A brief tutorial on SBML (the Systems Biology Markup Language)

Michael Hucka, Ph.D.

Department of Computing + Mathematical Sciences California Institute of Technology Pasadena, CA, USA

Email: mhucka@caltech.edu

Twitter: @mhucka

General background and motivations

Core features of SBML

A few additional details about SBML

Packages in SBML Level 3

A selection of resources for the SBML-oriented modeler

Closing

General background and motivations

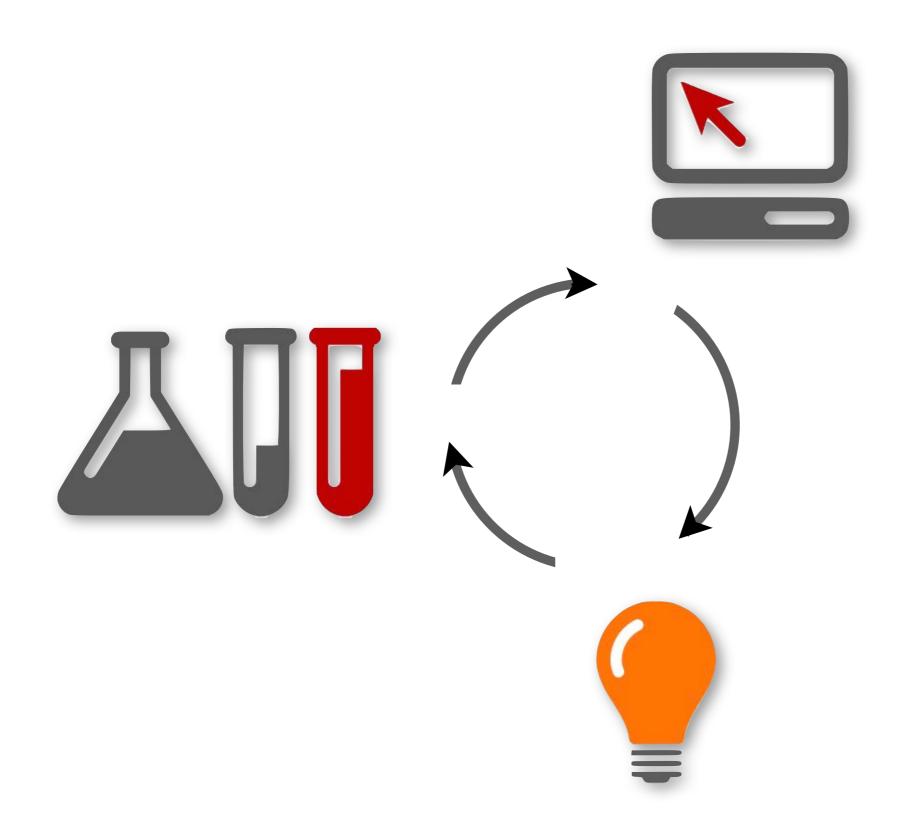
Core features of SBML

A few additional details about SBML

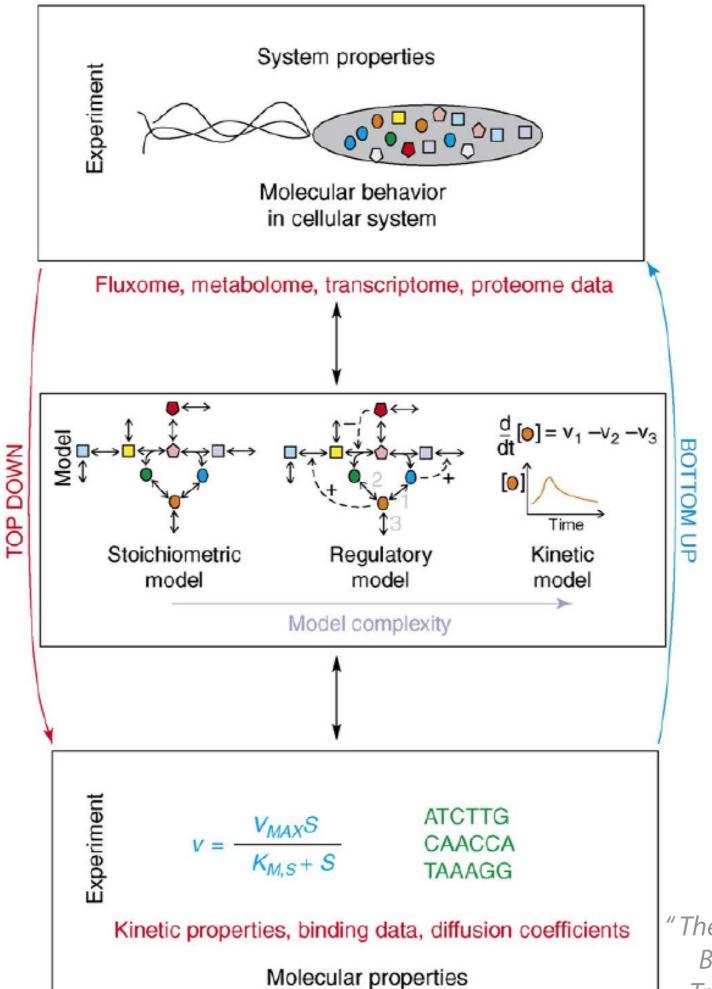
Packages in SBML Level 3

A selection of resources for the SBML-oriented modeler

Closing



Research today: experimentation, computation, cogitation



"The nature of systems biology" Bruggeman & Westerhoff, Trends Microbiol. 15 (2007).

Is it enough to communicate the model in a paper?

Traditional method of dissemination in the recent past

Problems:

- Errors in printing
- Missing information
- Dependencies on implementation
- Outright errors
- Larger model
 ⇒ more time & effort

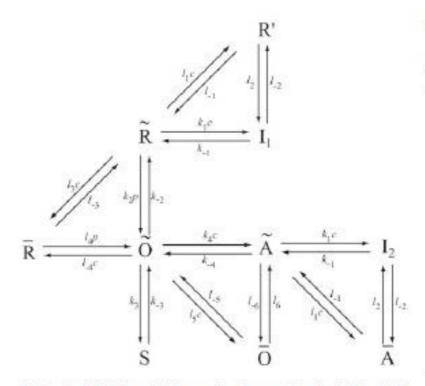


Fig. 2. The full IPR model. R, receptor; O, open; A, activated; S, shut; I, inactivated. c is [Ca²⁺]; p is [IP₃].

also has IP₃ and one other Ca²⁺ ion bound. For simplicity, we assume that the rate of Ca²⁺ binding to the inactivating site is independent of whether IP₃ is bound, or whether the receptor has been activated by Ca²⁺ (but see ref. 19 and later discussion). The R, R, O triangle models Ca²⁺-dependent binding of IP₃; Ca²⁺ modulates the interconversion of the receptor between two states, each of which can bind IP₃ with different kinetics. In type-3 IPR, Ca²⁺ enhances IP₃ binding but inhibits IP₃ binding in type-1 IPR (20–22). The reaction scheme shown here incorporates either possibility, depending on the parameter values. As we shall see, our model predicts that Ca²⁺ increases the rate of IP₃ binding to type-2 IPR.

To derive the model equations, we first define $K_i = k_{-i}/k_i$ and $L_i = l_{-i}/l_i$ for every appropriate integer i. We also let c and

$$R \xrightarrow{\phi_1(c)} I_1$$

$$\downarrow \phi_2(c) p \qquad \downarrow \phi_{-2}(c)$$

p denote [Ca²⁺] and [IP₃], respectively. Then, assuming the the transitions $\hat{R} \rightleftharpoons \hat{R}$, $\hat{O} \rightleftharpoons \hat{O}$, $\hat{A} \rightleftharpoons \hat{A}$ and $\hat{R} \rightleftharpoons \hat{R}'$ are fall and in instantaneous equilibrium, we get $c\hat{R} = L_3\hat{R}$, $c\hat{R} = L_1\hat{R}$. We now define the new variables \hat{R} $\hat{R} + \hat{R} + \hat{R}'$, $\hat{O} = \hat{O} + \hat{O}$, $\hat{A} = \hat{A} + \hat{A}$. Then

$$\frac{dR}{dt} = \phi_{-2}O - \phi_{2}pR + (k_{-1} + l_{-2})I_{1} - \phi_{1}R,$$

$$\frac{dO}{dt} = \phi_{2}pR - (\phi_{-2} + \phi_{4} + \phi_{3})O + \phi_{-4}A + k_{-3}S,$$

$$\frac{dA}{dt} = \phi_{4}O - \phi_{-4}A - \phi_{5}A + (k_{-1} + l_{-2})I_{2},$$

$$\frac{dI_{1}}{dt} = \phi_{1}R - (k_{-1} + l_{-2})I_{1},$$

$$\frac{dI_{2}}{dt} = \phi_{5}A - (k_{-1} + l_{-2})I_{2},$$

where

$$\phi_1(c) = \frac{(k_1L_1 + l_2)c}{L_1 + c(1 + L_1/L_3)},$$

$$\phi_2(c) = \frac{k_2L_3 + l_4c}{L_3 + c(1 + L_3/L_1)},$$

$$\phi_{-2}(c) = \frac{k_{-2} + l_{-4}c}{1 + c/L_5},$$

$$\phi_3(c) = \frac{k_3L_5}{L_5 + c},$$

$$\phi_4(c) = \frac{(k_4L_5 + l_6)c}{L_5 + c},$$

$$\phi_{-4}(c) = \frac{L_1(k_{-4} + l_{-6})}{L_1 + c},$$

$$\phi_5(c) = \frac{(k_1L_1 + l_2)c}{L_2 + c},$$

and where $R + O + A + S + I_1 + I_2 = 1$. Thus, given the fa equilibria above, Fig. 2 is equivalent to Fig. 3.

The model assumes that the binding of IP₃ and Ca²⁺ sequential, not independent. So, for instance, Ca²⁺ can bind to

Is it enough to make your (software X) code available?

It's vital for good science:

- Someone with access to the same software can try to run it, understand it, verify the computational results, build on them, etc.
- Opinion: you should always do this in any case

Is it enough to make your (software X) code available?

It's vital for good science—

- Someone with access to the same software can try to run it, understand it, verify the computational results, build on them, etc.
- Opinion: you should always do this in any case

But it's still not ideal for communication of scientific results:

- Doesn't necessarily encode biological semantics of the model
- What if they don't have access to the same software?
- What if they don't want to use that software?
- What if they want to use a different conceptual framework?
- And how will people be able to relate the model to other work?



Different tools ⇒ different interfaces & languages

General background and motivations

Core features of SBML

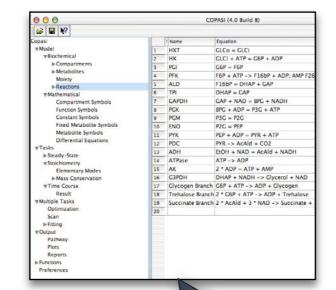
A few additional details about SBML

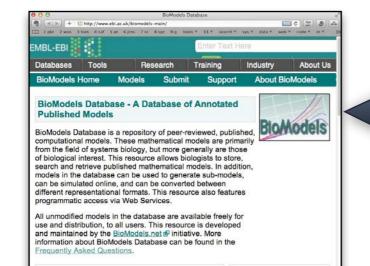
Packages in SBML Level 3

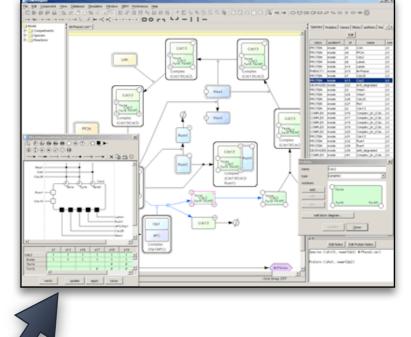
A selection of resources for the SBML-oriented modeler

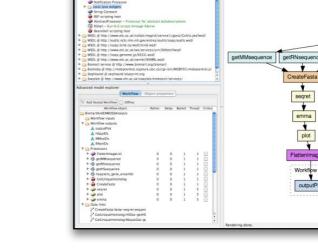
Closing

SBML. alinguation and solitons and solitons are solitons are solitons are solitons are solitons and solitons are solitons

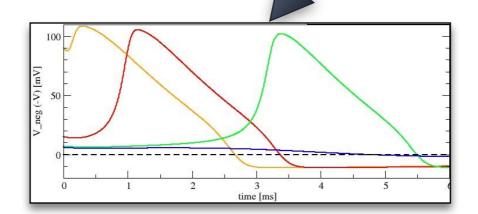








MMusiDs RNoriDs



SBML = Systems Biology Markup Language

Format for representing computational models of biological processes

- Data structures + usage principles + serialization to XML
- (Mostly) Declarative, not procedural—not a scripting language

Neutral with respect to modeling framework

E.g., ODE, stochastic systems, etc.

Important: **software** reads/writes SBML, **not humans**

The raw SBML (as XML)

```
st0fSpecies>
    <species compartment="cytosol" id="ES" initialAmount="0"</pre>
                                                                  name="ES"/>
    <species compartment="cytosol" id="P"</pre>
                                           initialAmount="0"
                                                                  name="P"/>
    <species compartment="cytosol" id="S"</pre>
                                           initialAmount="1e-20" name="S"/>
    <species compartment="cytosol" id="E"</pre>
                                            initialAmount="5e-21" name="E"/>
</listOfSpecies>
<listOfReactions>
    <reaction id="veq">
        <listOfReactants>
            <speciesReference species="E"/>
            <speciesReference species="S"/>
        </listOfReactants>
        st0fProducts>
            <speciesReference species="ES"/>
        <kineticLaw>
            <math xmlns="http://www.w3.org/1998/Math/MathML">
                <apply>
                    <times/>
                    <ci>ci>cytosol</ci>
                    <apply>
                        <minus/>
                        <apply>
                            <times/>
                            <ci>kon</ci>
                            <ci>E</ci>
                            <ci>S</ci>
                        </apply>
                        <apply>
                            <times/>
                            <ci>koff</ci>
                            <ci>ES</ci>
                        </apply>
                    </apply>
                </annlus
```

The process is central

- Literally called a "reaction" in SBML
- Participants are pools of entities (biochemical species)

$$n_a A + n_b B \xrightarrow{f([A],[B],[P],\ldots)} n_p P$$

$$n_c C \xrightarrow{f(...)} n_d D + n_e E + n_f F$$

Models can further include:

- Compartments
- Other constants & variables
- Discontinuous events
- Other, explicit math

- Unit definitions
- Annotations

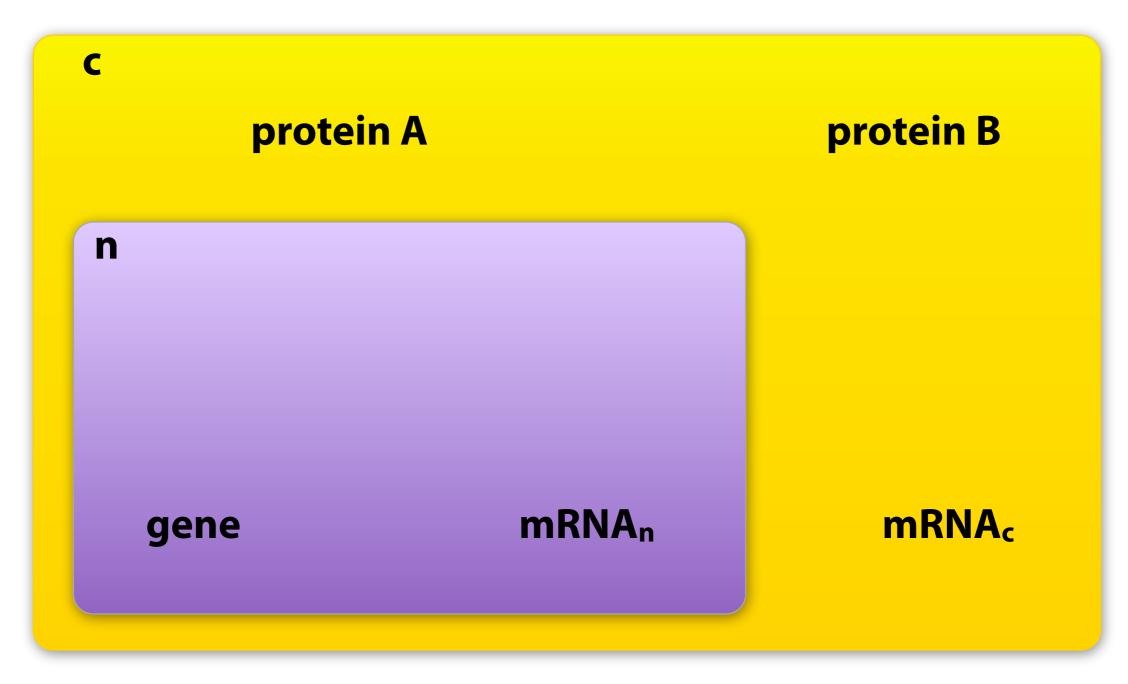
Core SBML concepts are fairly simple

Some basics of SBML core model encoding

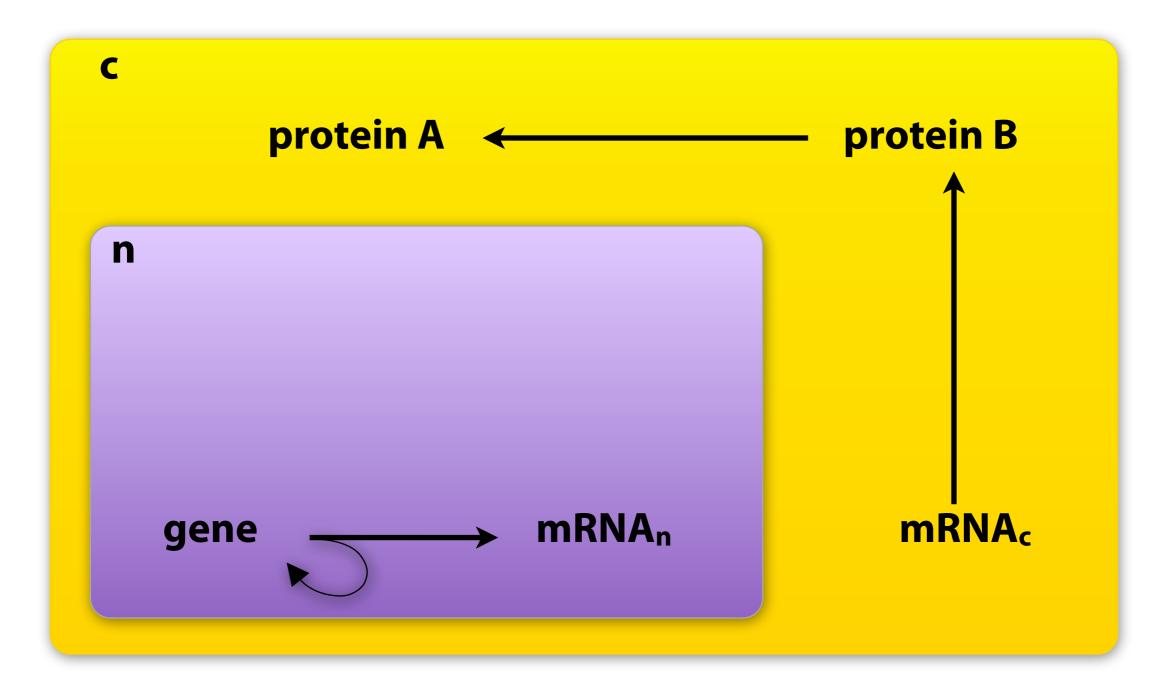
Well-stirred compartments



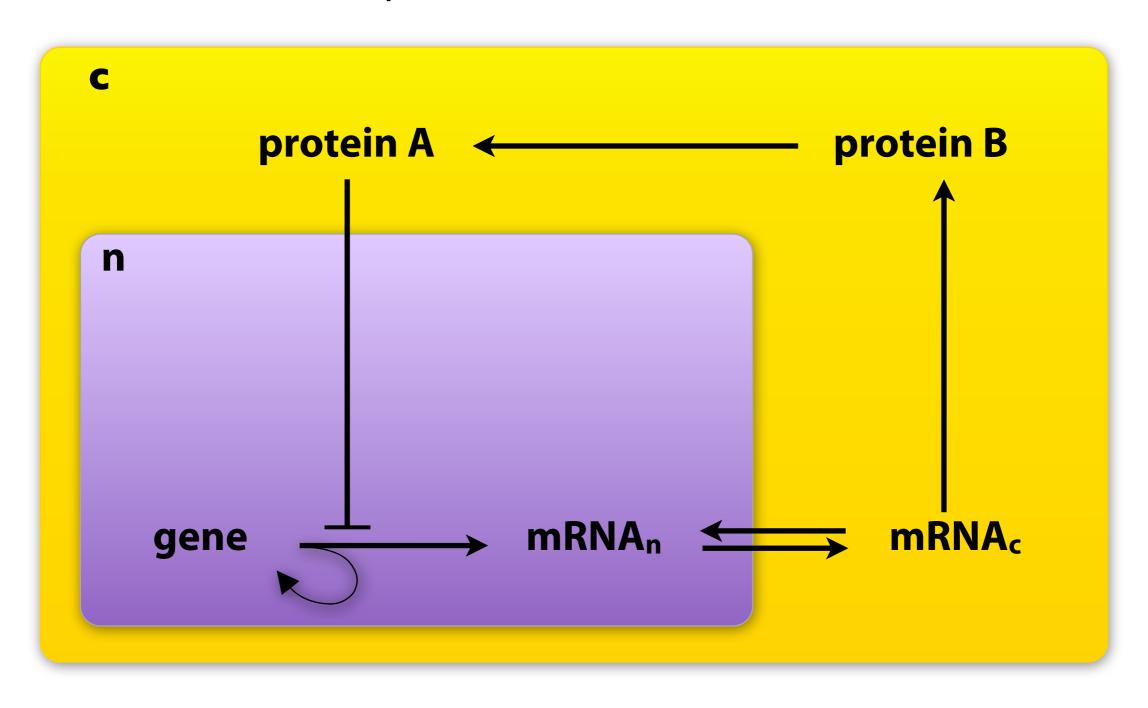
Species pools are located in compartments



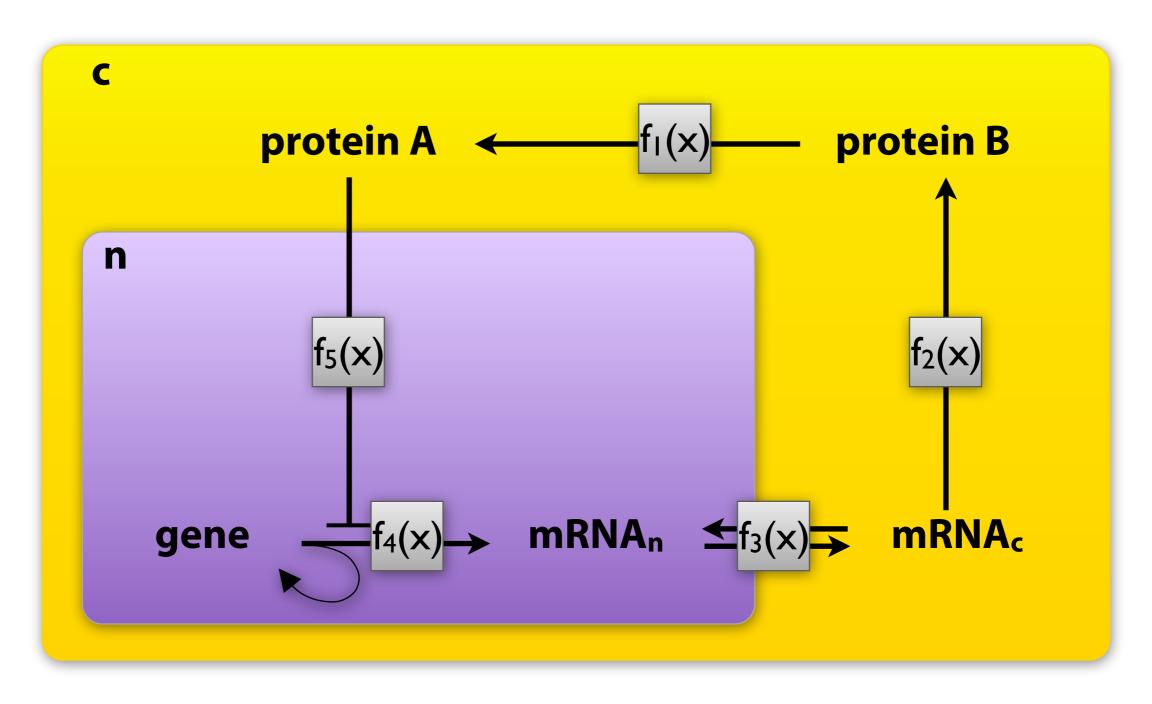
Reactions can involve any species anywhere



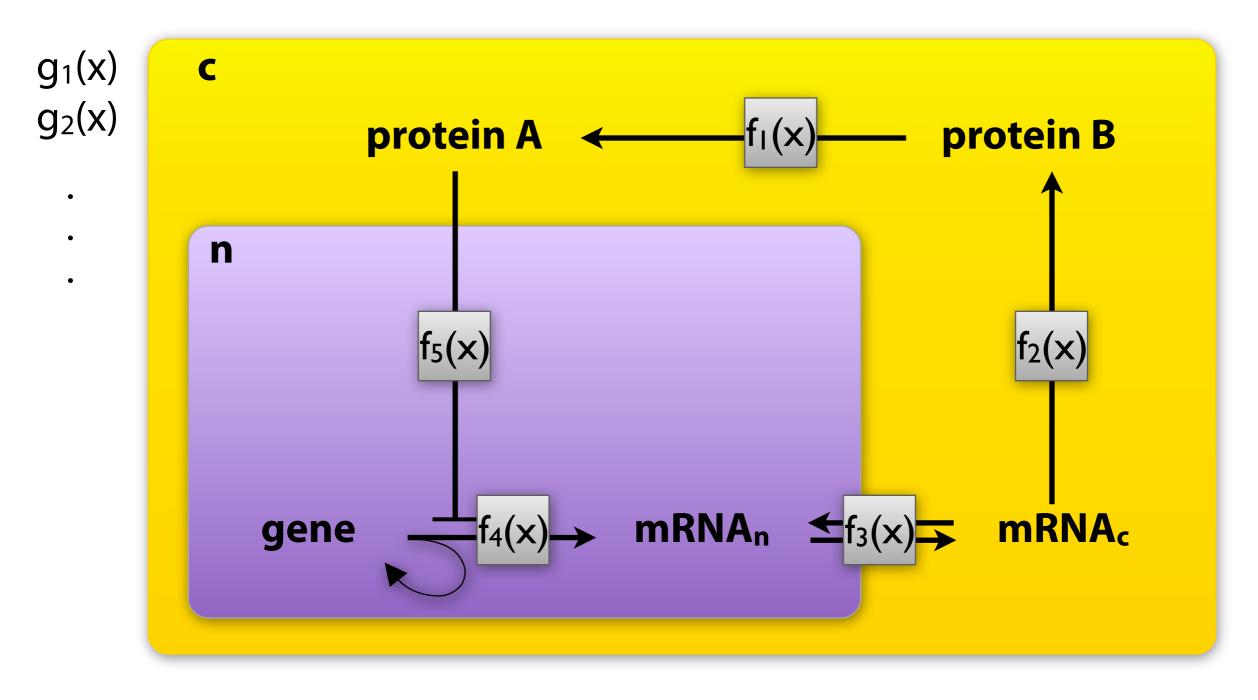
Reactions can cross compartment boundaries



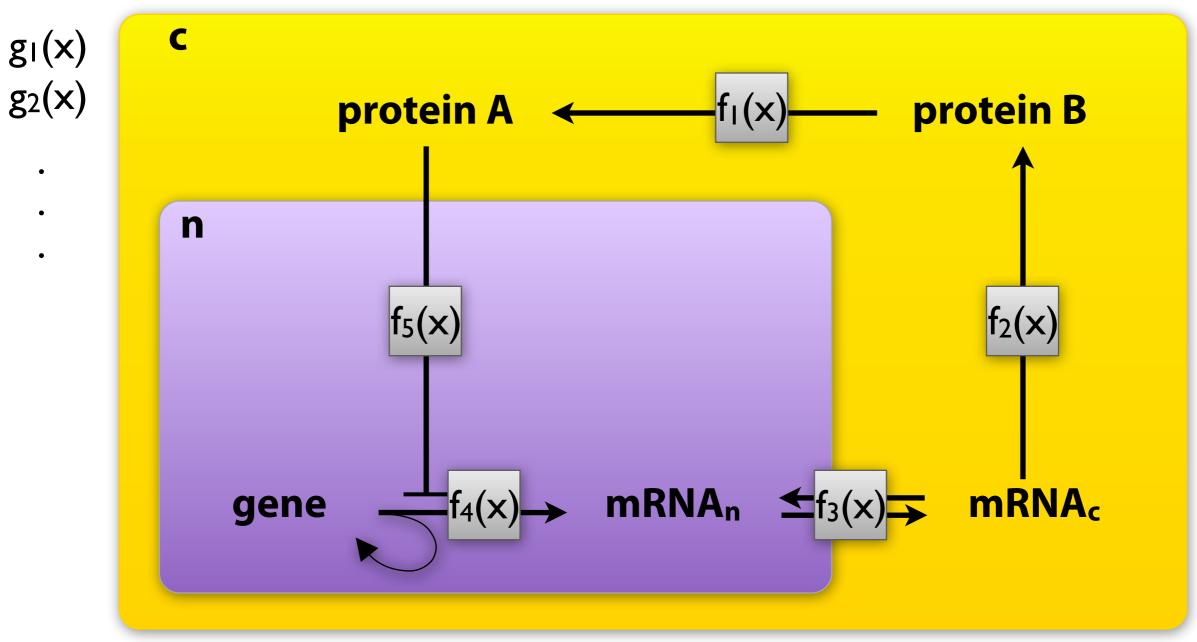
Reaction/process rates can be (almost) arbitrary formulas



"Rules": equations expressing relationships in addition to reaction sys.

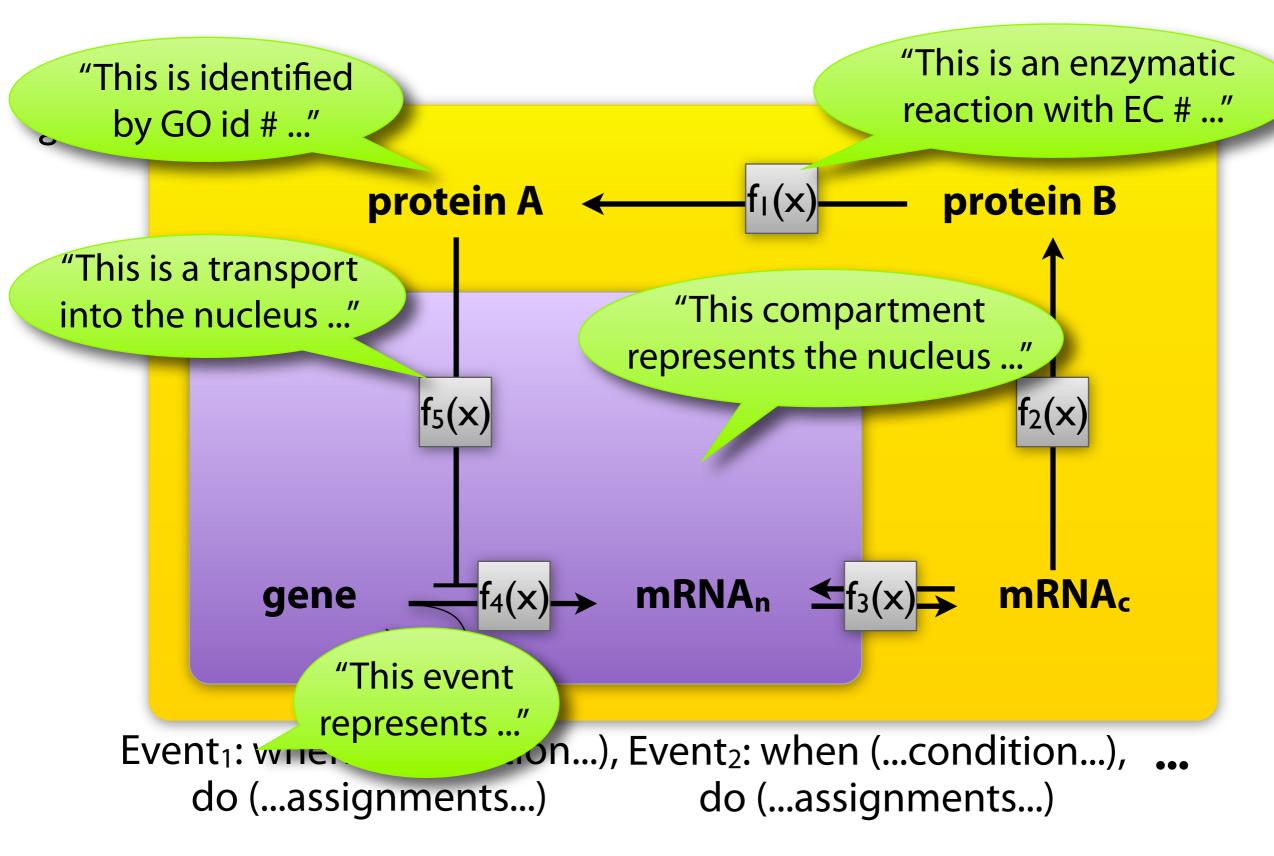


"Events": discontinuous actions triggered by system conditions



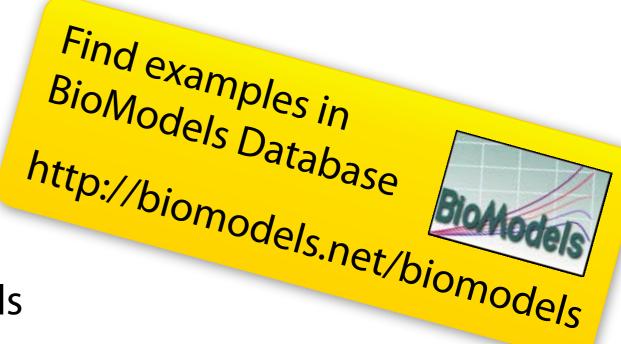
Event₁: when (...condition...), Event₂: when (...condition...), odo (...assignments...)

Annotations: machine-readable semantics and links to other resources



Many types of models can be encoded just using core SBML features

- Metabolic network models
- Signaling pathway models
- Conductance-based models
- Neural models
- Pharmacokinetic/dynamics models
- Infectious diseases



SBML Level 3 packages extend the core to support other types

• E.g.: Spatially inhomogeneous models, also qualitative/logical

Scope of SBML encompasses many types of models

SBML funding sources over the past 13+ years

National Institute of General Medical Sciences (USA)

European Molecular Biology Laboratory (EMBL)

JST ERATO Kitano Symbiotic Systems Project (Japan) (to 2003)

JST ERATO-SORST Program (Japan)

ELIXIR (UK)

Beckman Institute, Caltech (USA)

Keio University (Japan)

International Joint Research Program of NEDO (Japan)

Japanese Ministry of Agriculture

Japanese Ministry of Educ., Culture, Sports, Science and Tech.

BBSRC (UK)

National Science Foundation (USA)

DARPA IPTO Bio-SPICE Bio-Computation Program (USA)

Air Force Office of Scientific Research (USA)

STRI, University of Hertfordshire (UK)

Molecular Sciences Institute (USA)

General background and motivations

Core features of SBML

A few additional details about SBML

Packages in SBML Level 3

A selection of resources for the SBML-oriented modeler

Closing

SBML Files

Format: text (technically UTF-8)

Extension: usually .xml (not .sbml)

Does **not** store experimental data, or simulation descriptions

• But software may write proprietary metadata (annotations) in SBML

Applications usually have their own native format

Import/export SBML

Models exist in different "Levels and Versions" of SBML

Indicated at top of file

SBML identifiers and names

Most elements have both an "id" and a "name" field

- Identifier field has restricted syntax: abc123 or _abc123, etc.
 - This "id" field value is what you use in expressions
- Value of "name" field is almost completely unrestricted
 - Names with spaces, \$tr@nge and Fμηηγ characters!
- Must assign a value to "id", but "name" is optional

Some tools let you use the names and ignore id's (e.g., COPASI)—they generate the id's automatically

But some (especially those w/ script language features) expose id's

SBML "rules"

"Rules" in SBML define extra mathematical expressions

 E.g.: if need to express additional mathematical relationships beyond what is implied by the system of reactions

3 subtypes:

Rule type	General form	Example
algebraic	$0 = f(\mathbf{W})$	0 = S1 + S2
assignment	$x = f(\mathbf{V})$	x = y + z
rate	$dx/dt = f(\mathbf{W})$	dS/dt = 10.5

Rules define relationships that hold at all times

$$dS_1/dt = r_1 + r_2 + r_3 + \dots$$

$$dS_2/dt = -r_1 + r_5 + \dots$$

$$0 = f_1(\mathbf{W})$$

$$0 = f_2(\mathbf{W})$$

$$\dots$$

$$x = g_1(\mathbf{W})$$

$$y = g_2(\mathbf{W})$$

$$\dots$$

$$dm/dt = h_1(\mathbf{W})$$

$$dq/dt = h_2(\mathbf{W})$$

Equations derived from reaction definitions

$$dS_1/dt = r_1 + r_2 + r_3 + \dots$$

 $dS_2/dt = -r_1 + r_5 + \dots$

$$dS_2/dt = -r_1 + c$$
 $0 = f_1(\mathbf{W})$
 $0 = f_2(\mathbf{W})$
 \cdots
 $x = g_1(\mathbf{W})$
 $y = g_2(\mathbf{W})$
 \cdots
 $dm/dt = h_1(\mathbf{W})$
 $dq/dt = h_2(\mathbf{W})$

$$dS_1/dt = r_1 + r_2 + r_3 + \dots$$

 $dS_2/dt = -r_1 + r_5 + \dots$

Algebraic rules

$$0 = f_1(\mathbf{W})$$

$$0 = f_2(\mathbf{W})$$

• • •

$$x = g_1(\mathbf{W})$$

$$y = g_2(\mathbf{W})$$

• • •

$$dm/dt = h_1(\mathbf{W})$$

$$dq/dt = h_2(\mathbf{W})$$

. . .

$$dS_1/dt = r_1 + r_2 + r_3 + \dots$$

$$dS_2/dt = -r_1 + r_5 + \dots$$

$$0 = f_1(\mathbf{W})$$

$$0 = f_2(\mathbf{W})$$

$$\dots$$

Assignment rules

$$x = g_1(\mathbf{W})$$
 $y = g_2(\mathbf{W})$
 \dots
 $dm/dt = h_1(\mathbf{W})$
 $dq/dt = h_2(\mathbf{W})$

$$dS_1/dt = r_1 + r_2 + r_3 + \dots$$

$$dS_2/dt = -r_1 + r_5 + \dots$$

$$0 = f_1(\mathbf{W})$$

$$0 = f_2(\mathbf{W})$$

$$\dots$$

$$x = g_1(\mathbf{W})$$

$$y = g_2(\mathbf{W})$$

Rate rules

$$dm/dt = h_1(\mathbf{W})$$
$$dq/dt = h_2(\mathbf{W})$$

$$dS_1/dt = r_1 + r_2 + r_3 + \dots$$

$$dS_2/dt = -r_1 + r_5 + \dots$$

$$0 = f_1(\mathbf{W})$$

$$0 = f_2(\mathbf{W})$$

$$\dots$$

$$x = g_1(\mathbf{W})$$

$$y = g_2(\mathbf{W})$$

$$\dots$$

$$dm/dt = h_1(\mathbf{W})$$

$$dq/dt = h_2(\mathbf{W})$$

$$\dots$$

Rules and equations from reactions are taken together

SBML supports two annotation schemes

SBO (Systems Biology Ontology)

- For mathematical semantics
- One SBML object ← one SBO term
- Short, compact, tightly coupled but limited scope

MIRIAM (Minimum Information Requested In the Annotation of Models)

- For any kind of annotation
- One SBML object ← multiple MIRIAM annotations
- Larger, more free-form, wider scope

Both are externalized and independent of SBML

General background and motivations

Core features of SBML

A few additional details about SBML

Packages in SBML Level 3

A selection of resources for the SBML-oriented modeler

Closing

predefined math functions

SBML Level 3

SRMT revei i	2RML resei 5

user-defined functions

user-defined functions

MathML subset

RDF-based controlled

annotation scheme

text-string math notation

reserved namespaces for

annotations

no reserved namespaces

for annotations

MathML subset

no reserved namespaces for annotations

no controlled annotation scheme

RDF-based controlled annotation scheme

discrete events

discrete events

no discrete events

default values defined

default values defined

no default values

monolithic

monolithic

modular

Evolution of SBML continues

Today: SBML Level 3

- Level 3 Core provides framework for common models
- Level 3 packages add additional constructs to the Core



Level 3 package	What it enables	Status
Hierarchical model composition	Models containing submodels	✓
Flux balance constraints	Constraint-based models	✓
Qualitative models	Petri net models, Boolean models	✓
Graph layout	Diagrams of models	✓
Multicomponent/state species	Entities w/ structure; also rule-based models	draft
Spatial	Nonhomogeneous spatial models	draft
Graph rendering	Diagrams of models	draft
Groups	Arbitrary grouping of components	draft
Distributions	Numerical values as statistical distributions	in dev
Arrays & sets	Arrays or sets of entities	in dev
Dynamic structures	Creation & destruction of components	in dev
Annotations	Richer annotation syntax	

SBML Level 3 Qualitative models (released)

Some models use a **discrete logical formalism** – e.g., Boolean net, Petri

Often because there is not enough data for a quantitative model

SBML species are quantities, not states or levels

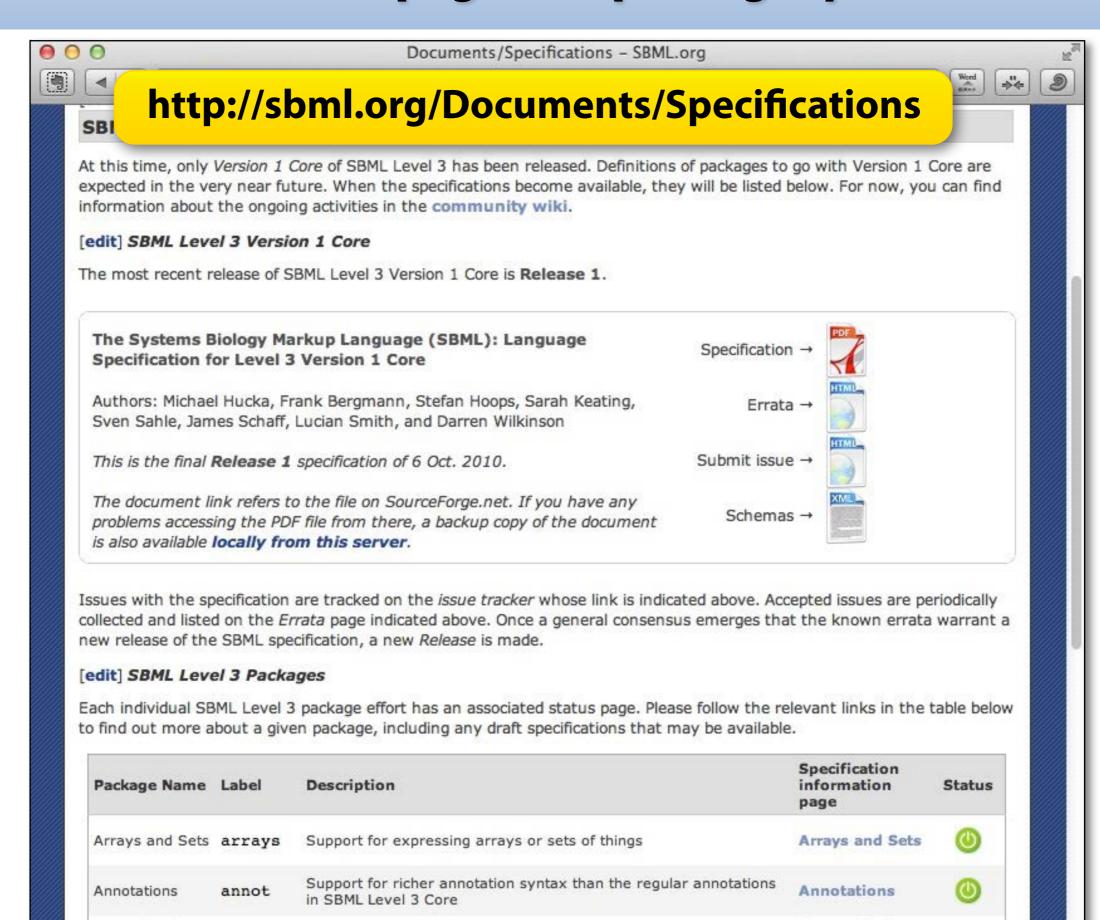
Yes, you could ignore that, but...

SBML Level 3 *Qualitative Models* package provides proper support

- Adds data structures for:
 - **Qualitative species**, with assume discrete values (e.g., 0 or 1)
 - **Transitions**, with inputs, outputs and function terms
 - At each time step in a simulation, all function terms evaluated
 - ▶ Terms evaluating to *true* dictate resulting state changes

Supported in GINsim, CellNOpt, Cell Collective

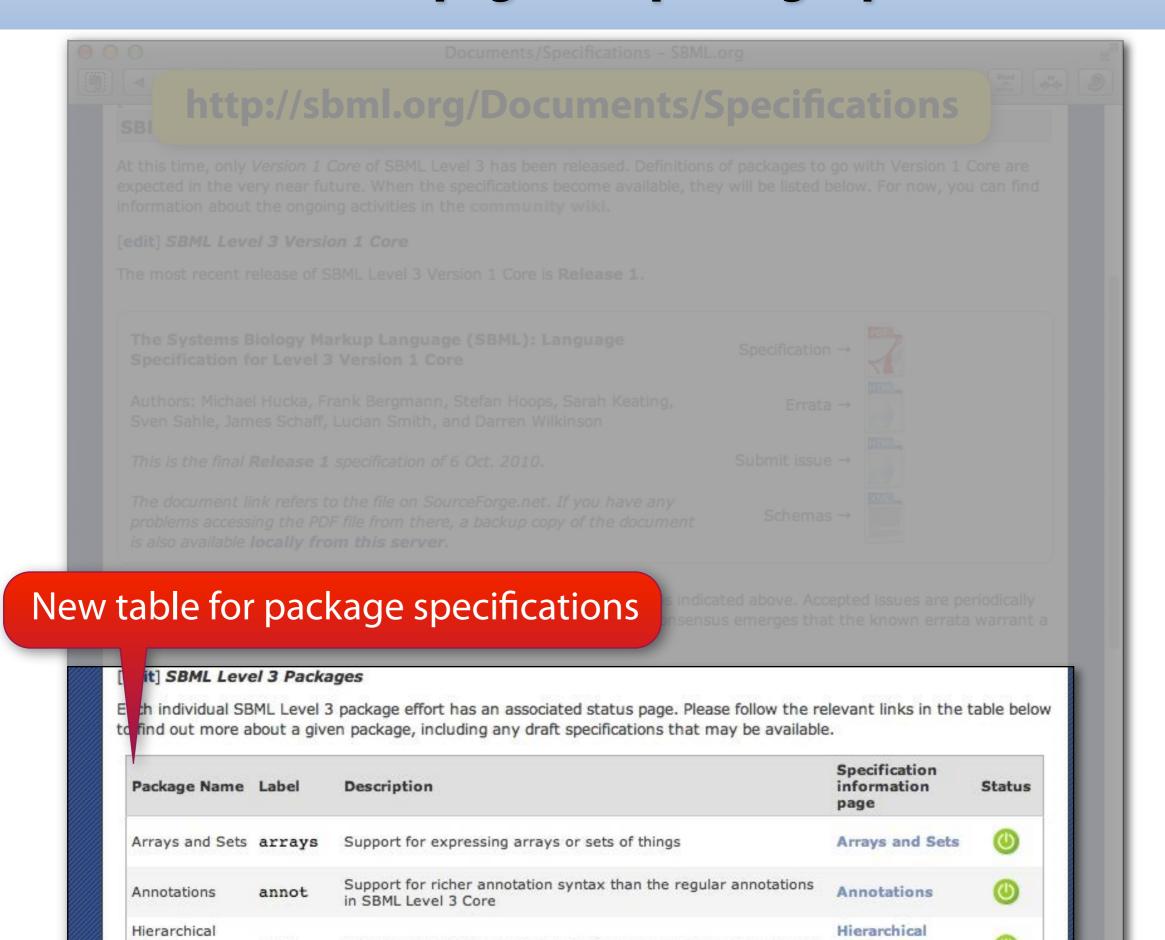
New table & status pages for package specifications



Hierarchical

Hierarchical

New table & status pages for package specifications



General background and motivations

Core features of SBML

A few additional details about SBML

Packages in SBML Level 3

A selection of resources for the SBML-oriented modeler

Closing

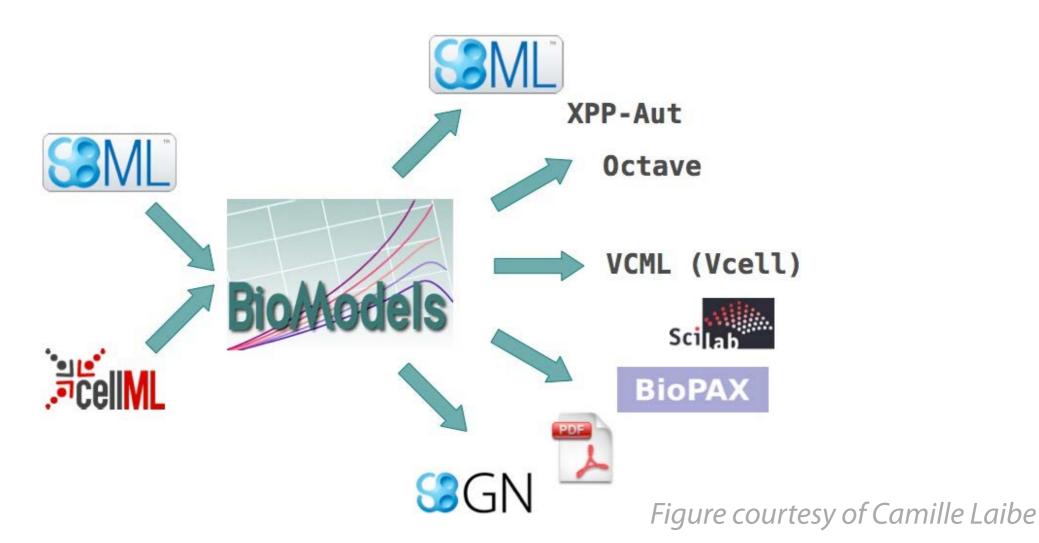
BioModels Database

Stores & serves quantitative models of biological interest

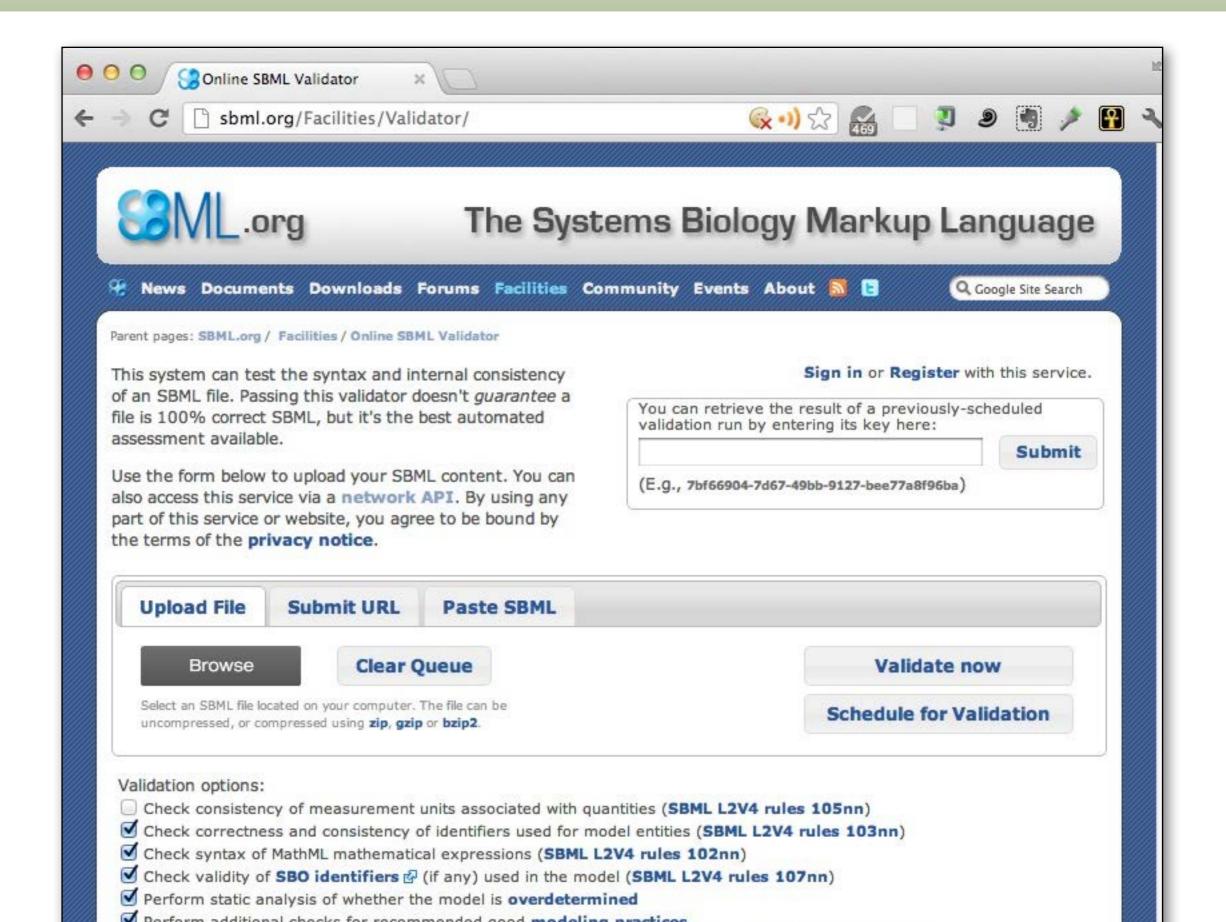
- Free, public resource
- Models must be described in peer-reviewed publication(s)

Hundreds of models are curated by hand

Imports & exports models in several formats



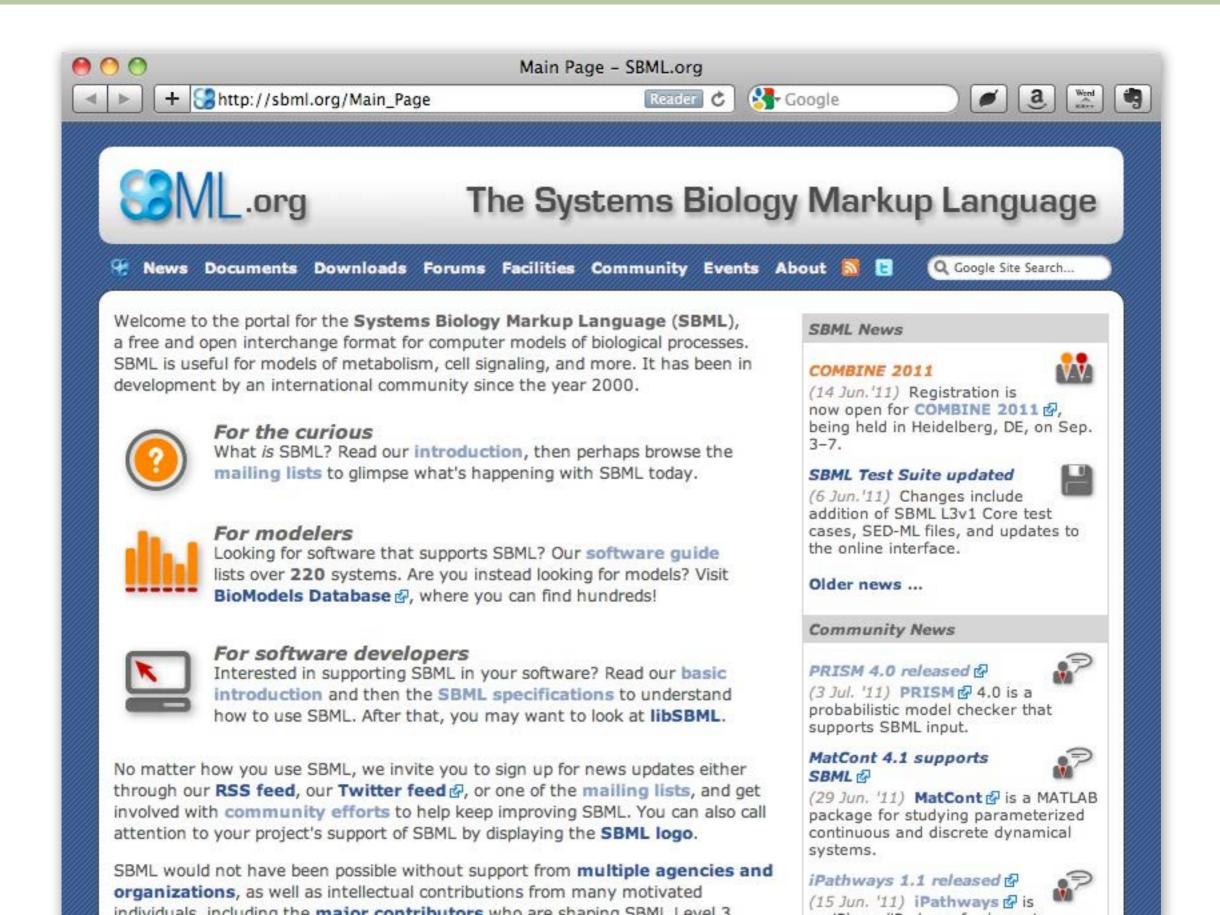
The Online SBML Validator



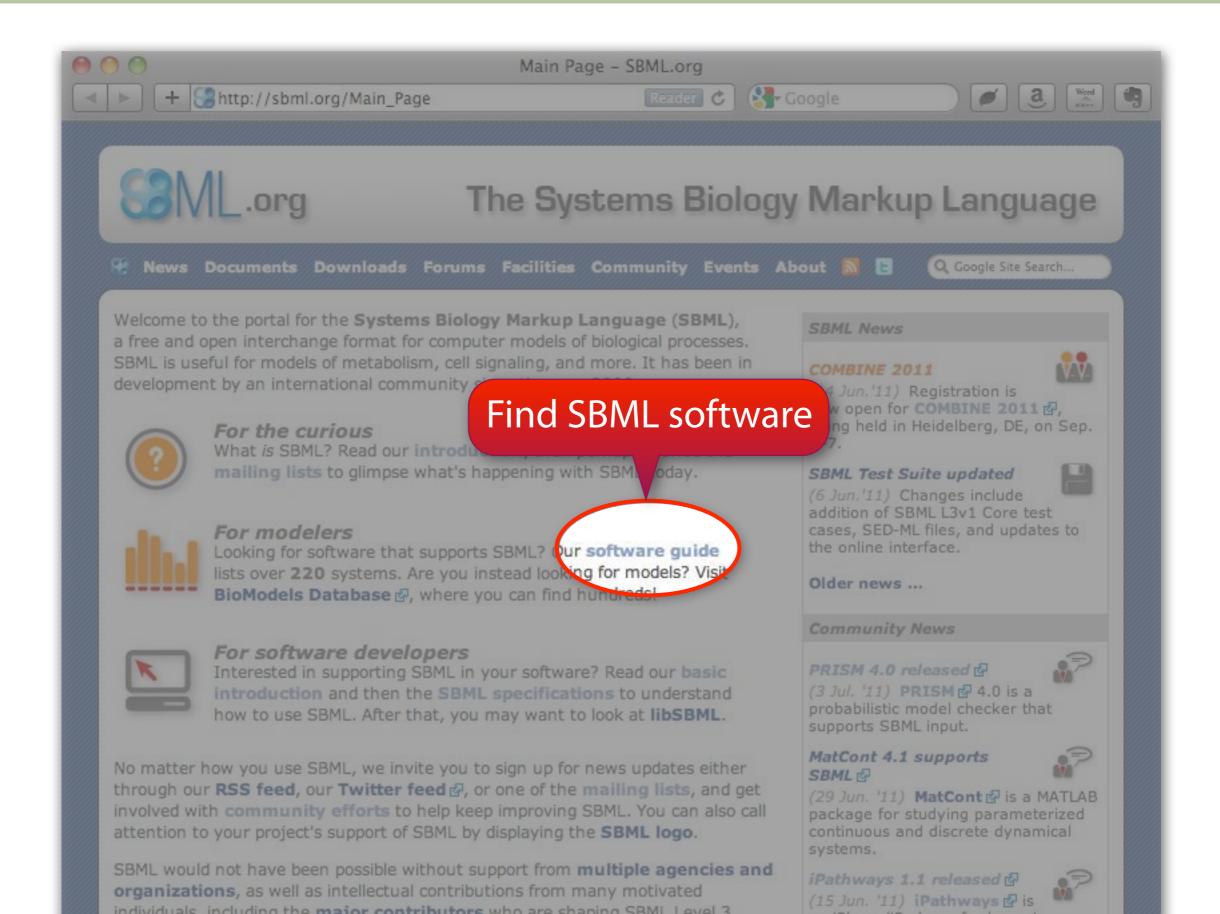
The Online SBML Validator



Find software in the SBML Software Guide



Find software in the SBML Software Guide







The Systems Biology Markup Language

News Documents Downloads Forums Facilities Community Events About 🔝 🖹 🔍 Google Site Search...

Parent pages: SBML.org

SBML Software Guide

The following pages describe SBML-compatible software packages known to us. We offer different ways of viewing the information, all drawn from the same underlying data collected from the systems' developers via our **software survey**. The *Matrix* provides a table listing all known software and a variety of their features; the *Summary* provides general descriptions of most of the software; and the *Showcase* provides a sequential slideshow of a subset of the software.

Number of software packages listed in the matrix today: 229.

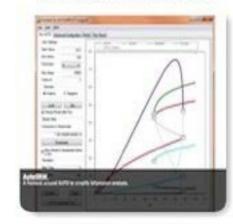
Go to the SBML Software Matrix



Go to the SBML Software Summary



Go to the SBML Software Showcase



Please use the survey form to notify us about additions and suggestions.

[edit] Historical trend

The following graph shows the total number of known SBML-compatible software packages each year, as counted by the SBML Team. The counts shown are for approximately the middle of each year.

200

libSBML

Reads, writes, validates SBML

Can check & convert units

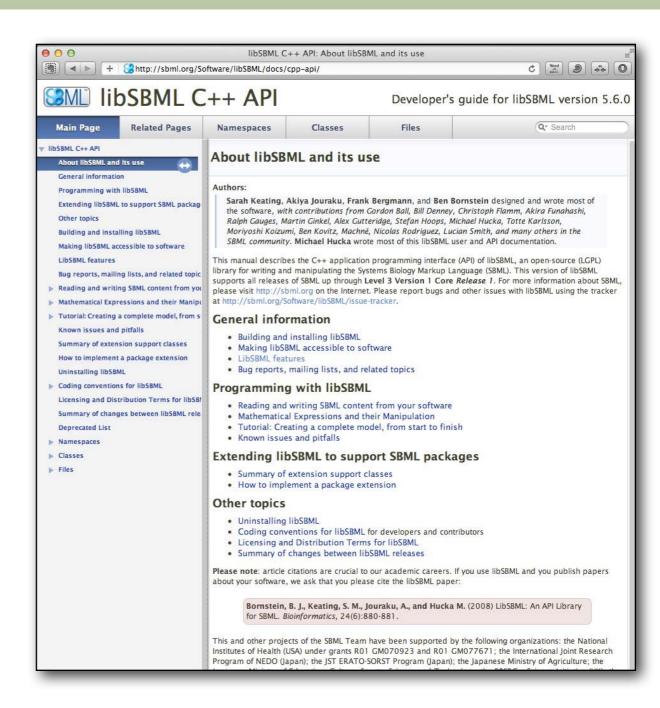
Written in portable C++

Runs on Linux, Mac, Windows

APIs for C, C++, C#, Java, Octave, Perl, Python, R, Ruby, MATLAB

Well documented API

Open-source (LGPL)



http://sbml.org/Software/libSBML

JSBML

User Guide for JSBML

Version: 1.0.0 (preview)

Authors:

Andreas Dräger ^a Alexander Dörr ^a Florian Mittag ^a Nicolas Rodriguez ^b Clemens Wrzodek ^a Michael Hucka ^c

Marine Dumousseau ^b Finja Büchel ^a

Principal investigators:

Nicolas Le Novère b

Andreas Zella

Michael Hucka^c

Institutional affiliations:

^a Center for Bioinformatics Tuebingen University of Tuebingen Tübingen, Germany

^b European Bioinformatics Institute Wellcome Trust Genome Campus Hinxton, Cambridge, UK

^c Computing and Mathematical Sciences California Insitute of Technology Pasadena, CA, USA

SBML (the Systems Biology Markup Language) is an XML-based format for storing and exchanging computational descriptions of biological processes. To read, write, manipulate, and perform higher-level operations on SBML files and data streams, software applications need to map SBML entities to suitable software objects. JSBML provides a pure Java library for this purpose. It supports all Levels and Versions of SBML, and provides many powerful features, including facilities to help migrate from the use of libSBML (a popular library for SBML that is not written in Java).

This document provides an introduction to JSBML and its use. It is aimed at both developers writing new Java-based applications as well as those who want to adapt libSBML-based applications to using JSBML. This user guide is a companion to the JSBML API documentation.

The JSBML home page is http://sbml.org/Software/JSBML/.



Pure Java implementation

API is compatible with libSBML but more Java-like

Functionality is subset of libSBML

Open source (LGPL)

http://sbml.org/Software/JSBML

What is the SBML Test Suite?

System for testing SBML support in software

- Currently aimed at simulators (easiest to assess)
- Extensible architecture—easy to add more test cases

Components:

- Test models + simulation run parameters + expected results
 - Each case is labeled with tags that indicate tested features
- Online assessment system
- Standalone test runner run an application through all tests
- Online database of test results

What is the SBML Test Suite?

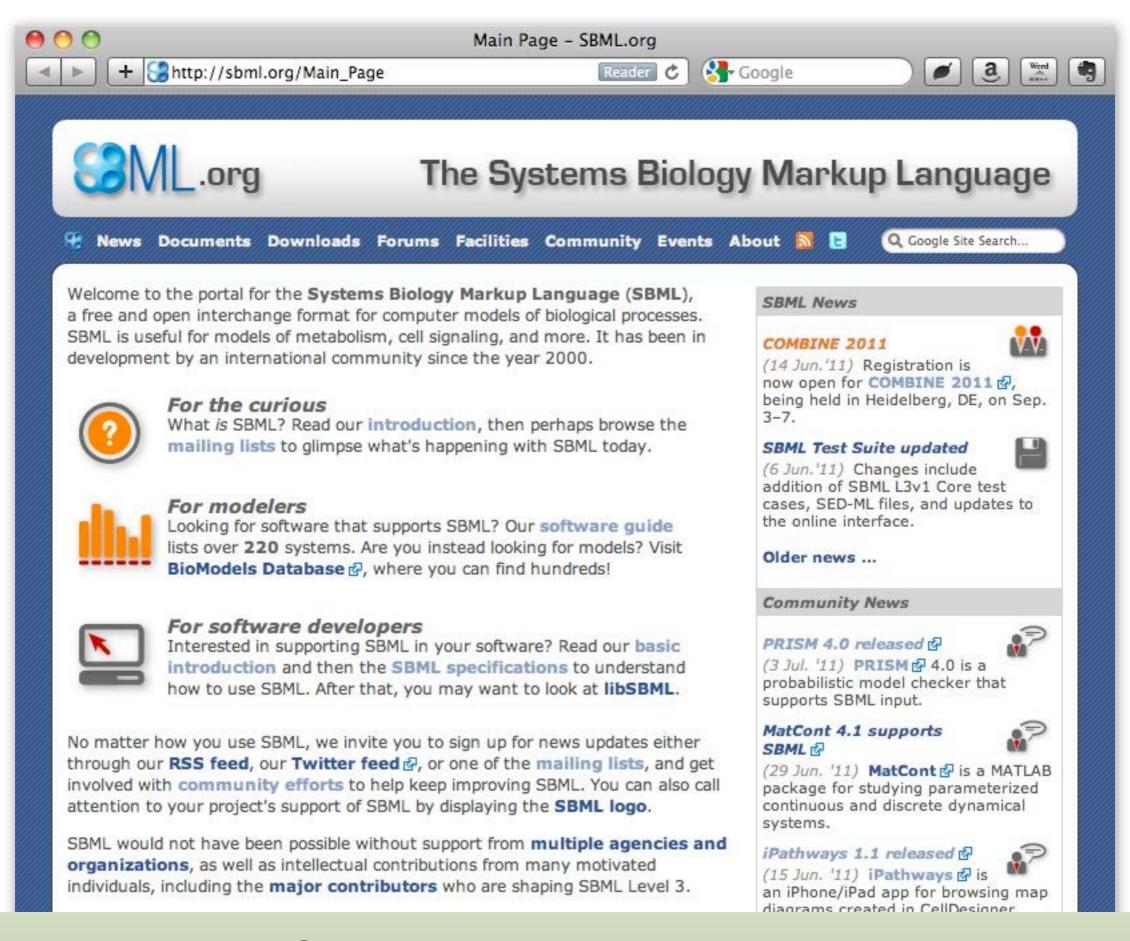
System for testing SBML support in software

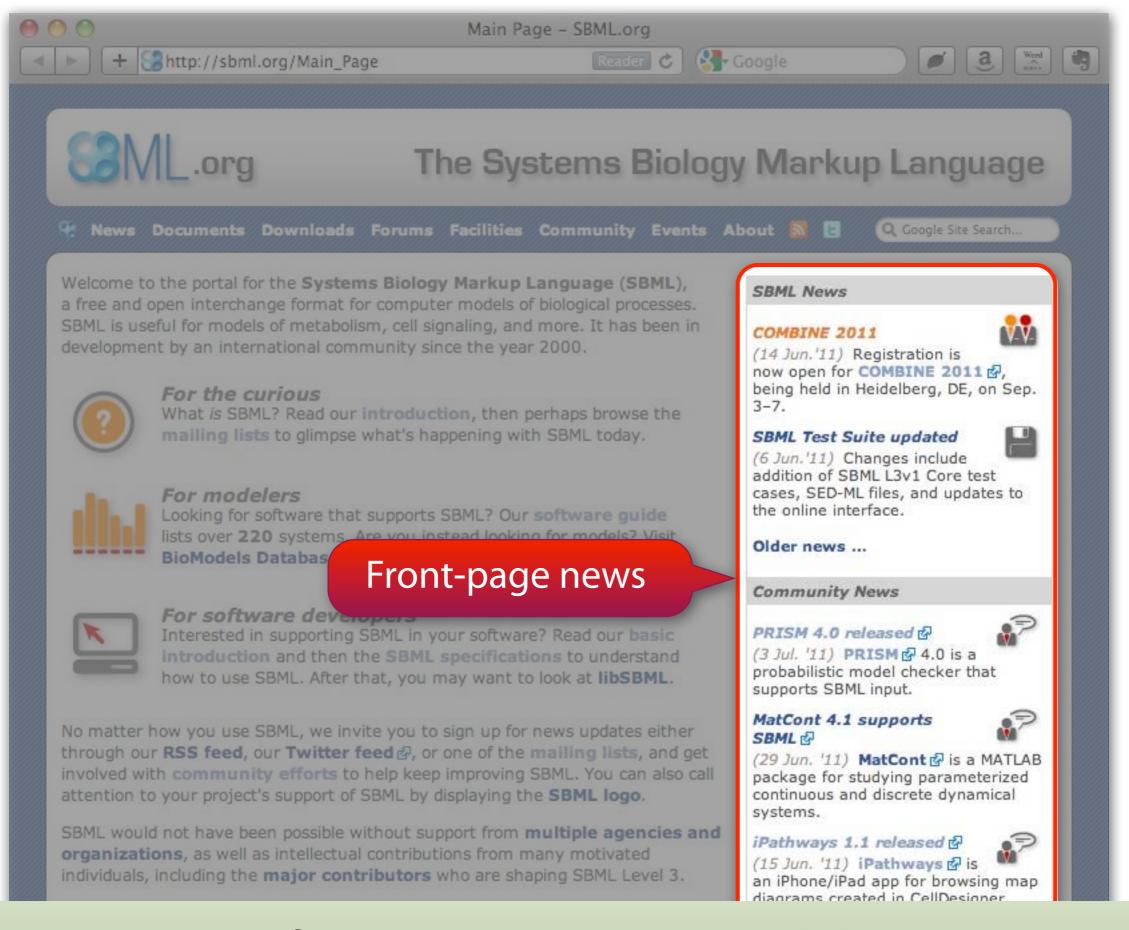
- Currently aimed at simulators (easiest to assess)
- Extensible architecture—easy to add more test cases

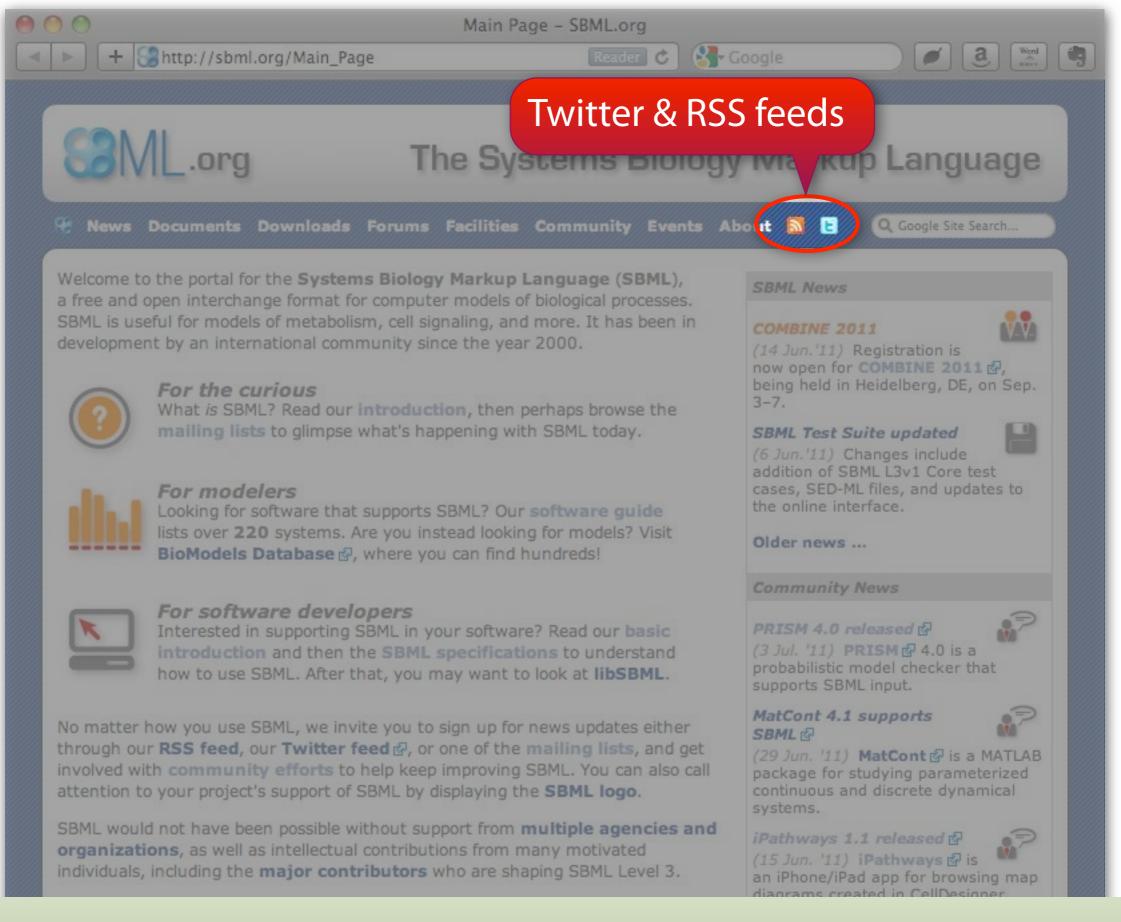
Components:

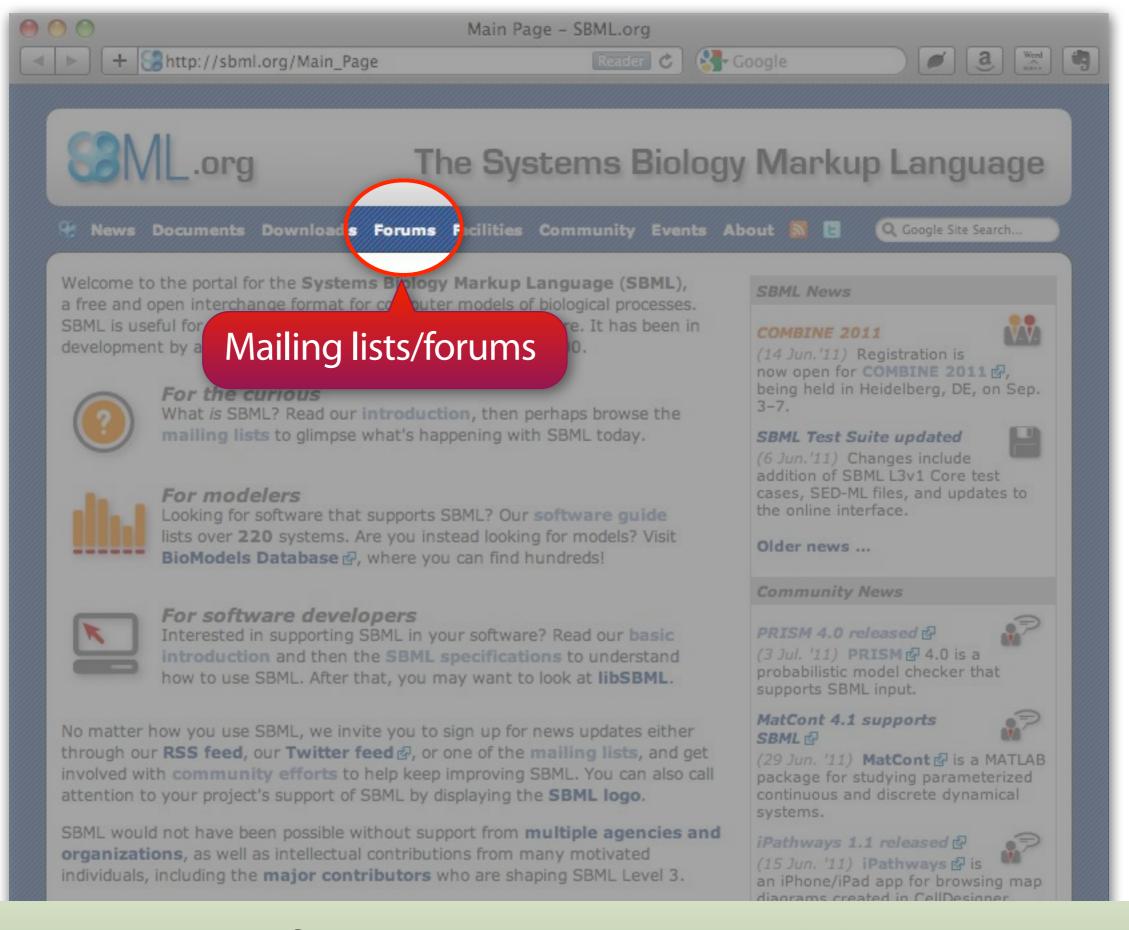
- Test models + simulation run parameters + expected results
 - Each case is labeled with tags that indicate tested features
- Online assessment system
- Standalone test runner run an application through all tests
- Online database of test results

http://sbml.org/Facilities/Online_SBML_Test_Suite









General background and motivations

Core features of SBML

A few additional details about SBML

Packages in SBML Level 3

A selection of resources for the SBML-oriented modeler

Closing

SBML Team: Mike Hucka, Sarah Keating, Frank Bergmann, Lucian Smith, Andrew Finney, Herbert Sauro, Hamid Bolouri, Ben Bornstein, Bruce Shapiro, Akira Funahashi, Akiya Juraku, Ben Kovitz

SBML Editors: Mike Hucka, Nicolas Le Novère, Sarah Keating, Frank Bergmann, Lucian Smith, Chris Myers, Stefan Hoops, Sven Sahle, James Schaff, Darren Wilkinson

BioModels DB: Nicolas Le Novère, <u>Henning Hermjakob</u>, <u>Camille Laibe</u>, Chen Li, Lukas Endler, <u>Nico Rodriguez</u>, Marco Donizelli, <u>Viji Chelliah</u>, Mélanie Courtot, Harish Dharuri

Original Pl's: John C. Doyle, Hiroaki Kitano

And a huge thanks to many others in the COMBINE community



This work was made possible thanks to a great community

SBML http://sbml.org

BioModels Database http://biomodels.net/biomodels

MIRIAM http://biomodels.net/miriam

identifiers.org http://identifiers.org

SED-ML http://biomodels.net/sed-ml

SBO http://biomodels.net/sbo

SBGN http://sbgn.org

COMBINE http://co.mbine.org

I'd like your feedback! You can use this anonymous form:

http://tinyurl.com/mhuckafeedback