

Standard Formats in Computational Modeling

Matthias König

THIS WEEK

EDITORIALS

New agreement to tackle pharmaceutical pollution [p184](#)Vaccination: the best way to measure health care [p185](#)

How to save

Let's think about cognitive bias

The human brain's habit of finding what it wants to find is a key problem for research. Establishing robust methods to avoid such bias will make results more reproducible.

“Ever since I first learned about confirmation bias I've been... Some researchers already do this well, so one relatively simple strategy is to encourage them to do more.”

Reproducibility: Seek out stronger science

• Monya Baker

Nature 537, 703–704

An open mind on open data

The move to make scientific findings transparent can be a major boon to research, but it can be tricky to embrace the change.

Power failure: why small sample size undermines the reliability of neuroscience

Katherine S. Button^{1,2}, John D. A.

Repetitive flaws

Strict guidelines to improve the reproducibility of experiments are a welcome move.

From next week, scientists who submit grant applications to National Institutes of Health (NIH) will be asked to take more care. As part of an increasing drive to boost the reliability of research, the NIH will require applicants to explain the science behind their proposals and defend the quality of their research designs. They must also account for biological variability, by including both male and female mice in planned experiments. This will authenticate experimental material at all times and artifacts.

These demands are timely, sensible and, if researchers have followed the advice of their scientific societies, will sound familiar. Over the past year, a string of organizations have published the statements and guidelines to boost the reproducibility of research.

Collectively, the message is: show your work, don't fool yourself with unreliable results or obfuscating data. Unstated guidelines from the Federation of American Societies for Experimental Biology, for example, go further. The organization's code of ethics states that “it is not studies infrastructure. Society has done And the Ameri-

Fewer numbers, better science

Scientific quality is hard to define, and numbers are easy to look at. But bibliometrics are warping science – encouraging quantity over quality. Leaders at two research institutions describe how they do things differently.



Acknowledging and Overcoming Nonreplicability in Basic and Preclinical Research

ence for nonreplicability in basic and pre-

original study vs 0.71 (95% CI,

COMMENT

nature

[Home](#) | [News & Comment](#) | [Research](#) | [Careers & Jobs](#) | [Current Issue](#) | [Archive](#) | [Audio & Video](#)
[News & Comment](#) > [News](#) > [2016](#) > [May](#) > Article

NATURE | NEWS

Missing mice: gaps in data plague animal research

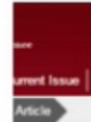
Reports of hundreds of biomedical experiments lack essential information.

Hide results to seek the truth

More fields should, like particle physics, adopt blind analysis to thwart bias, urge Robert MacCann and Saul Perlmutter.

Believe it or not: how much do we rely on published data on protein?

Low statistical power in biomedical science: a review of three human research domains



ROYAL SOCIETY OPEN SCIENCE

[rsos.royalsocietypublishing.org](#)

Research



Scientific method: Statistical errors

P values, the ‘gold standard’ of statistical validity, are not as reliable as we assume.

Confidence in preclinical research

For decades, model organisms have provided an important reductionist approach for understanding

making strides in their efforts to understand and for the complexity of the microbiome in rodent

Perrin S (2014) *Nature* 407:423

Results of rigorous animal tests by the Amyotrophic Lateral Sclerosis Therapy Development Institute (ALS) are less promising than those published. All these compounds have disappointed in human testing,

Raise standards for preclinical cancer research

THE LANCET

“85% of health research is wasted.”

DUE DILIGENCE, OVERDUE

Results of rigorous animal tests by the Amyotrophic Lateral Sclerosis Therapy Development Institute (ALS) are less promising than those published. All these compounds have disappointed in human testing,

The Breakdown in Biomedical Research

Contaminated samples, faulty studies and have created a crisis in laboratories and in quest for new treatments and cures



the Atlantic

Lies, Damned Lies, and Medical Science

MUCH OF WHAT MEDICAL RESEARCHERS CONCLUDE IN THEIR STUDIES IS MISLEADING,¹ FLAT-OUT WRONG. SO WHY ARE DOCTORS—TO A STRIKING EXTENT—STILL DRAWING US

Editorial: “Is science in big trouble?”

The
Economist

OCTOBER 2016 | \$4.99 | economist.com

Washington's lawyer surplus
How to do a nuclear deal with Iran
Investment tips from Nobel economist
Junk bonds are back
The meaning of Sachin Tendulkar

SUBSCRIBE

THE NEW YORKER

IN THE MAGAZINE | BLOGS | AUDIO & VIDEO | LISTINGS

POLITICS | PROFILES | THE TALK OF THE TOWN | COMMENT | TABLE OF CONTENTS | THE FINANCIAL PAGE

THE NEW YORKER | PERMITTING & ESSAYS

ANNALS OF SCIENCE

THE TRUTH WEARS OFF

Is there something wrong with the scientific method?

BY JONAH LEHRER

DECEMBER 12, 2011

HOW SCIENCE GOES WRONG.

Einsteinium

T

FEATURES

How many scientific papers just aren't true?

Enough that basing government policy on 'peer-reviewed studies' isn't all it's cracked up to be

Donna Laframboise

The New York Times

SCIENCE

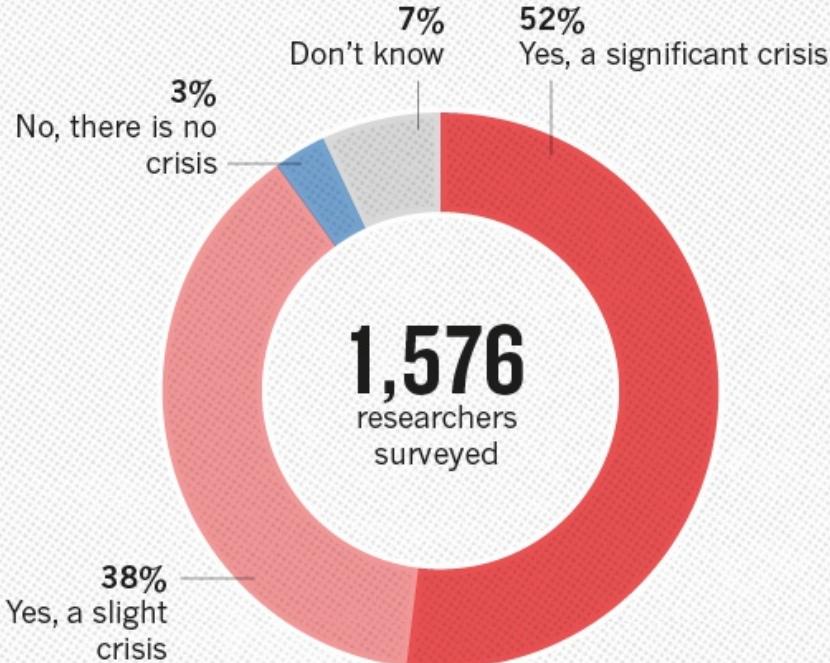
Science, Now Under Scrutiny Itself

The science 'reproducibility crisis' – and what can be done about it

March 15, 2017 by Ottoline Leyser, Danny Kingsley And Jim Grange, The Conversation

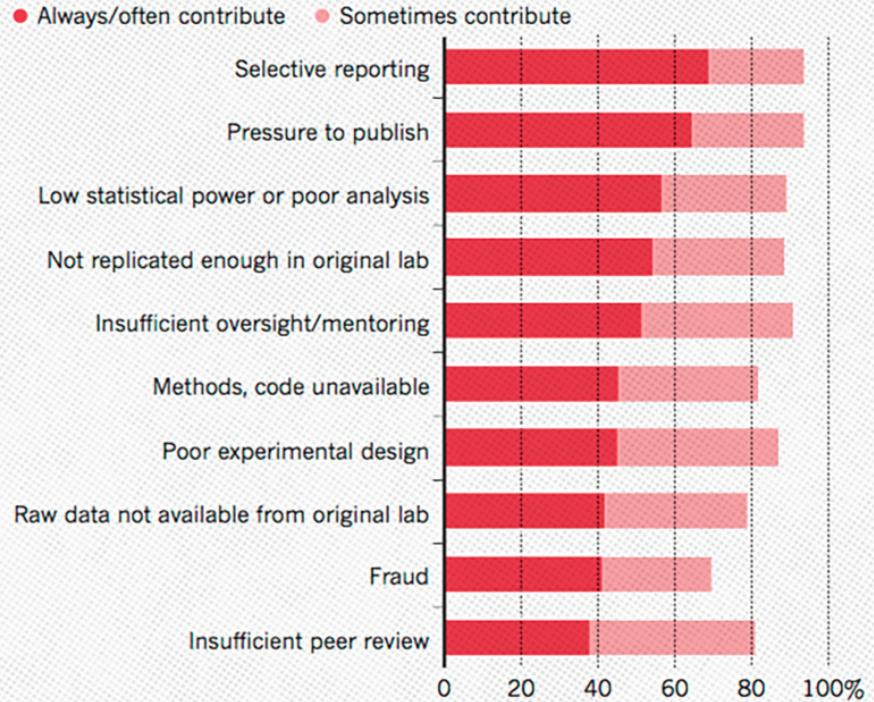
Reproducibility crisis

IS THERE A REPRODUCIBILITY CRISIS?

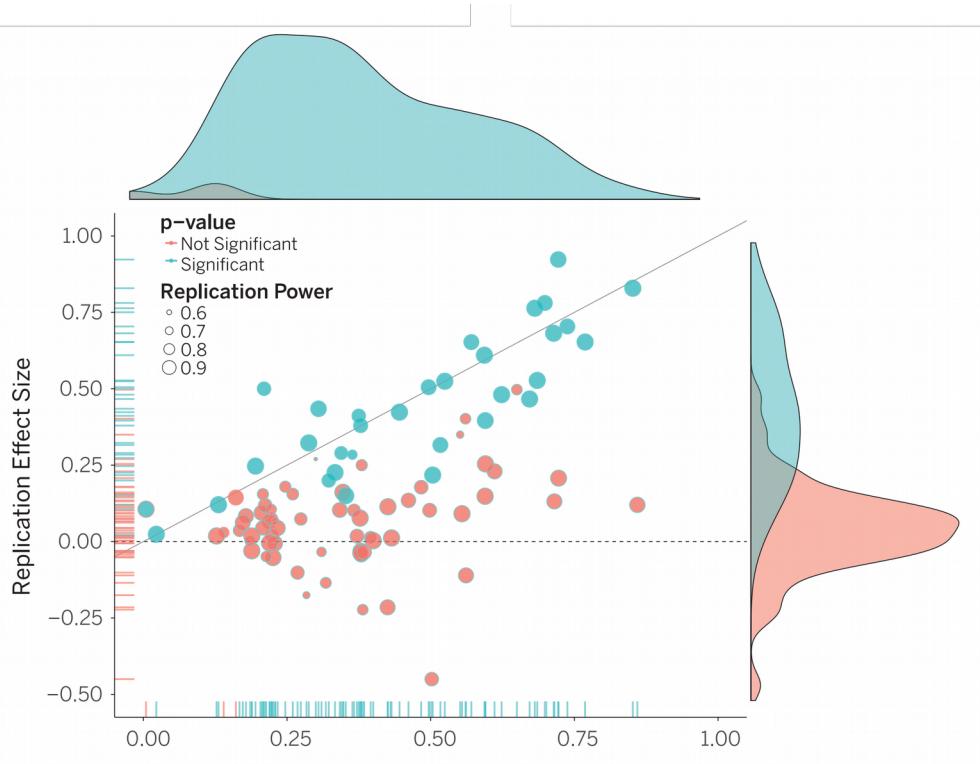


WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.



Reproducibility efforts



- Replication studies of 100 experimental and correlation studies (psychology studies)
- **97% original statistically significant results, 37% of replications**
- **Replication effects half the magnitude of original**

Original study effect size versus replication effect size (correlation coefficients). Diagonal line represents replication effect size equal to original effect size. Dotted line represents replication effect size of 0. Points below the dotted line were effects in the opposite direction of the original. Density plots are separated by significant (blue) and nonsignificant (red) effects.

Publications are advertisement

“An article about (computational) science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship.

**The actual scholarship is the complete ... set of instructions
and data which generated the figures.”**

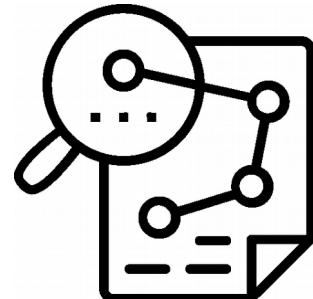
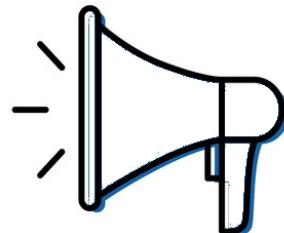
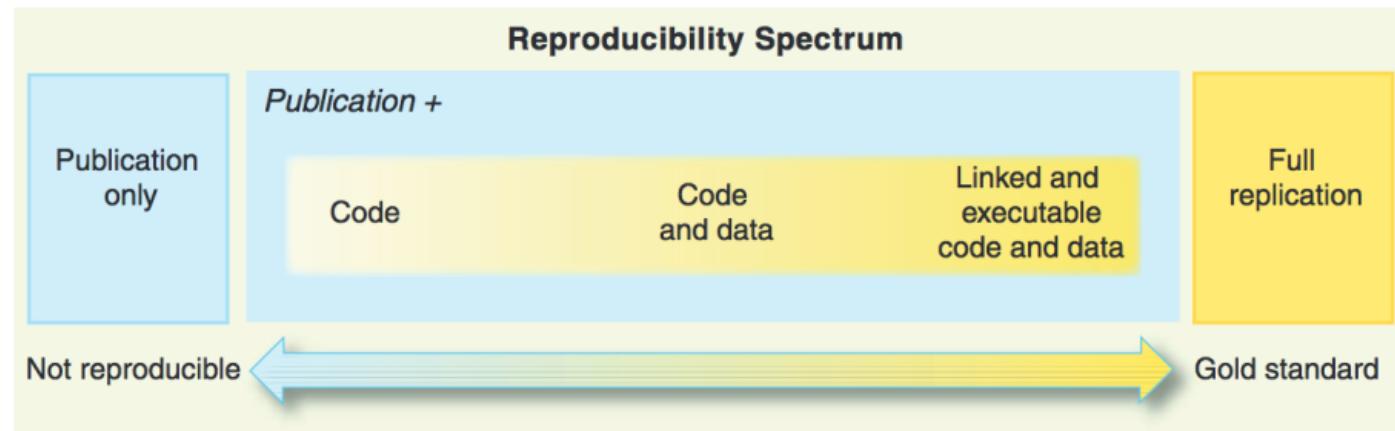
David Donoho, 1998

Publications are advertisement

“An article about (computational) science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship.

The actual scholarship is the complete ... set of instructions and data which generated the figures.”

David Donoho, 1998



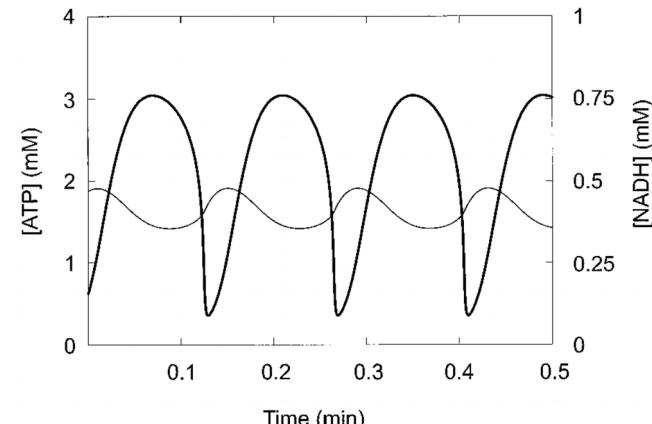
Standard Formats

- Exchangability, reusability, interoperability
 - Encoding information in computer readable format
- Annotations to ontologies
 - Knowledge integration (biological, computational)
 - Documentation (what is my model component)
- Reproducibility
 - Identical results in multiple tools (roadrunner, COPASI, JWS)
- Model quality
 - Minimal Information for models and simulation (MIRIAM, MIASE)
 - Automatic validation (unit checking, model consistency)
- Large ecosystems of tools
 - Simulation, parameter fitting, model analysis, visualization, ...

TABLE 1 Parameter Values

Parameter	Value
J_0	50.0 mM · min ⁻¹
k_1	550.0 mM ⁻¹ · min ⁻¹
K_i	1.0 mM
k_2	9.8 min ⁻¹
$k_{\text{GAPDH}+}$	323.8 mM ⁻¹ · min ⁻¹
$k_{\text{GAPDH}-}$	57823.1 mM ⁻¹ · min ⁻¹
$k_{\text{PGK}+}$	76411.1 mM ⁻¹ · min ⁻¹
$k_{\text{PGK}-}$	23.7 mM ⁻¹ · min ⁻¹
k_4	80.0 mM ⁻¹ · min ⁻¹
k_5	9.7 min ⁻¹
k_6	2000.0 mM ⁻¹ · min ⁻¹
k_7	28.0 min ⁻¹
k_8	85.7 mM ⁻¹ · min ⁻¹
κ	375.0 min ⁻¹
φ	0.1
A	4.0 mM
N	1.0 mM
n	4

$$v_3 = \frac{k_{\text{GAPDH}+} k_{\text{PGK}+} S_3 N_1 (A - A_3) - k_{\text{GAPDH}-} k_{\text{PGK}-} S_4 A_3 N_2}{k_{\text{GAPDH}-} N_2 + k_{\text{PGK}+} (A - A_3)}$$





- initiative to coordinate the development of the various community standards and formats for computational models
- COMBINE meeting & HARMONY hackathon
- Core standards
 - **SBML** Systems Biology Markup Language
 - **SED-ML** Simulation Experiment Description Language (SED-ML)
 - **SBGN** Systems Biology Graphical Notation
 - **CellML**
 - **SBOL** Synthetic Biology Open Language Data
 - **NeuroML**



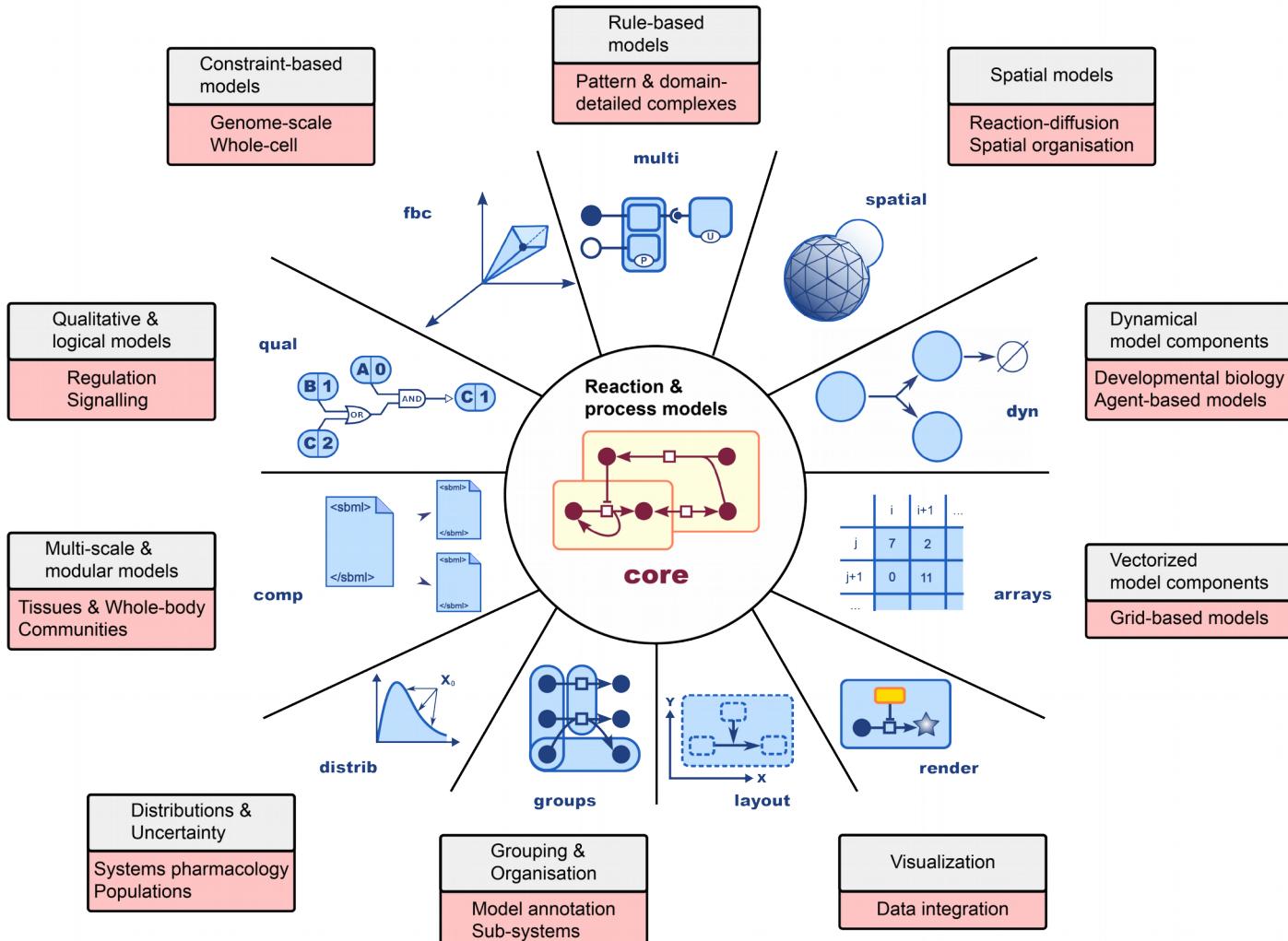
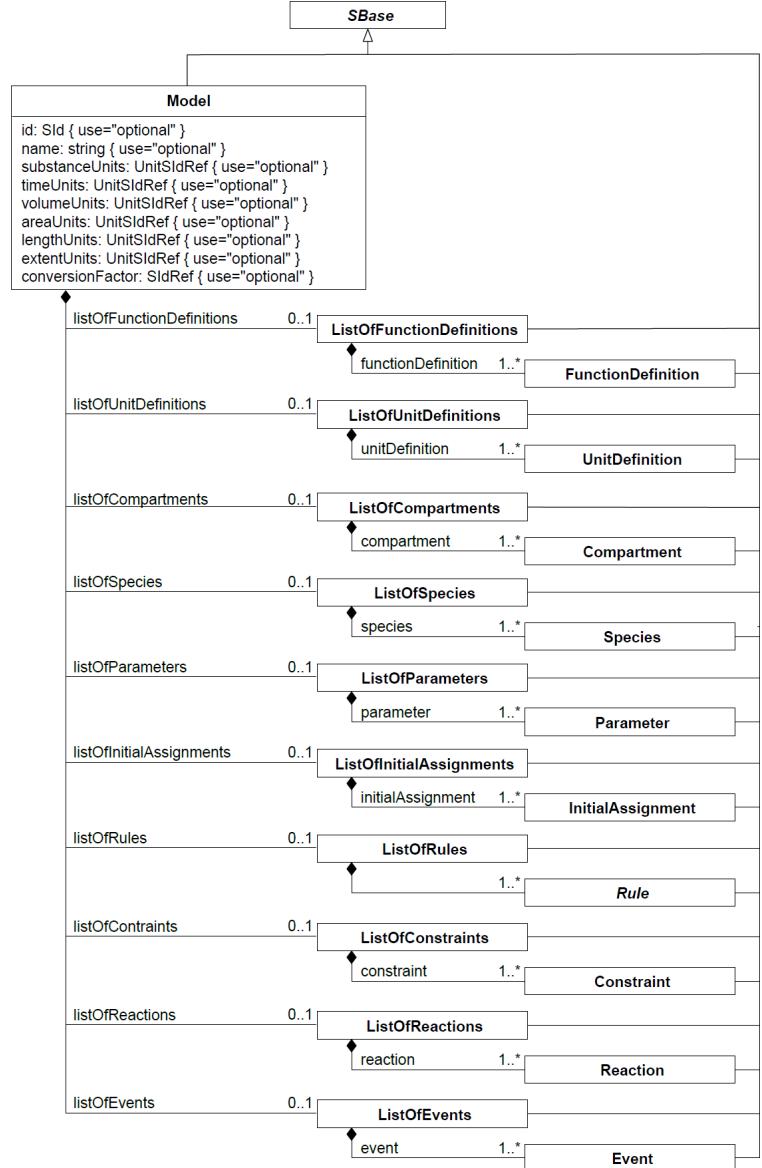


Figure 1: SBML Level 3 consists of a core (center) and specialized SBML Level 3 Packages (in blue) providing new syntactical constructs and cover new modeling approaches. The Packages support new types of modeling (in gray) needed for large and complex models such as used in various domains and fields of biology (in red).



- **De facto standard** for encoding computational models
- **Libraries:** libsbml (C++, python, R, JavaScript, ... & JSBML (Java))
- **TestSuite and Validators**
- **Components**
 - UnitDefinitions
 - FunctionDefinitions
 - Compartments
 - Species
 - Parameters
 - InitialAssignments
 - Rules
 - Constraints
 - Reactions
 - Events



SBML comp

- Hierarchical model composition
- Coupling of models

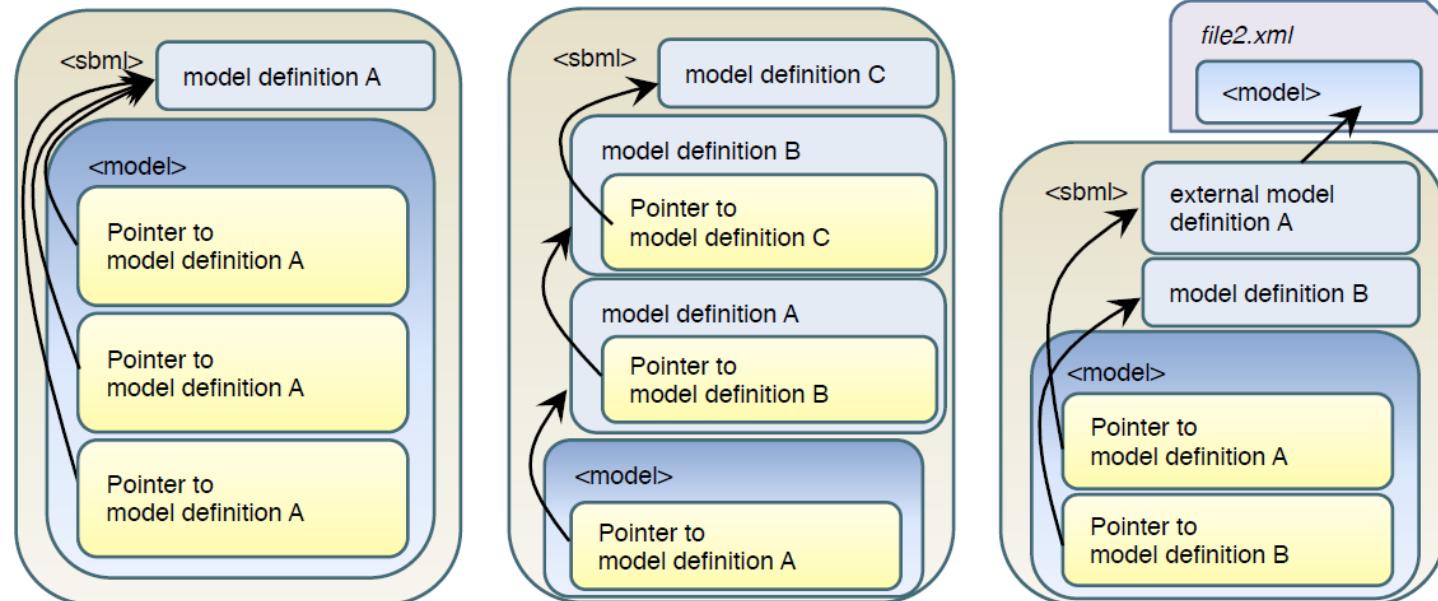
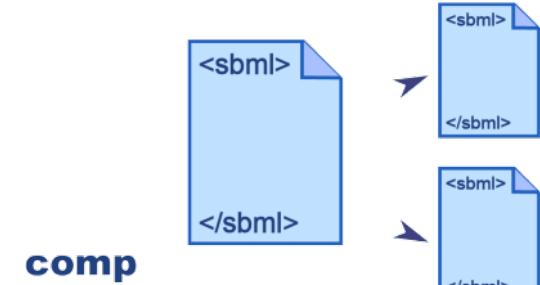


Figure 1: Three different examples of model composition scenarios. From left to right: (1) a model composed of multiple instances of a single, internally-defined submodel definition; (2) a model composed of a submodel that is itself composed of submodels; and (3) a model composed of submodels, one of which is defined in an external file.



Annotations

- **Ontology**
 - definition of controlled vocabulary with clear meaning and relationships
 - allows to precisely describe objects
- **Annotation**
 - process of attaching ontology terms to objects
 - important for mapping data onto models
 - important for automatic methods (model merging, reuse of components)
- **RDF triples**
 - (subject, verb, object)
- **Examples**
 - CHEBI (chemical entities)
 - UniProt (proteins)
 - Ontology Lookup Service

CHEBI:4167 - D-glucopyranose

Main ChEBI Ontology Automatic Xrefs Reactions Pathways Models

ChEBI Name **D-glucopyranose**
ChEBI ID **CHEBI:4167**
ChEBI ASCII Name D-glucopyranose
Definition A glucopyranose having d-configuration.
Stars This entity has been manually annotated by the ChEBI Team.
Supplier Information eMolecules:711823, eMolecules:29536451, MolPort-021-782-999
Download Molfile XML SDF

• Find compounds which contain this structure
• Find compounds which resemble this structure
• Take structure to the Advanced Search

more structures >>

Wikipedia License

Glucose (also called dextrose) is a simple sugar with the molecular formula C₆H₁₂O₆. Glucose is the most abundant monosaccharide, a subcategory of carbohydrates. Glucose is mainly made by plants and most algae during photosynthesis from water and carbon dioxide, using energy from sunlight. There it is used to make cellulose in cell walls, which is the most abundant carbohydrate. In energy metabolism, glucose is the most important source of energy in all organisms. Glucose for metabolism is partially stored as a polymer, in plants mainly as starch and amylopectin and in animals as glycogen. Glucose circulates in the blood of animals as blood sugar. The naturally occurring form of glucose is D-glucose, while L-glucose is produced synthetically in comparably small amounts and is of lesser importance. Glucose, as intravenous sugar solution, is on the World Health Organization's List of Essential Medicines, the most important medications needed in a basic health system. The name glucose derives through the French from the Greek γλυκός, which means "sweet," in reference to must, the sweet, first press of grapes in the making of wine. The suffix "-ose" is a chemical classifier, denoting a sugar.

[Read full article at Wikipedia](#)

Formula	C ₆ H ₁₂ O ₆
Net Charge	0
Average Mass	180.15588
Monoisotopic Mass	180.063
InChI	InChI=1S/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2/h2-11H,1H2/t2-,3-,4+,5-,6?/m1/s1
InChIKey	WQZGKKKJIJFFOK-GASJEMHNSA-N
SMILES	OC[C@H]1OC(O)[C@H](O)[C@@H](O)[C@H](O)[C@@H]1O

```
// -- Begin Antimony block converted from MAPKcascade.xml
// Created by libAntimony v2.9.3
model *MAPKcascade()
...
// Reactions:
J0: MKKK => MKKK_P; J0_V1*MKKK/((1 + (MAPK_PP/J0_Ki)^J0_n)*(J0_K1 + MKKK));
J1: MKKK_P => MKKK; J1_V2*MKKK_P/(J1_KK2 + MKKK_P);
J2: MKK => MKK_P; J2_K3*MKKK_P*MKK/(J2_KK3 + MKK);
J3: MKK_P => MKK_PP; J3_K4*MKKK_P*MKK_P/(J3_KK4 + MKK_P);
J4: MKK_PP => MKK_P; J4_V5*MKKK_PP/(J4_KK5 + MKK_PP);
J5: MKK_P => MKK; J5_V6*MKKK_P/(J5_KK6 + MKK_P);
J6: MAPK => MAPK_P; J6_k7*MKKK_PP*MAPK/(J6_KK7 + MAPK);
J7: MAPK_P => MAPK_PP; J7_k8*MKKK_PP*MAPK_P/(J7_KK8 + MAPK_P);
J8: MAPK_PP => MAPK_P; J8_V9*MAPK_PP/(J8_KK9 + MAPK_PP);
J9: MAPK_P => MAPK; J9_V10*MAPK_P/(J9_KK10 + MAPK_P);
...
end
-- End Antimony block

// -- Begin PhraSEDML block converted from main.xml
// Created by libphrasedml v1.0.7
// Models
model1 = model "MAPKcascade"

// Simulations
sim1 = simulate uniform(0, 4000, 1000)

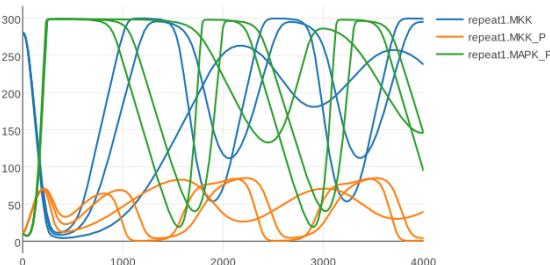
// Tasks
task1 = run sim1 on model1

// Repeated Tasks
repeat1 = repeat task1 for model1.J1_KK2 in [1, 10, 40], reset=true

// Outputs
plot "Sampled Simulation" repeat1.time vs repeat1.MKK, repeat1.MKK_P, repeat1.MAPK_PP
// -- End PhraSEDML block
```



Sampled Simulation



Modeling Tools

- **libRoadRunner**: High performance SBML simulator
- **tellurium**: Python based modeling environment (library & notebook)
- **COPASI**: GUI based tool for working with SBML models
- **JWS**: web based tool for simulations

Analysis and Visualization

Session: /home/mkoenig/git/cy3sbml/src/main/resources/sessions/Koenig_demo_10.cys

File Edit View Select Layout Apps Tools Help

Control Panel Network Style Select

Network Nodes Edges

Koenig_demo_10
Koenig_demo_10 36(0) 69(0)
Main: Koenig_demo_10 13(0) 14(0)

Koenig_demo_10
Koenig_demo_10 36(0) 69(0)
Main: Koenig_demo_10 13(0) 14(0)

Main: Koenig_demo_10

Enter search term...

Results Panel cy3sbml

Model : Koenig_demo_10 (Koenig_demo_10)
L3v1

Koenig Demo Metabolism

Description

This is a demonstration model in [SBML](#) format.
The content of this model has been carefully created in a manual research effort.
This file has been produced by [Matthias Koenig](#).

Terms of use

Copyright © 2016 Matthias Koenig.
Redistribution and use of any part of this model, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of this SBML file must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in a different form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

This model is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.

Table Panel f(x)

shared name name id sbml-type sbo metaid biomodels.sbo go fma label value units derivedUnits constant

external compartment	external c...	e	compartment	SBO:0000...	meta_22d897...	GO:0005...	FMA:70022	external co...	1.0E-6	m3	m^-3	<input type="checkbox"/>
cell compartment	cell comp...	c	compartment	SBO:0000...	meta_78b0e7...	SBO:0000290	FMA:6646	cell compar...	1.0E-6	m3	m^-3	<input type="checkbox"/>
plasma membrane	plasma me...	m	compartment	SBO:0000...	meta_bcdbe7...	SBO:0000290	GO:0005...	plasma me...	1.0	m2	m^-2	<input type="checkbox"/>
metabolic scaling factor	metabolic ...	Km_C	parameter	SBO:0000...	meta_c63c69...	SBO:0000027	FMA:63841	Km_C	3.0	mm	mol^m^-3)	<input checked="" type="checkbox"/>
metabolic scaling factor	metabolic ...	scale_f	parameter	SBO:0000...	meta_871a28...	SBO:0000186		metabolic s...	1.0E-6	dimensionl...	dimensionless	<input checked="" type="checkbox"/>
	Vmax_B		parameter	SBO:0000...	meta_ad89f6...	SBO:0000186		Vmax_B	2.0	mole_per_s	mol*s^-1)	<input checked="" type="checkbox"/>
	Vmax_A		parameter	SBO:0000...	meta_351d07...	SBO:0000186		Vmax_A	5.0	mole_per_s	mol*s^-1)	<input checked="" type="checkbox"/>
	Vmax_C		parameter	SBO:0000...	meta_074616...	SBO:0000186		Vmax_C	0.5	mole_per_s	mol*s^-1)	<input checked="" type="checkbox"/>
	Vmax_V3		parameter	SBO:0000...	meta_le2e9b...	SBO:0000186		Vmax_V3	0.5	mole_per_s	mol*s^-1)	<input checked="" type="checkbox"/>
	Vmax_V1		parameter	SBO:0000...	meta_78fe37...	SBO:0000186		Vmax_V1	1.0	mole_per_s	mol*s^-1)	<input checked="" type="checkbox"/>
	Km_A		parameter	SBO:0000...	meta_98f0e1...	SBO:0000027		Km_A	1.0	mm	mol*m^-3)	<input checked="" type="checkbox"/>
	Vmax_V4		parameter	SBO:0000...	meta_20f045...	SBO:0000186		Vmax_V4	0.5	mole_per_s	mol*s^-1)	<input checked="" type="checkbox"/>

Node Table Edge Table Network Table

Memory

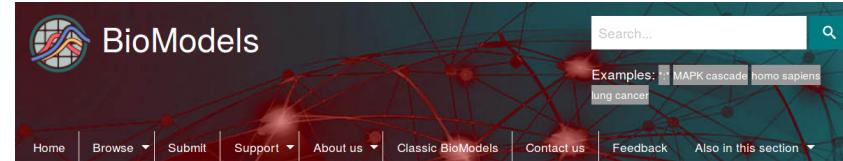
Model databases

▪ Biomodels

- large collection of freely available models (SBML and others)
- curated & uncurated
- <https://biomodels.org>

▪ JWS

- similar database, allows for online simulations
- https://jjj.bio.vu.nl/models/experiments/elowitz2000_fig1c/simulate



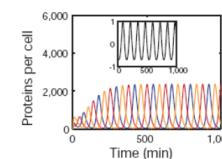
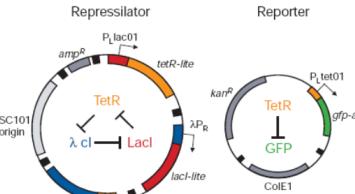
Elowitz and Leibler (2000), The Repressilator

July 2006, model of the month by Dominic P. Tolle

Original model: BIOMD00000000012

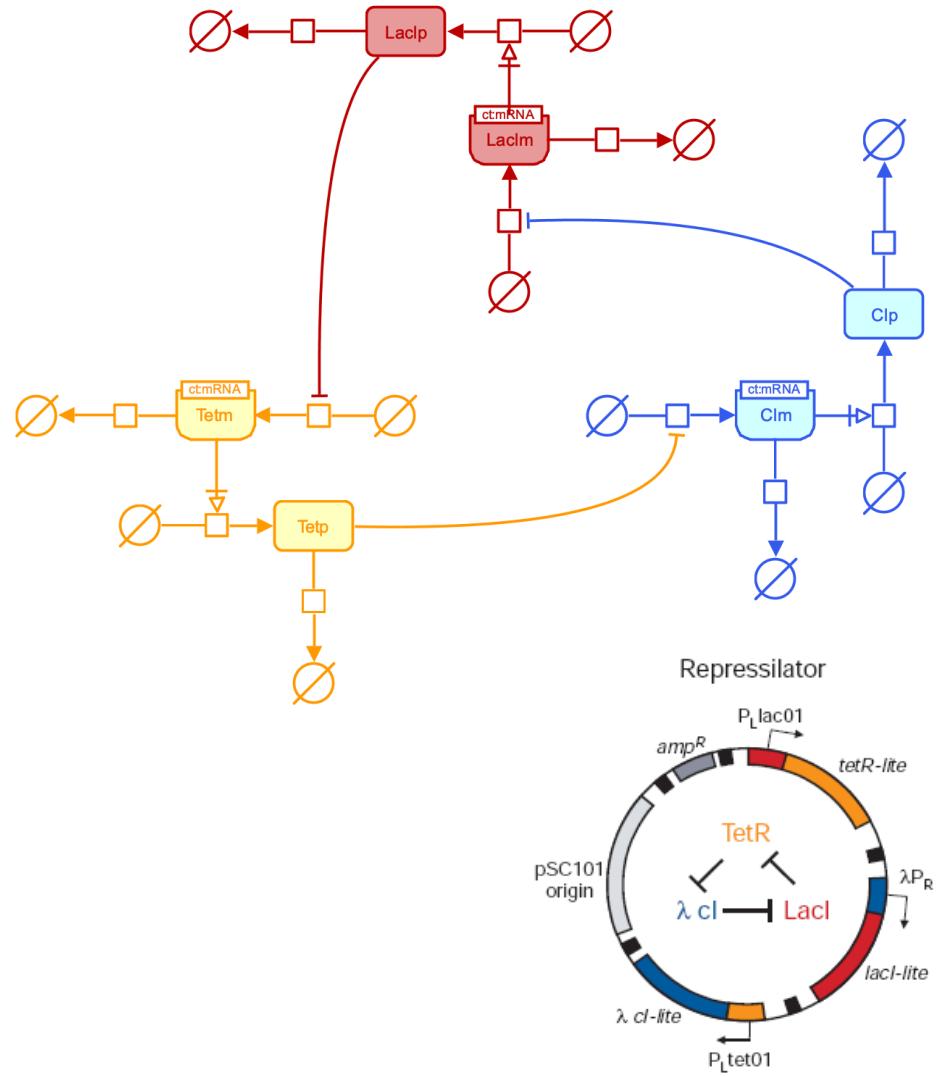
One of the major goals of Systems Biology is the elucidation of the control logic which determines the behaviour of naturally occurring biological systems[1]. To this end, Systems Biologist often create mathematical models designed to mimic a carefully observed biological system. Traditionally, the modeller acquires data, creates his model and tests the model against the available data. In an interesting take on the conventional way of modelling, Elowitz and Leibler[2] built a mathematical model of transcription regulation describing a cyclic negative-feedback loop made up of three repressor genes and their promoters. They used this model to determine the important parameters of the system and predict the systems behaviour, paying particular attention to parameter values that would cause the system to enter an unstable state leading to oscillatory behaviour. Finally the authors artificially reconstruct the system in *E. coli* using standard molecular biological approaches. In effect, rather than observing a natural system and explaining it in mathematical terms, the authors create a mathematical model to aid construction of an artificial control circuit. The result is an oscillating network which does not occur in nature, which the authors termed the Repressilator (see also the Brusselator[3] and the Oregonator[4] (BIOMD0000000040)).

The authors created a simple mathematical model of transcription regulation. The mathematical model was composed of six molecular species: three mRNA concentrations and three corresponding repressor protein concentrations. Each species was involved in transcription, translation and degradation reactions. Six coupled first-order differential equations described the dynamic behaviour of the system. Using the model, the authors predicted what parameters the stability of the steady state would be dependent on. In particular, the authors used the model to determine how to induce stable oscillations. Parameters that would favour oscillations were strong promoters, strong repression of transcription, cooperativity of repressor binding and similar lifetimes of mRNA and Proteins. The actual synthetic biological system was constructed from natural components using molecular biological techniques. Two alterations two the natural components were made to bring the system in line with the parameter space which favours oscillations: strong but tightly repressible hybrid promoters, and carboxy-terminal tags for the repressor proteins thus targetting them for protease degradation. A compatible reporter plasmid expressing GFP was also inserted into the system.

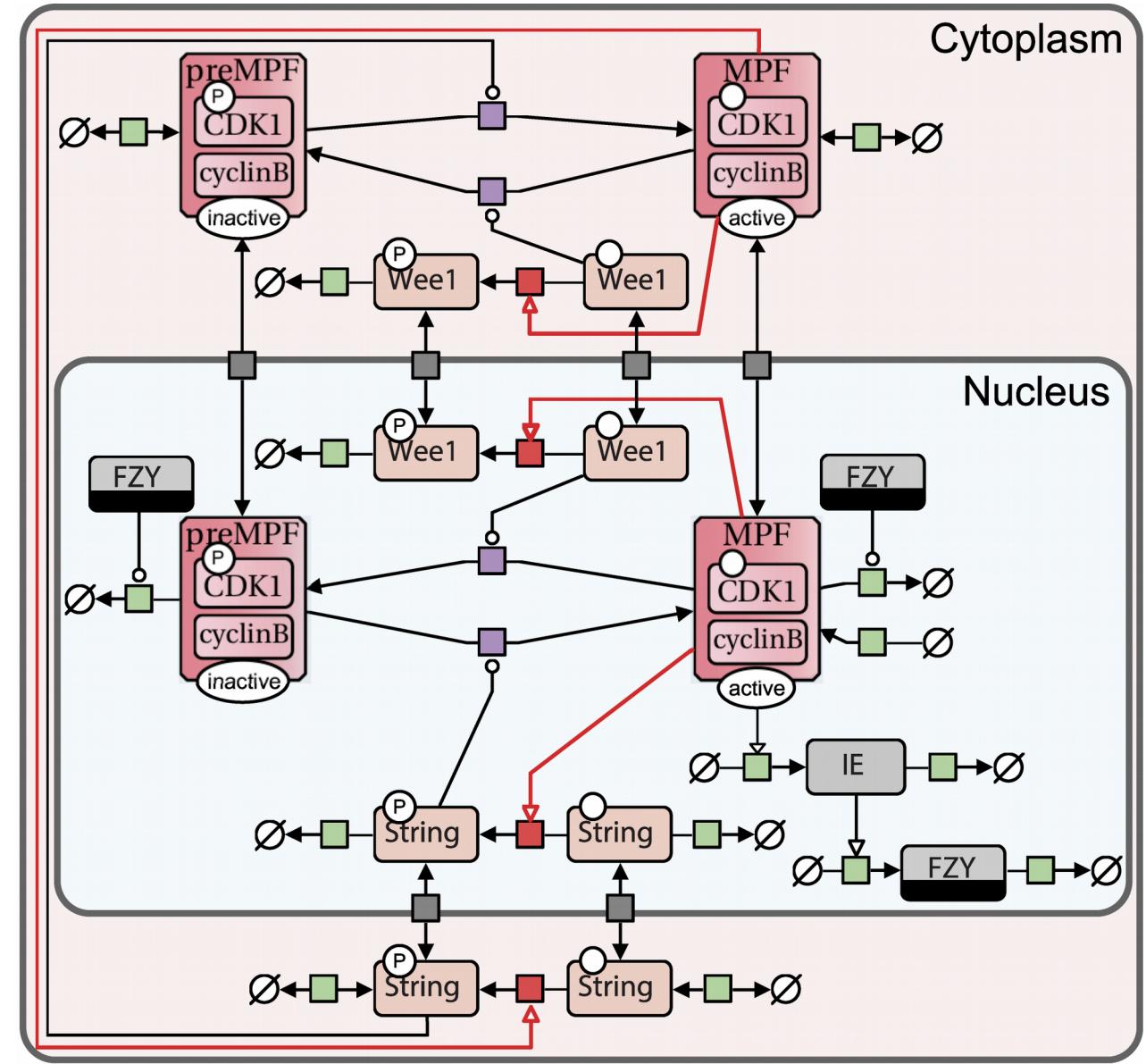


SBGN

- high quality, standard graphical languages for representing biological processes and interactions
 - PD: process description
 - AF: activity flow
 - ER: entity relationship
- <http://sbgn.github.io/sbgn/about>



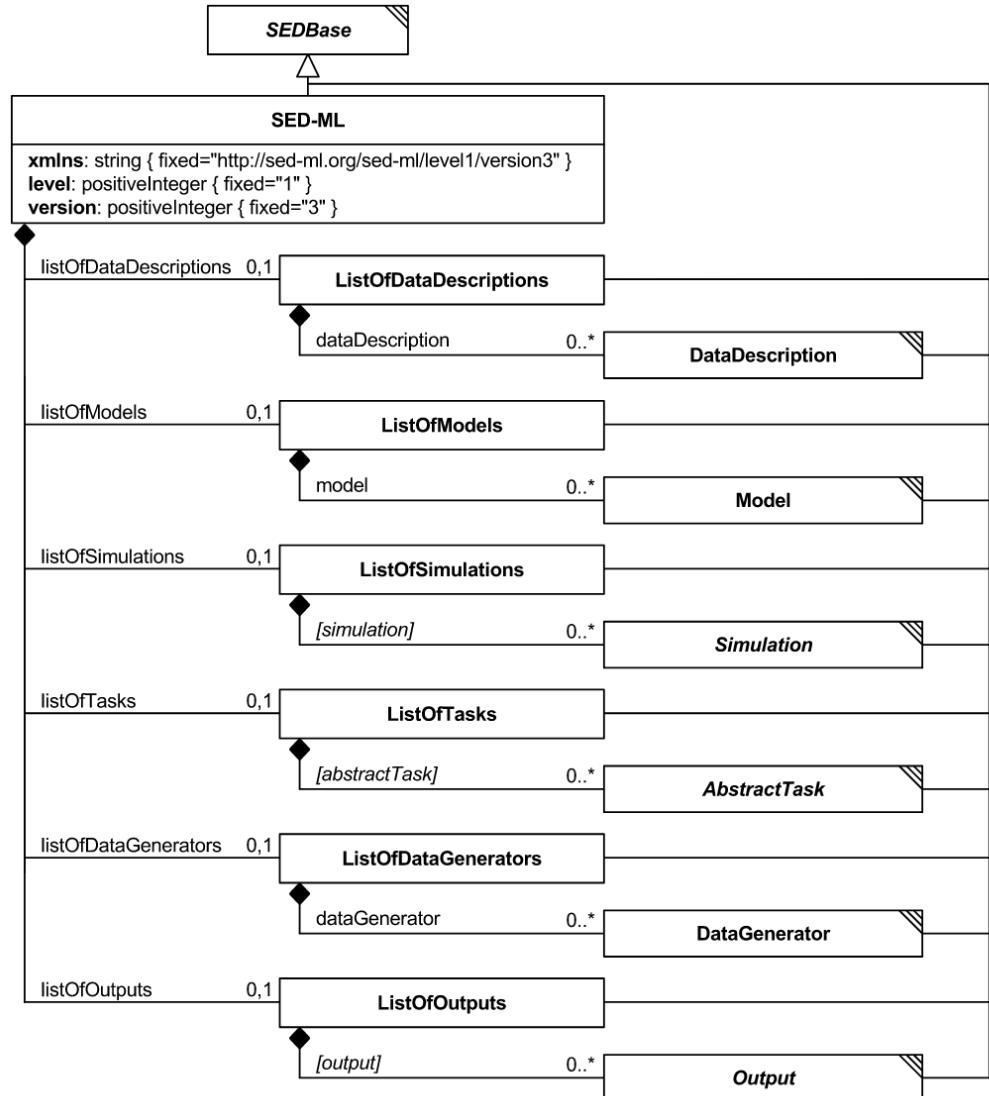
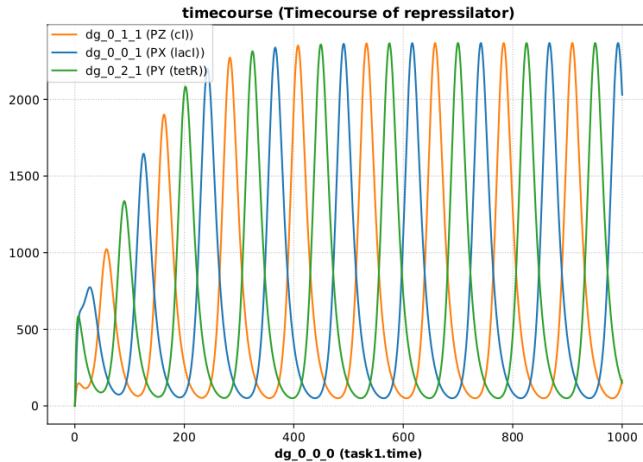
- Map of drosophila cell cycle



SED-ML

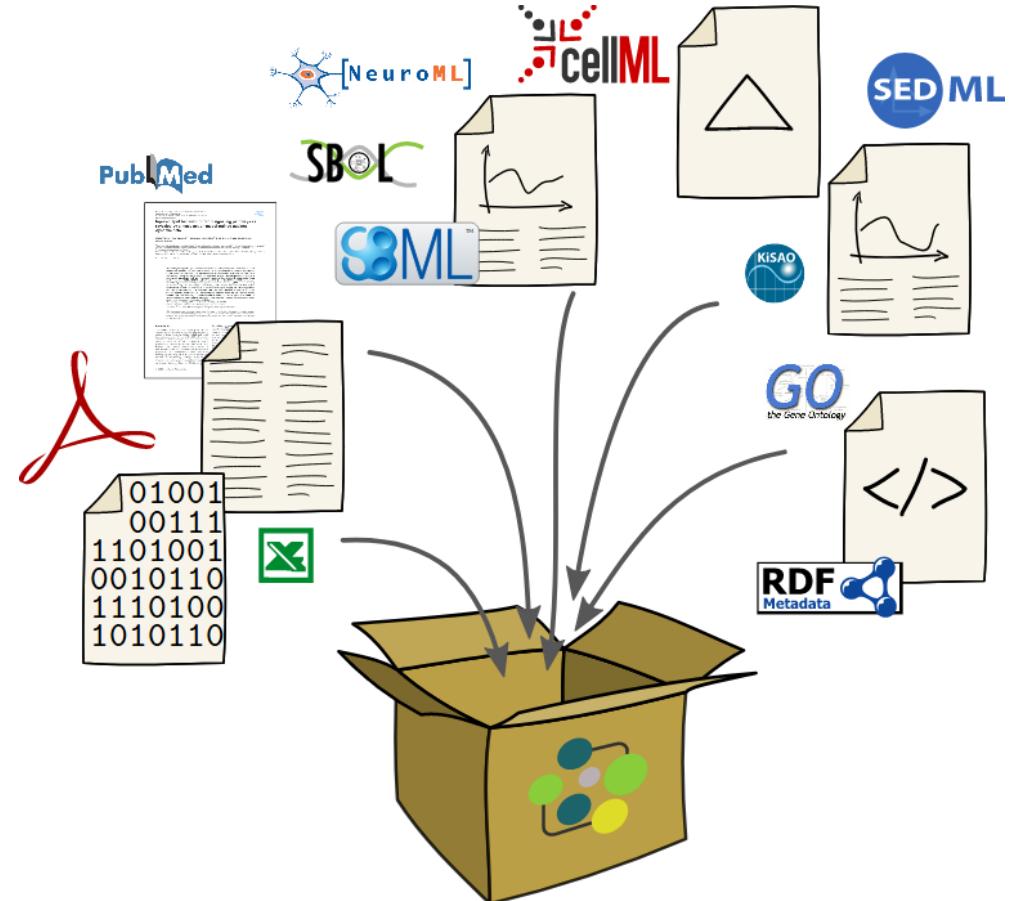


- **Simulation Experiment Description Markup Language (SED-ML)**
<https://sed-ml.github.io>
- SED-ML is an XML-based format for encoding simulation setups, to ensure exchangeability and reproducibility of simulation experiments.



COMBINE archive

- A COMBINE archive is a single file bundling the various documents necessary for a modeling and simulation project.
- The archive is encoded using the **Open Modeling EXchange format (OMEX)**.



Executable simulation model

<https://matthiaskoenig.github.io/exsimo/>

