8, Henning Hermjakob9,10, Randall K. Julian Je11, Matthew Kane11,12, Carsten Kettenring13, Christopher Kirschner14, Eugene Kolker13,14, Mark Kulp3,15, Martin Kulp3,15, Nicoletta Leonardi1,2, Jim Leebens-Mack9,4, Susanna E. Lewis1,2,5, Phillip Lord1,2, Ann Marie Mallon16, Nils-Niklas Marhandan17,4, Hiroaki Masuyama18,5, Ruth McNally5,19, Alexander Mehlke20,19, Norman Morrissey21,19, Sandra Orchard1, Johan Quakenbush1,5, James M. Reece1,23, Donald G. Robertson1,24, Philippe Rocca-Serra1,4, Henno Rodenkirchen1,2, Heiko Rosenfelder1,2, Javier Santoyo-Alvarez25, Richard H. Schumacher26, Daniel Schober2, Barry Smith5,27, Jason Snape27, Christian Stoeckert Jr28, Keith Tipton28, Peter Stark28, Andreas Untergasser28, Jo Vanderboom29 & Stefan Wiemann30

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.
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

http://sbgn.org/
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>
      ...
      ...
      ...
    </listOfSpecies>
    <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>
  </model>
</sbml>

SBML main site  http://sbml.org/

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/
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/
Annotating the model elements
Biological annotations in SBML

Species called “enzyme_R00001” represents an enzyme

Species called “ATP” represents KEGG C06262 (ATP)

Species called “ATP”

Species "ATP"

Species "enzyme_R00001"
SemanticSBML: Annotation and merging of SBML models

Stand-alone version

Online version

www.semanticsbml.org

... included in SSBML
Own efforts
Data conversion in model building: table formats

Stoichiometric model (table format)

Omics data (table format)

Kinetic parameters (table format)

SBML model

Requirements for data table format
- Completeness
- Annotations
- Defined structure
- Compatibility
- Intuitive format
SBtab: our proposal for standard table formats

<table>
<thead>
<tr>
<th>Reaction*</th>
<th>ReactionFormula</th>
<th>ID:kegg.reaction</th>
<th>GeneName</th>
<th>EnzymeRegulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>ATP + F6P ⇄ ADP + F16P</td>
<td>R00658</td>
<td>pfk</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>F16P + H2O ⇄ F6P + Pi</td>
<td>R01015</td>
<td>fbp</td>
<td>+ PEP – AMP</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound*</th>
<th>Name</th>
<th>ID:kegg.compound</th>
</tr>
</thead>
<tbody>
<tr>
<td>F6P</td>
<td>Fructose-6-phosphate</td>
<td>C05345</td>
</tr>
<tr>
<td>ATP</td>
<td>ATP</td>
<td>C00002</td>
</tr>
<tr>
<td>ADP</td>
<td>ADP</td>
<td>C00008</td>
</tr>
<tr>
<td>F16P</td>
<td>Fructose-1,6-bisphosphate</td>
<td>C00354</td>
</tr>
<tr>
<td>H2O</td>
<td>Water</td>
<td>C00001</td>
</tr>
<tr>
<td>Pi</td>
<td>Inorganic phosphate</td>
<td>C00009</td>
</tr>
<tr>
<td>PEP</td>
<td>Phosphoenolpyruvate</td>
<td>C00074</td>
</tr>
<tr>
<td>AMP</td>
<td>AMP</td>
<td>C00020</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound*</th>
<th>CompoundID:obo.chebi</th>
<th>'s1 Mean'</th>
<th>'s1 Std'</th>
<th>'s2 Mean'</th>
<th>'s2 Std'</th>
</tr>
</thead>
<tbody>
<tr>
<td>4abut</td>
<td>CHEBI:16865</td>
<td>27.5</td>
<td>0.1</td>
<td>28.2</td>
<td>0.1</td>
</tr>
<tr>
<td>fum</td>
<td>CHEBI:18012</td>
<td>0.13</td>
<td>0.1</td>
<td>0.15</td>
<td>0.1</td>
</tr>
<tr>
<td>succ</td>
<td>CHEBI:15741</td>
<td>0.17</td>
<td>0.1</td>
<td>0.19</td>
<td>0.1</td>
</tr>
</tbody>
</table>
SBtab: our proposal for standard table formats

### Defined column names
- **Reaction**: ATP + F6P <=> ADP + F16P
- **GeneName**: pfk
- **EnzymeRegulation**: + PEP – AMP

### Defined syntax for reactions and regulation
- **ID**: kegg.reaction
- **GeneName**: pfk

### MIRIAM-compliant annotations
- **Compound**
  - F6P: Fructose-6-phosphate
  - ATP: ATP
  - ADP: ADP
  - F16P: Fructose-1,6-bisphosphate
  - H2O: Water
  - Pi: Inorganic phosphate
  - PEP: Phosphoenolpyruvate
  - AMP: AMP

### Similar to existing formats
- **Structured**: easy to parse

### 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**  
\[ A + B \rightleftharpoons C \]

<table>
<thead>
<tr>
<th>Method</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mass-action</strong></td>
<td>[ v = k^+ a b - k^- c ]</td>
</tr>
<tr>
<td><strong>Linlog</strong></td>
<td>[ \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) ]</td>
</tr>
<tr>
<td><strong>TKM</strong></td>
<td>[ v = R^{-1} (\xi_A \xi_B - \xi_C) \quad \xi_i = e^{\mu_i/RT} ]</td>
</tr>
<tr>
<td><strong>Convenience kinetics</strong></td>
<td>[ 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} ]</td>
</tr>
</tbody>
</table>

**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

4 types of regulation
- Inhibition (non-competitive)
- Inhibition (competitive)
- Activation (essential)
- Activation (non-essential)

Cooperativity (exponents)
- Thermodynamically correct formulas with Hill-like exponents

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
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
And don't worry.

Individualism will never die ...

Thank you !!!

Heike Bollig, “Produktionsfehler”
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