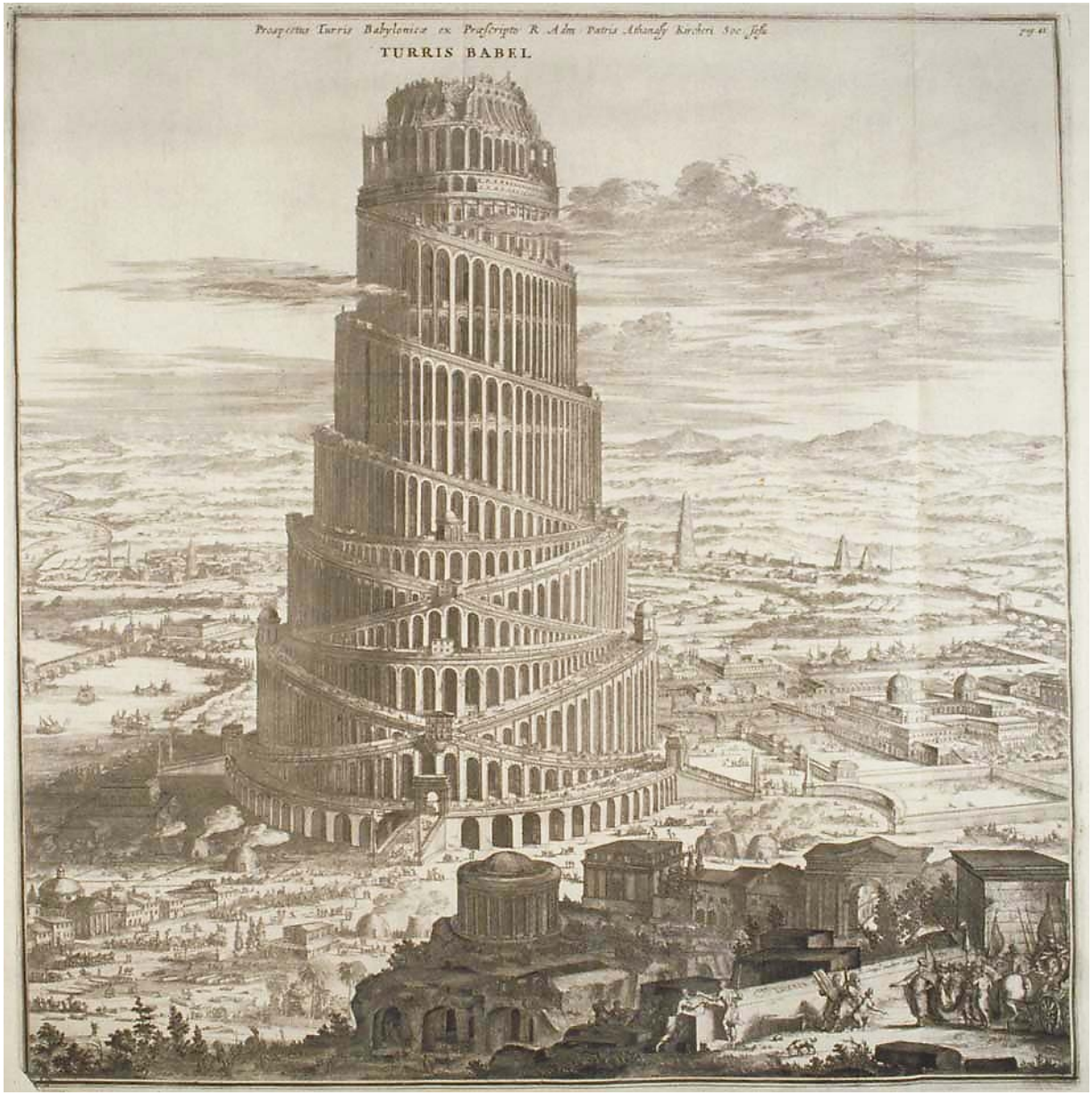
Processing of SBML models based on semantic annotations

Wolfram Liebermeister

Managing and curating Biological Pathways

Humboldt University Berlin, January 8, 2010

The mother of all failed projects ...



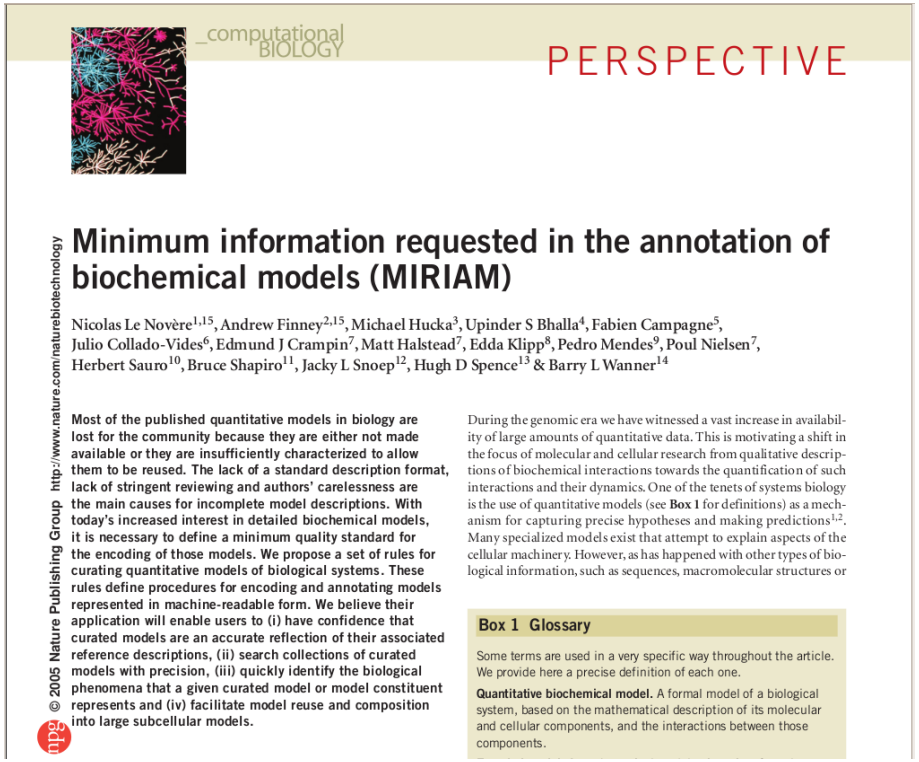
… and one of its children

Mars climate orbiter

How to waste time and money in Systems Biology

"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)



_computational BIOLOGY

PERSPECTIVE

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

Nicolas Le Novère[1,15], Andrew Finney[2,15], Michael Hucka[3], Upinder S Bhalla[4], Fabien Campagne[5], Julio Collado-Vides[6], Edmund J Crampin[7], Matt Halstead[7], Edda Klipp[8], Pedro Mendes[9], Poul Nielsen[7], Herbert Sauro[10], Bruce Shapiro[11], Jacky L Snoep[12], Hugh D Spence[13] & Barry L Wanner[14]

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

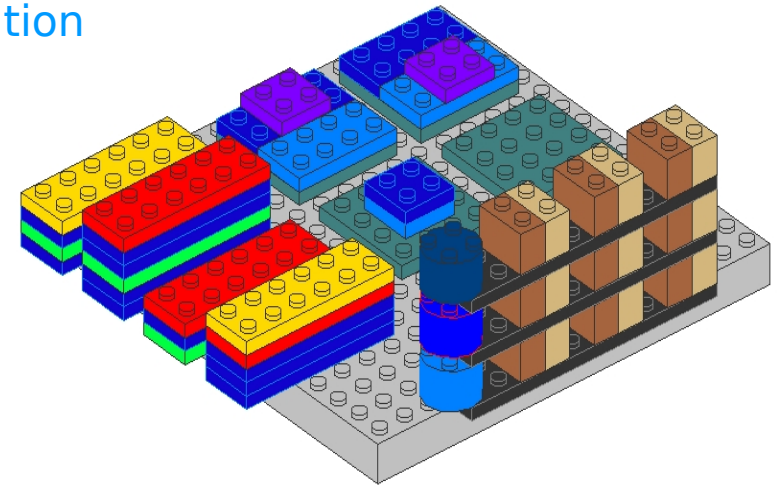
Box 1 Glossary

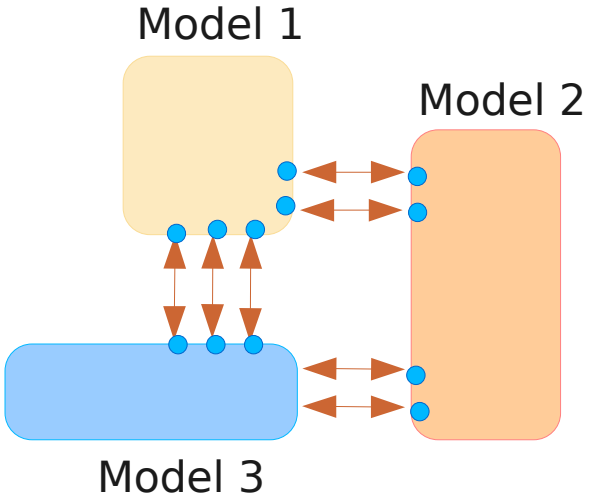
Some terms are used in a very specific way throughout the article. We provide here a precise definition of each one.

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.

Playing LEGO with biochemical models ?

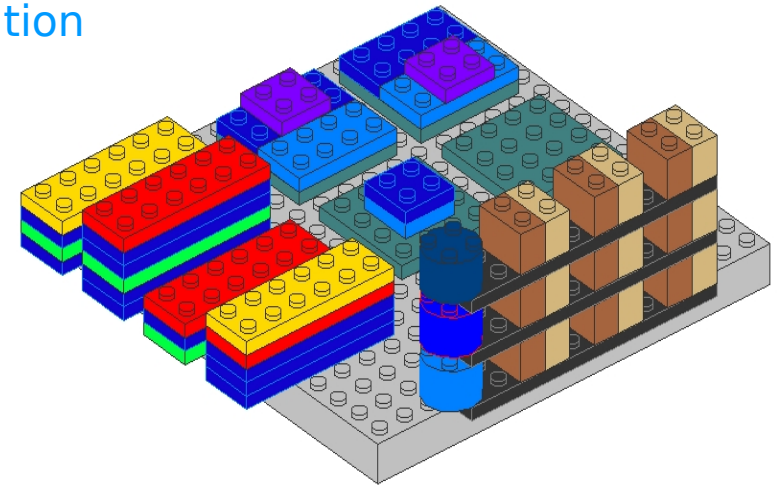
Model composition

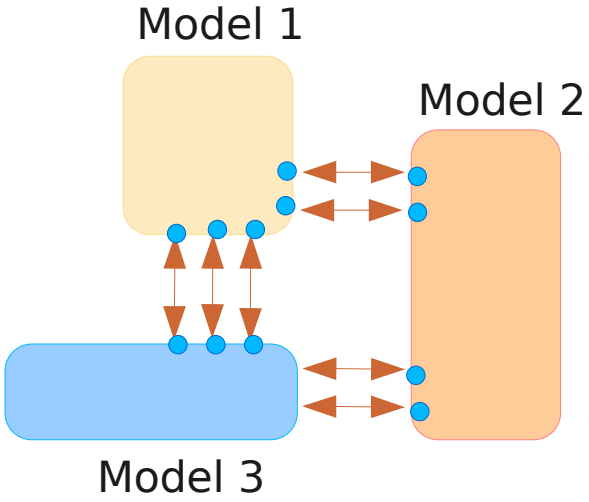




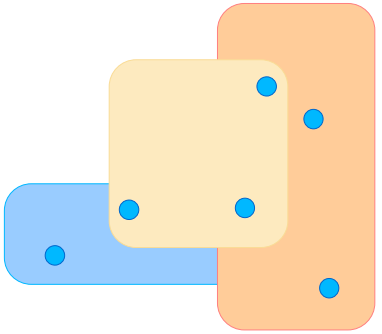
Playing LEGO with biochemical models ?

Model composition





Model merging

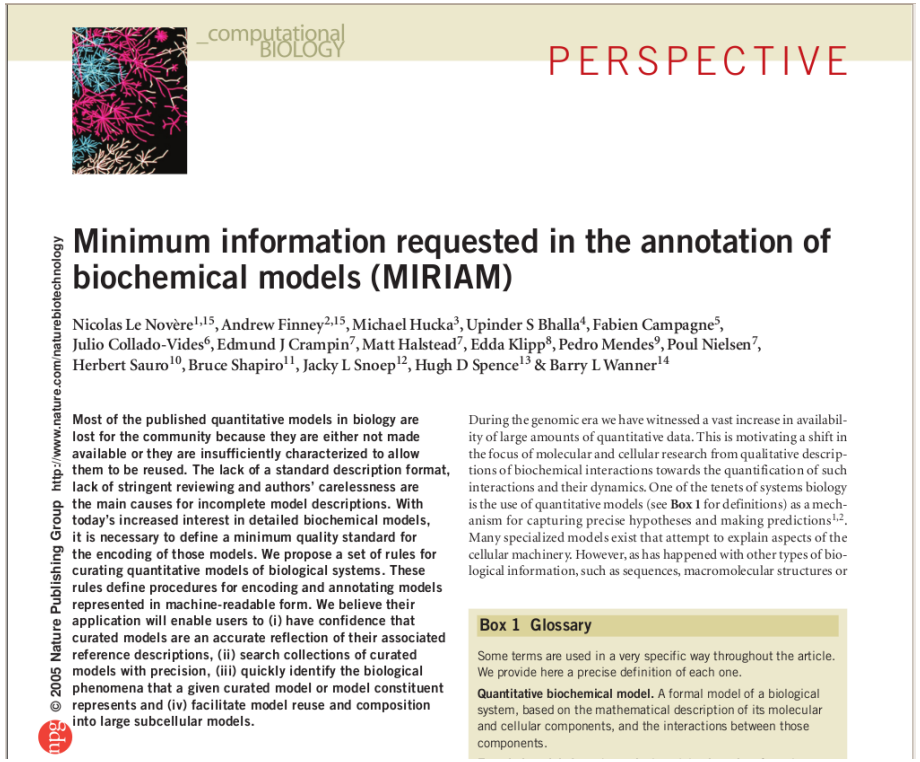


Overview

- MIRIAM guidelines and SBML format for systems biology models
- Semantic annotations in SBML
- Model building and standard rate laws
- Open questions

MIRIAM and the SBML format

The MIRIAM rules for model publishing



_computational BIOLOGY

PERSPECTIVE

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

Nicolas Le Novère[1,15], Andrew Finney[2,15], Michael Hucka[3], Upinder S Bhalla[4], Fabien Campagne[5], Julio Collado-Vides[6], Edmund J Crampin[7], Matt Halstead[7], Edda Klipp[8], Pedro Mendes[9], Poul Nielsen[7], Herbert Sauro[10], Bruce Shapiro[11], Jacky L Snoep[12], Hugh D Spence[13] & Barry L Wanner[14]

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.

_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 Aerts[4], Rolf Apweiler[1], Michael Ashburner[5], Catherine A Ball[6], Pierre-Alain Binz[7,8], Molly Bogue[9], Tim Booth[2], Alvis Brazma[1], Ryan R Brinkman[10], Adam Michael Clark[11], Eric W Deutsch[12], Oliver Fiehn[13], Jennifer Fostel[14], Peter Ghazal[15], Frank Gibson[16], Tanya Gray[2,3], Graeme Grimes[15], John M Hancock[17], Nigel W Hardy[18], Henning Hermjakob[1], Randall K Julian Jr[19], Matthew Kane[20], Carsten Kettner[21], Christopher Kinsinger[22], Eugene Kolker[23,24], Martin Kuiper[25], Nicolas Le Novère[1], Jim Leebens-Mack[26], Suzanna E Lewis[27], Phillip Lord[16], Ann-Marie Mallon[17], Nishanth Marthandan[28], Hiroshi Masuya[29], Ruth McNally[30], Alexander Mehrle[31], Norman Morrison[2,32], Sandra Orchard[1], John Quackenbush[33], James M Reecy[34], Donald G Robertson[35], Philippe Rocca-Serra[1,36], Henry Rodriguez[22], Heiko Rosenfelder[31], Javier Santoyo-Lopez[15], Richard H Scheuermann[28], Daniel Schober[1], Barry Smith[37], Jason Snape[38], Christian J Stoeckert Jr[39], Keith Tipton[40], Peter Sterk[1], Andreas Untergasser[41], Jo Vandesompele[42] & Stefan Wiemann[31]

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 communities[8,9]; 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

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}

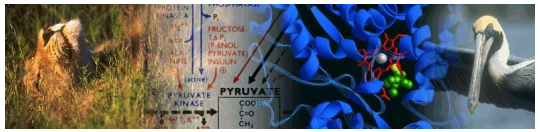
SBML, the Systems Biology Markup Language

One exchange format - about 180 tools
that understand each other

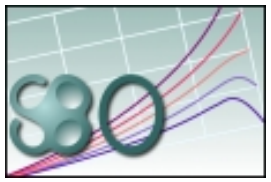
http://sbml.org/

BIOMODELS.NET

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/

```
<?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>
```

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

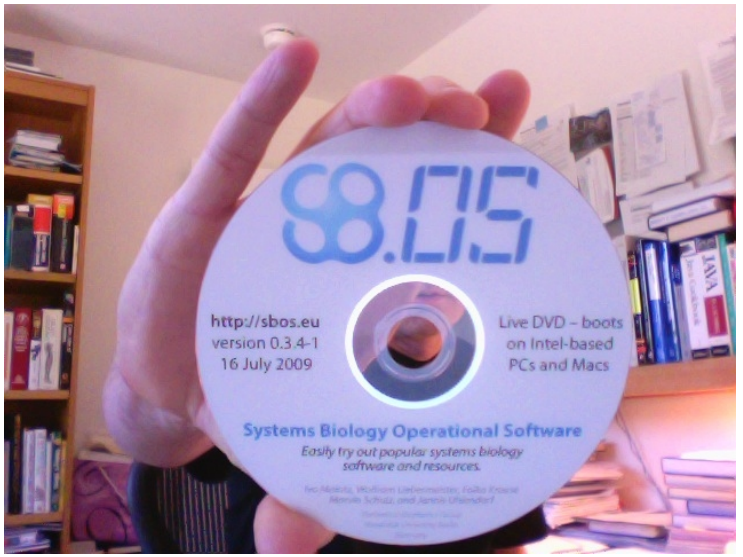
Allows for

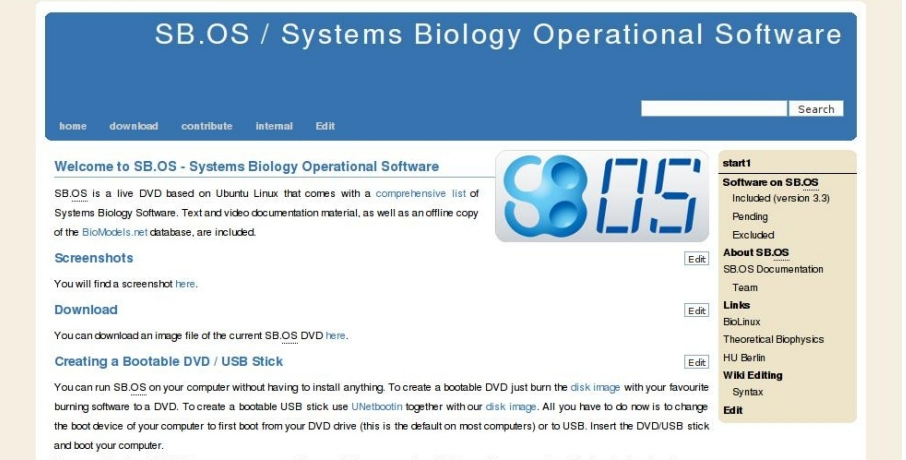
model building, layout, simulation, fitting, annotation, merging

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/

Annotations and semanticSBML

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)

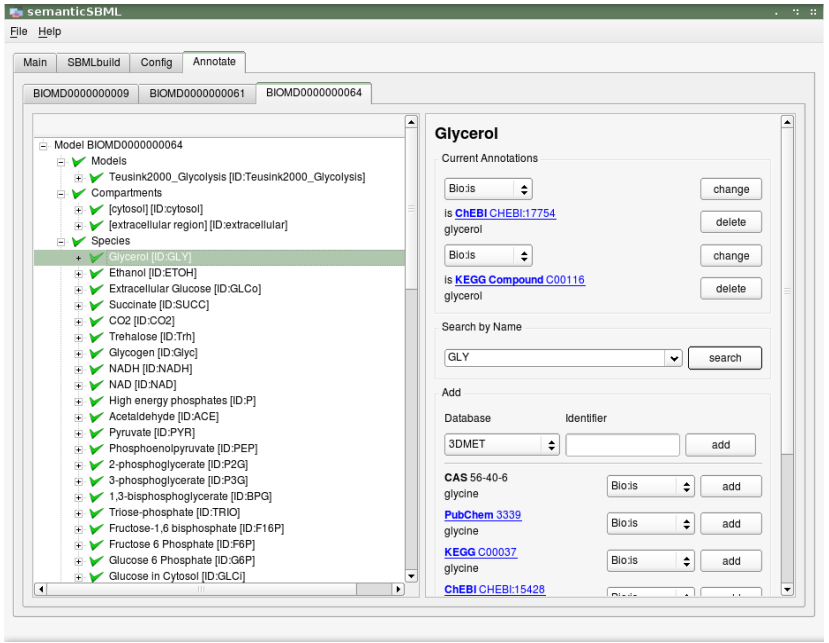
```
<species metaid=".." id="ATP" name="ATP concentration" compartment="cytosol">
 <annotation>
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
           xmlns:bqbiol=”http://biomodels.net/biology-qualifiers/"
           xmlns:bqmodel="http://biomodels.net/model-qualifiers/">
   <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 Web resource Identifier

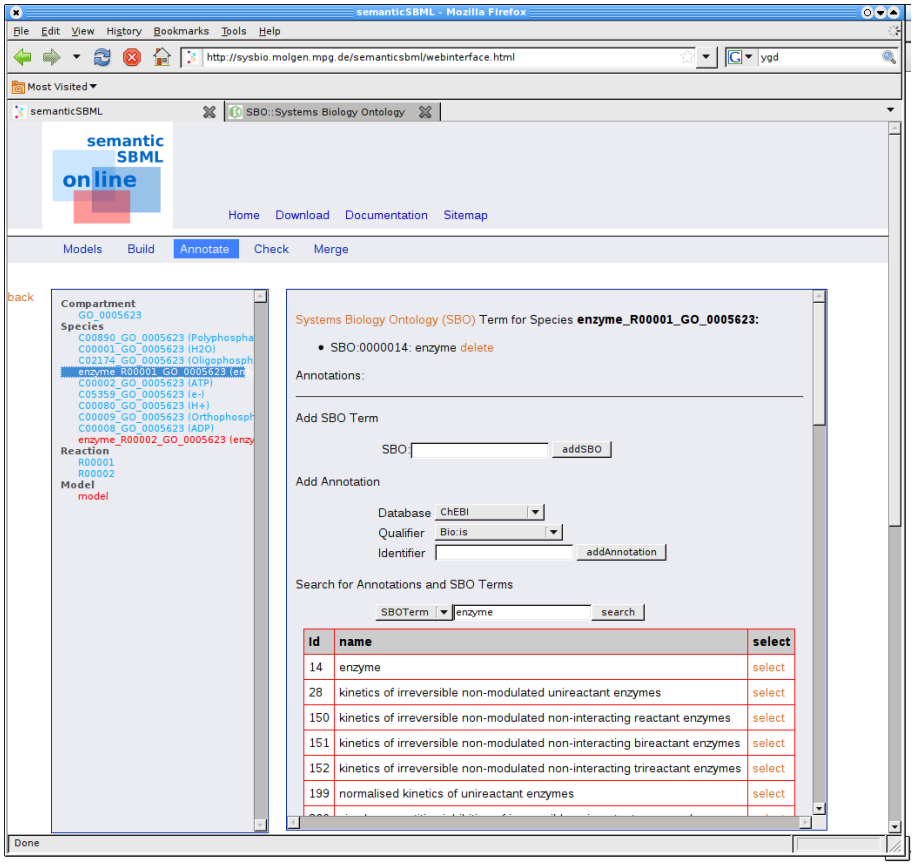
"A simple scheme for annotating SBML with references to controlled vocabularies and database entries" Le Novere and Finney, 2005

SemanticSBML: Annotation and merging of SBML models

Stand-alone version



Online version



semantic SBML www.semanticsbml.org

... included in

Annotating Source-Code Models

Python, C/C++, Matlab, R, ...

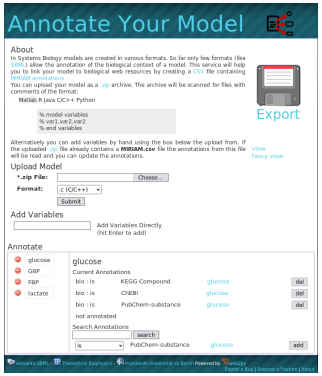
www.semanticsbml.org/aym/

1

```
function [glucose_EE,G6P_EE,FBP_EE,trioseP_EE,BPG_EE,PEP_EE,pyruvate_EE,l
% model variables
% glucose,G6P,FBP,trioseP
% end variables

% ~~ Initial conditions ~~
ini_cond        = zeros (1,13);
ini_cond (1)    = 25;   % glucose
ini_cond (2)    = 0;    % G6P/F6P pool

% Variable
% -- Enzyme concentrations
E_pts               = parpick (0.027,vector (1));
E_pfk               = parpick (0.027,vector (2));

% -- kcats
kcat_pts            = parpick (69.2593,vector (12));
kcat_pfk            = parpick (171.481,vector (13));

options = odeset ('RelTol',1e-8,'AbsTol',1e-8, 'InitialStep',0.01);
[T,Y] = ode15s (@glycolysis_eq,0:5000,ini_cond,options,param_vector);
glucose_EE  = Y (end,1);
G6P_EE      = Y (end,2);
FBP_EE      = Y (end,3);
trioseP_EE  = Y (end,4);
BPG_EE      = Y (end,5);
PEP_EE      = Y (end,6);
pyruvate_EE = Y (end,7);
lactate_EE  = Y (end,8);
mixedac_EE  = Y (end,9);
ATP_EE      = Y (end,10);
```

Raw Source-Code

2



AYM Web Service

3

```
function v = CK_ldh (E,kcat,pyr,                    NAD,          ,Pi,Km_pyr,Km_NADH,  Km_l
                     NADH
                     is ChEBI NADH
                     is PubChem-substance NADH
v = E.*((          /Pi)./((          /Pi) + Km_FBPPi)).*(kcat.*(pyr./Km_pyr).*(          ./K

function v = CK_mixed _acid _lump (E,kcat,pyr,          ,          ,Pi,mixedacid,NAD,
v = E.*((Km_trioseP.^n_trioseP)./((Km_trioseP.^n_trioseP) + (trioseP.^n_trioseP)))...

.*(kcat.* ((pyr./Km_pyr).^2).*((          ./Km_NADH).^2).*(          ./Km_ADP).*(Pi./K

((1 + (pyr./Km_pyr) + (pyr./Km_pyr).^2 ).*( 1 + (          ./Km_NADH) + ((          ./
```

Annotated Source-Code

4

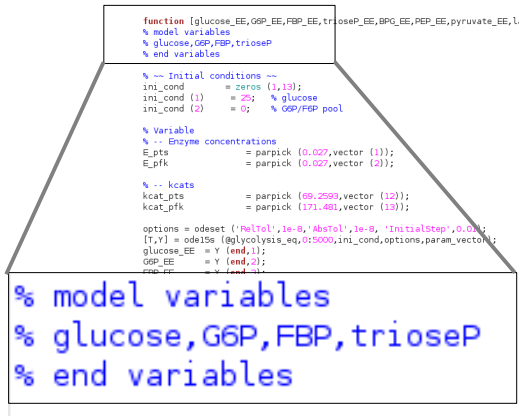
MIRIAM.csv

Annotating Source-Code Models

Python, C/C++, Matlab, R, …

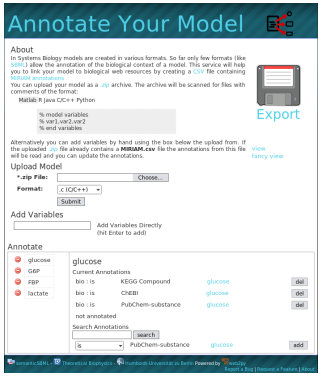
www.semanticsbml.org/aym/

1



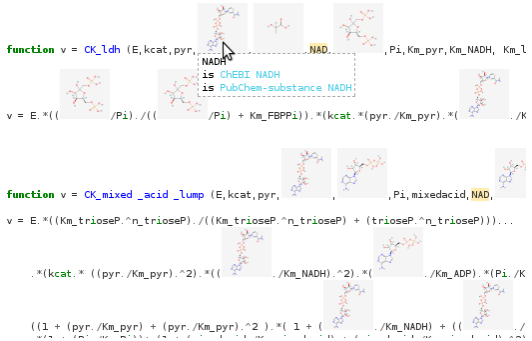
Raw Source-Code

2



AYM Web Service

3



Annotated Source-Code

4

MIRIAM.csv

Annotating Source-Code Models

Python, C/C++, Matlab, R, …

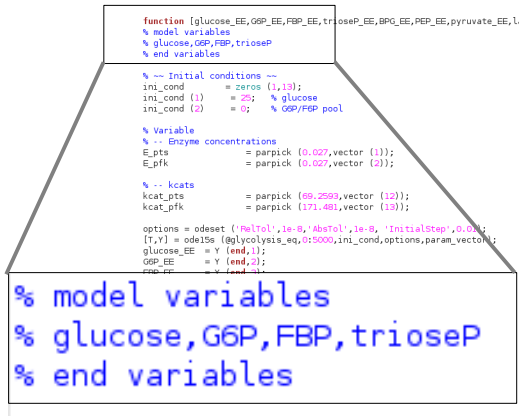
www.semanticsbml.org/aym/

1

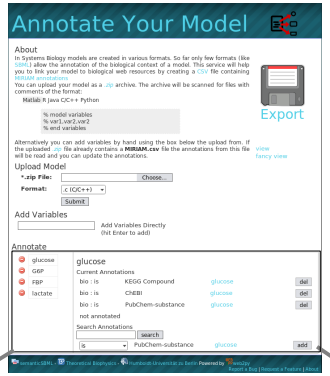
2

3

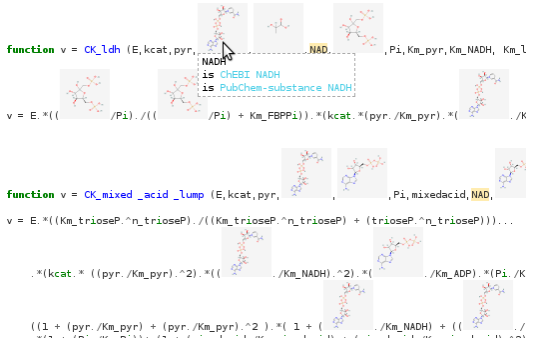
4



Raw Source-Code

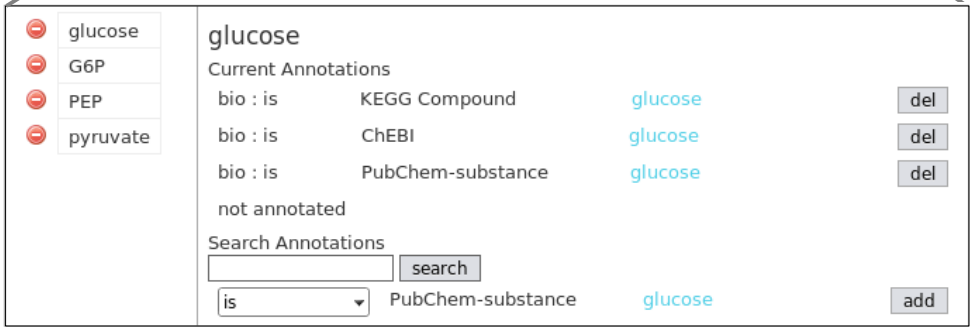


AYM Web Service



Annotated Source-Code

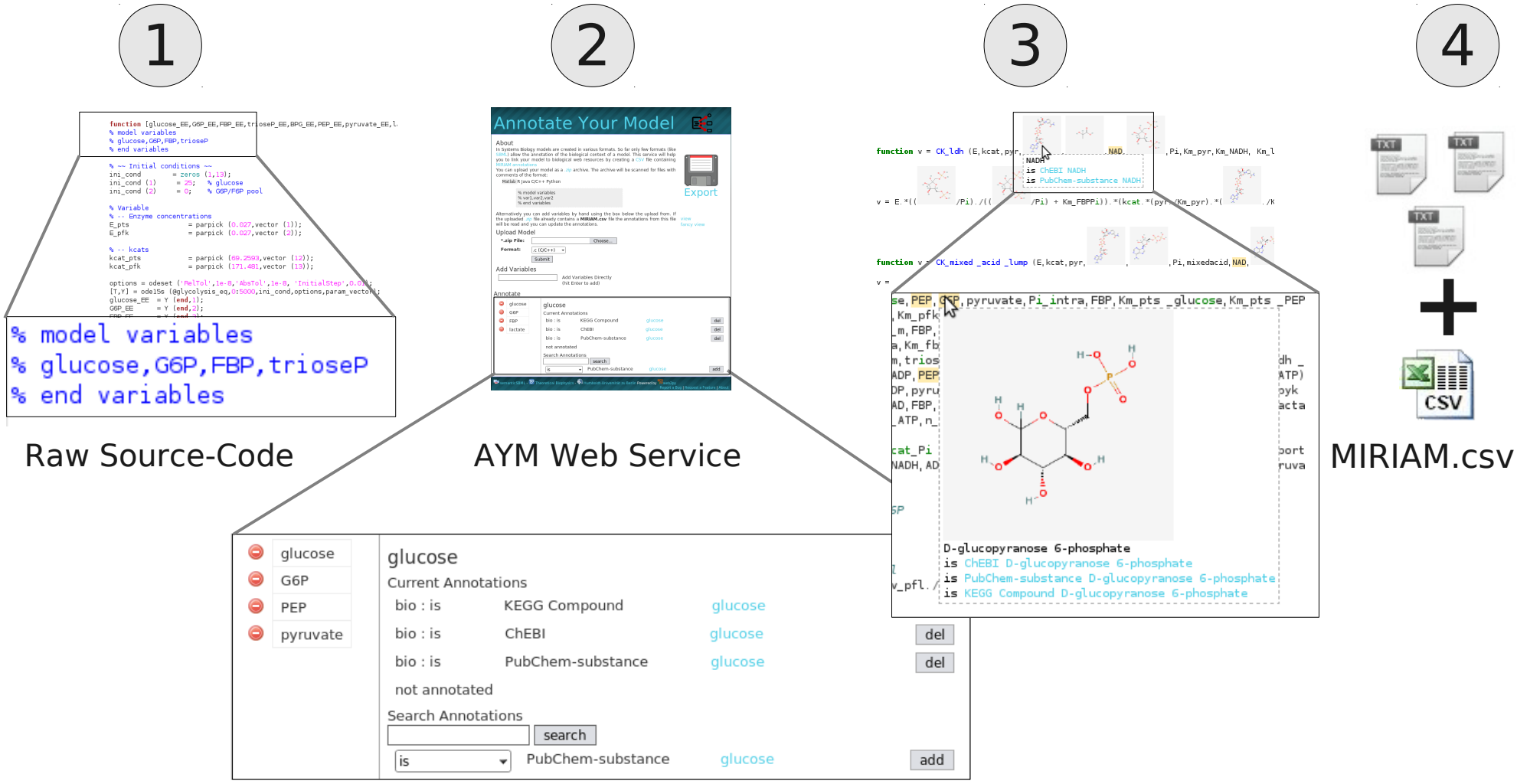
MIRIAM.csv



Annotating Source-Code Models

Python, C/C++, Matlab, R, …

www.semanticsbml.org/aym/



Annotating Source-Code Models

Python, C/C++, Matlab, R, …

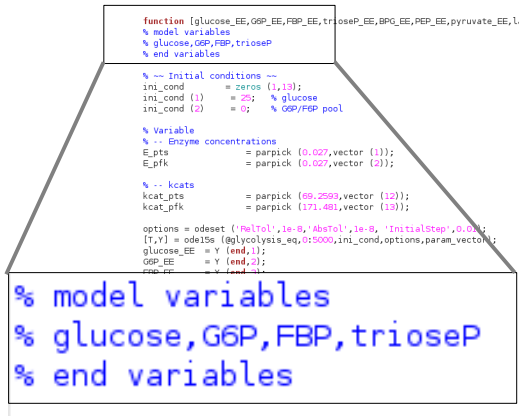
www.semanticsbml.org/aym/

1

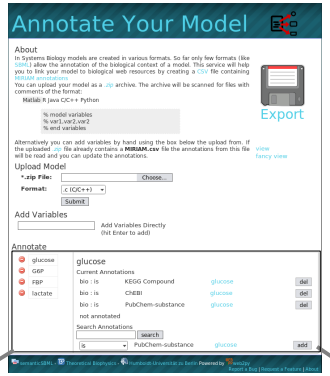
2

3

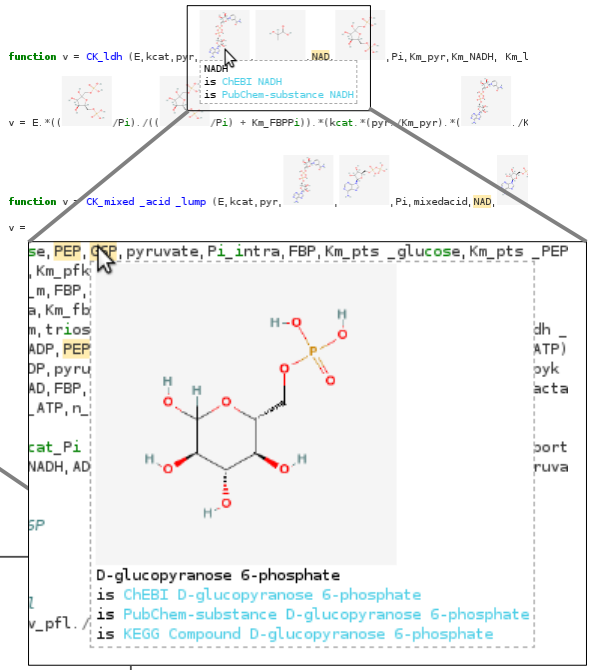
4

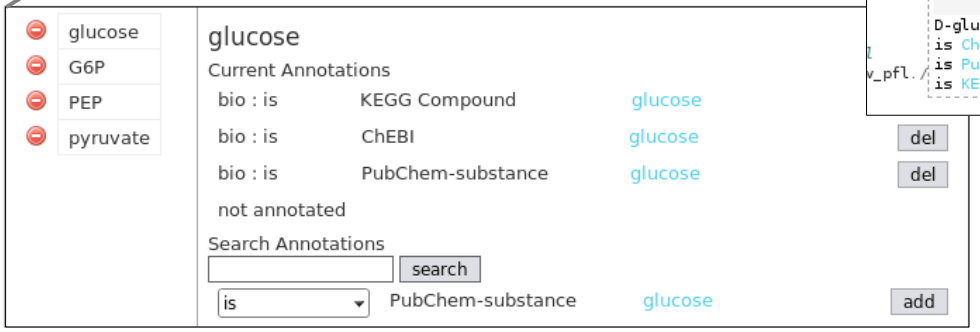


Raw Source-Code



AYM Web Service

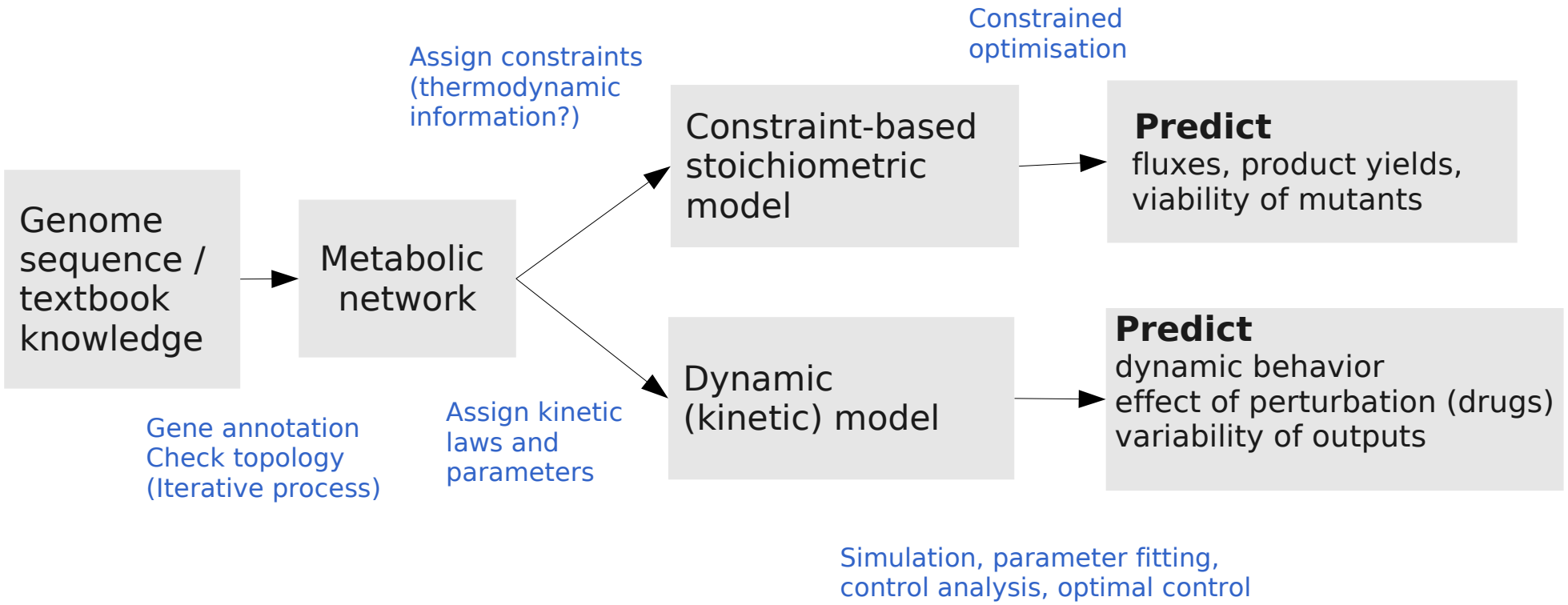




MIRIAM.csv

Creating kinetic models

A dream: from genomes to models ...



Creating SBML models with semanticSBML

Reaction list
(KEGG Ids and reaction formulas)

```
R00014
R00200
R00224
R00268
...
...
...
R00243
Glycerol[ext] <=> Glycerol
Ethanol[ext] <=> Ethanol
beta-D-Glucose[ext] <=> beta-D-Glucose
Acetate[ext] <=> Acetate
ADP + GTP <=> ATP + GDP
ADP + C00030 + 0.50 O2 <=> ATP + H2O + C00028
ADP + NADH + 0.50 O2 <=> ATP + NAD+ + H2O
ATP <=> ADP
0.03 Pyruvate + ... + 0.01 O2 <=> 0.81 ADP + ... + Biomass
```

semanticSBML

SBML with annotations

```
<?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>
```

Problems in building large kinetic models

1. Kinetic laws are often unknown

Use simple yet plausible standard rate laws

2. Models should obey the laws of thermodynamics

Be aware of relevant constraints
Use independent parameters in fitting, sampling, optimisation etc

3. Parameters show variation and may be uncertain

Describe parameters by probability distributions
Infer probabilistic statements about model outputs, dynamics etc

4. Data may not suffice to determine the parameters

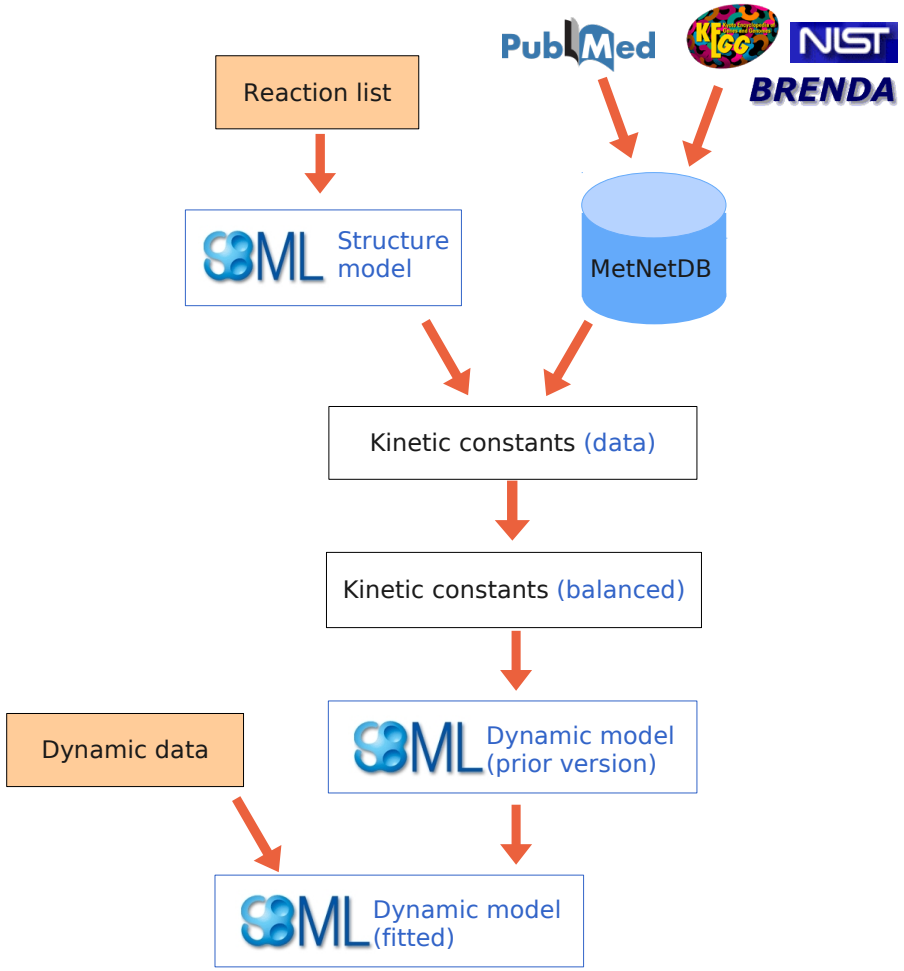
Use prior distributions and Bayesian statistics for estimation

5. Parameter estimation in large models is expensive

Use direct sampling methods that avoid steady-state calculation

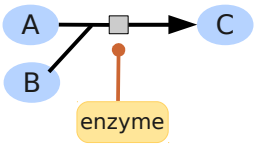
Collecting and combining kinetic data

Modelling workflow

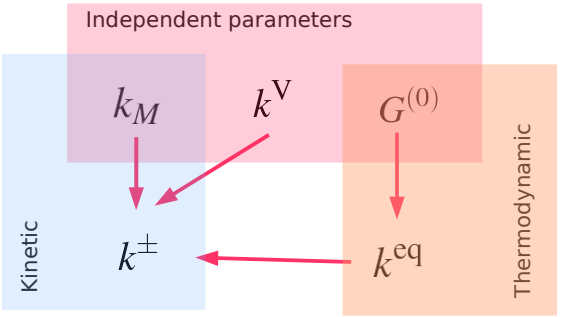


Convenience kinetics

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



Thermodynamic balancing



Modular rate laws – form of rate laws

3 Thermodynamic versions (numerator)

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

Apparent cooperativity

Thermodynamically correct formulas
with Hill-like exponents

Reaction $A + B \Leftrightarrow C$

$$v(a,b,c,u) = u\frac{k^{+}\left(a/k_{\mathrm{A}}^{\mathrm{M}}\right)\left(b/k_{\mathrm{B}}^{\mathrm{M}}\right) - k^{-}\left(c/k_{\mathrm{C}}^{\mathrm{M}}\right)}{\left(1+a/k_{\mathrm{A}}^{\mathrm{M}}\right)\left(1+b/k_{\mathrm{B}}^{\mathrm{M}}\right) + \left(1+c/k_{\mathrm{C}}^{\mathrm{M}}\right) - 1}$$

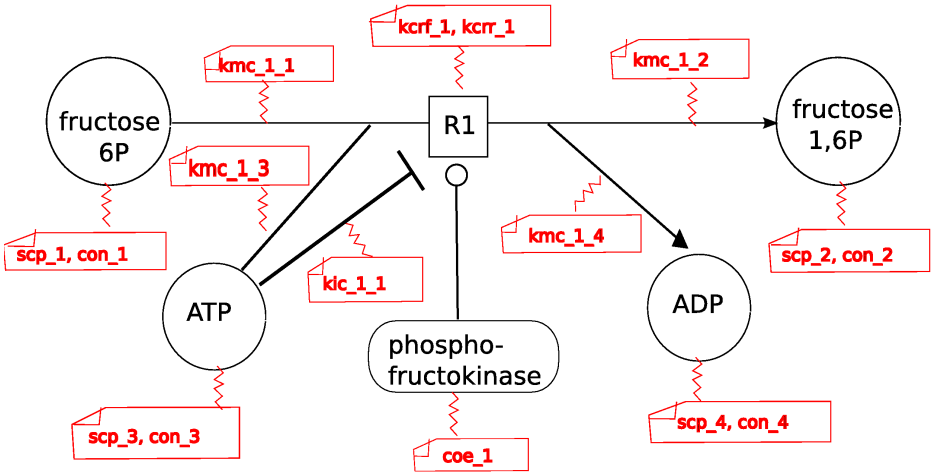
5 Types of regulation

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

5 Types of rate laws (denominator)

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

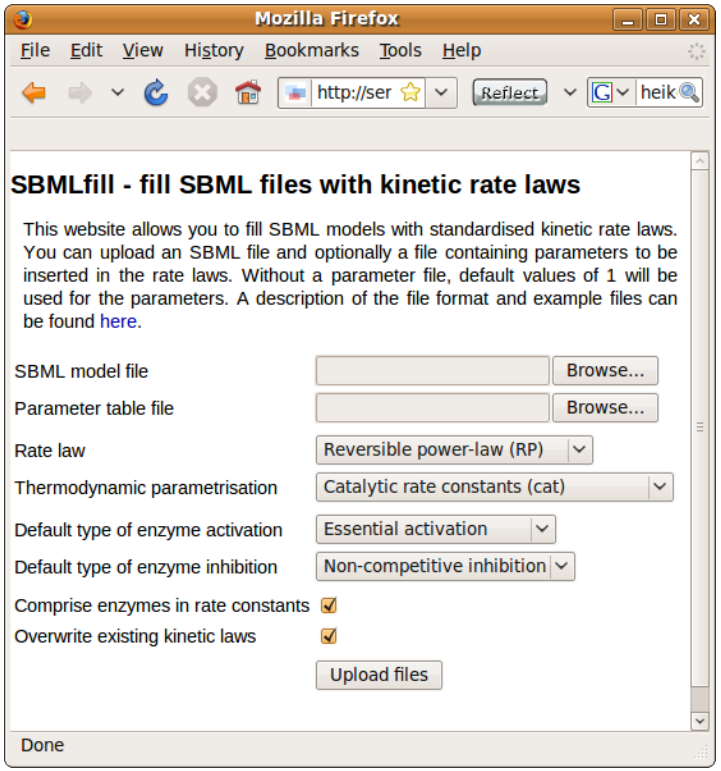
Modular rate laws – parameter tables



Parameter table in SBtab format

Quantity	QuantityType	Value	Unit	CompoundName	EnzymeName
scp 1	standard biochemical potential	-500	kJ/mol	fructose 6P	
scp 2	standard biochemical potential	-500	kJ/mol	fructose 1,6P	
scp 3	standard biochemical potential	-500	kJ/mol	ATP	
scp 4	standard biochemical potential	-500	kJ/mol	ADP	
hco 1	reaction cooperativity	1	1		posphofructokinase
kcrf 1	substrate catalytic rate constant	0.1	1/s		posphofructokinase
kcrr 1	product catalytic rate constant	0.1	1/s		posphofructokinase
Kmc 1 1	Michaelis constant	1	mM	fructose 6P	posphofructokinase
Kmc 1 2	Michaelis constant	1	mM	fructose 1,6P	posphofructokinase
Kmc 1 3	Michaelis constant	1	mM	ATP	posphofructokinase
Kmc 1 4	Michaelis constant	1	mM	ADP	posphofructokinase
kic 1 5	inhibitory constant	1	mM	ATP	posphofructokinase
win 1 5	inhibition cooperativity	1	1	ATP	posphofructokinase
con 1 1	concentration	1	mM	fructose 6P	
con 1 2	concentration	1	mM	fructose 1,6P	
con 1 3	concentration	1	mM	ATP	
con 1 4	concentration	1	mM	ADP	
coe 1	enzyme concentration	0.001	mM		posphofructokinase

Modular rate laws - implementation



Web interface for SBML models
http://semanticsbml.org/sbmlfill

SBMLsqueezer

Plug-in for CellDesigner by A. Dräger, Uni Tübingen

SemanticSBML: summary of features

- Match and translate of database identifiers, names, relations
- Add and edit semantic annotations in SBML models
- Annotate non-SBML models
- *Check model for semantics-related quality criteria
- Merge SBML models; detect possible inconsistencies
- Similarity/distance scores between annotations, model elements, models
- *Clustering of SBML models
- Build models from reaction sum formulas / database IDs
- Insert kinetic rate laws (with parameters)
- *Recognise kinetic rate laws by their formula
- *Determine consistent parameter sets from kinetic data

Open questions / future directions / collaborations ?

- Wider usage of SBtab format
- Parametrisation of large-scale kinetic models with given flux patterns
- Combination of constraint-based and kinetic models
- Annotating the history of a model
- Further validity checks on models
- Computer-assisted model extension (with models from the public domain)
- Logging and replay of model processing
- Combinatorial model variants and optimisation
- Mapping of quantitative data (concentrations, fluxes, expression)
- Replacing submodels by effective reactions / reaction modules

Advertisement: text book “Systems Biology”

Acknowledgements

Humboldt University Berlin, Theoretical Biophysics

Edda Klipp
Jannis Uhlendorf
Falko Krause
Timo Lubitz
Marvin Schulz
Dirk Wiesental
Ivo Mainz
...

!!! The SBML community !!!

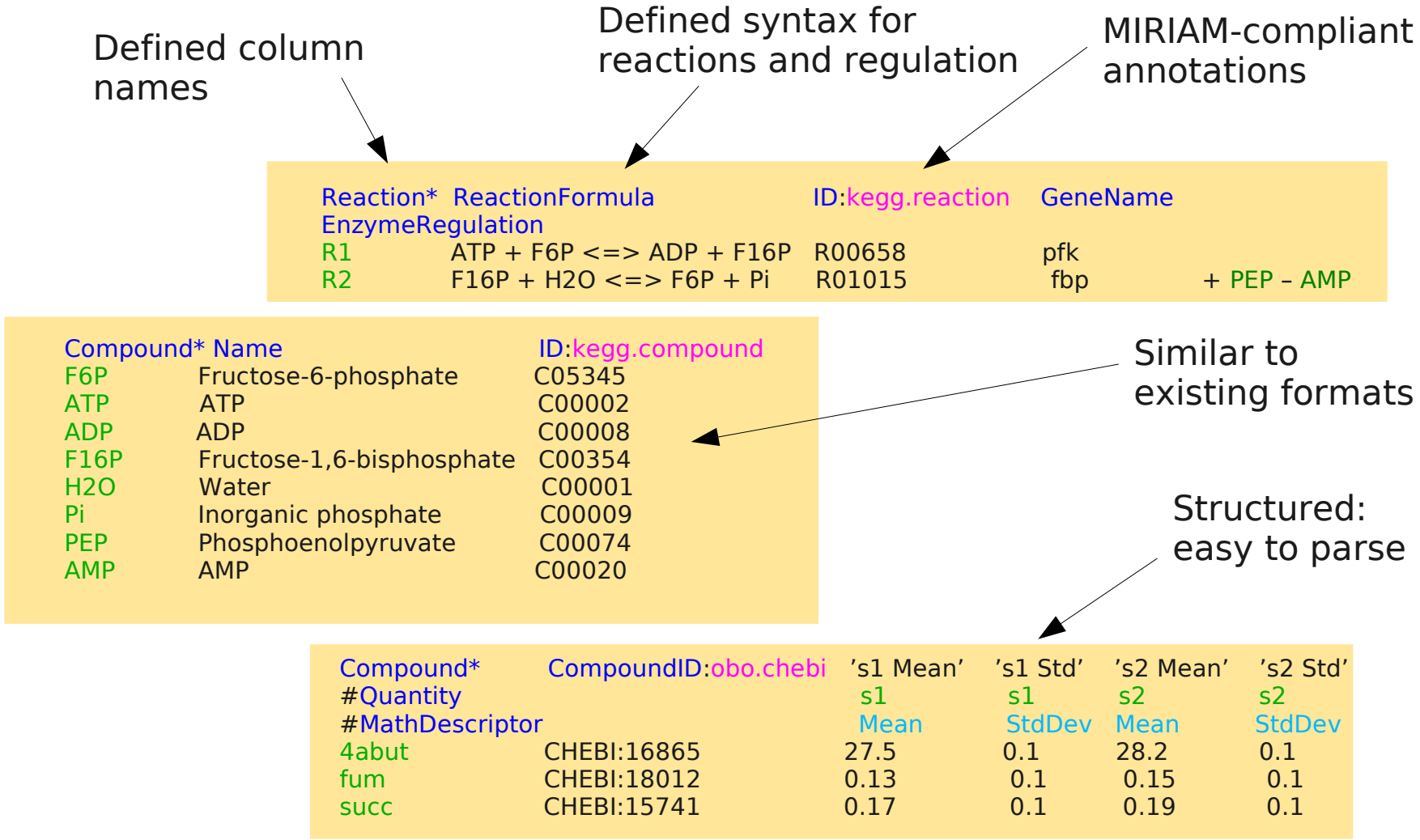
Financial support by BaSysBio

And don't worry

Individualism will never die ...

Thank you !!!

SBtab: a proposal for standard table formats



- Online validator is still under construction ...
- We are still working on the format, suggestions are appreciated!!