SBML, BioModels.net, and SBGN

Michael Hucka

Co-director—Biological Network Modeling Center (BNMC), Beckman Institute
Senior Research Fellow—Control and Dynamical Systems

California Institute of Technology
Pasadena, California, USA
SBML background
SBML background

- Conviction that computational modeling becoming crucial
- Mechanistic modeling and dynamical simulation enables quantitative hypothesis testing
SBML background

- Conviction that computational modeling becoming crucial
- Mechanistic modeling and dynamical simulation enables **quantitative hypothesis testing**
- Not a new idea—dates to 1940’s if not earlier
SBML background

- Conviction that computational modeling becoming crucial
  - Mechanistic modeling and dynamical simulation enables quantitative hypothesis testing
- Not a new idea—dates to 1940’s if not earlier
- Today software tool support is better than ever
Specialized software tools for computational modeling in biology

- > 100 available
- Range of capabilities
  - Editing/creating models
  - Simulating/analyzing
  - Visualizing
  - Databasing
Specialized software tools for computational modeling in biology

- > 100 available
- Range of capabilities
  - Editing/creating models
  - Simulating/analyzing
  - Visualizing
  - Databasing

CellDesigner
Specialized software tools for computational modeling in biology

- > 100 available
- Range of capabilities
  - Editing/creating models
  - Simulating/analyzing
  - Visualizing
  - Databasing

JDesigner
Specialized software tools for computational modeling in biology

- > 100 available
- Range of capabilities
  - Editing/creating models
  - Simulating/analyzing
  - Visualizing
  - Databasing
Specialized software tools for computational modeling in biology

- > 100 available
- Range of capabilities
- Editing/creating models
- Simulating/analyzing
- Visualizing
- Databasing

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Name</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ma→Mi</td>
<td>MPF Inactivation</td>
<td>Mass Action</td>
</tr>
<tr>
<td>Mi→Ma</td>
<td>MPF activation</td>
<td>l_c*Mi</td>
</tr>
<tr>
<td>Ca→Cl</td>
<td>Cdc25 inactivation</td>
<td>Michaelis-Menten</td>
</tr>
<tr>
<td>Cl→Ca</td>
<td>Cdc25 activation</td>
<td>l_c*Cl</td>
</tr>
<tr>
<td>Wa→Wt</td>
<td>Wee1 inactivation</td>
<td>Michaelis-Menten</td>
</tr>
<tr>
<td>Wt→Wt</td>
<td>Wee1 activation</td>
<td>l_w*Wt</td>
</tr>
<tr>
<td>L→L2</td>
<td>Labelled inactive MFP affected by Cdc25</td>
<td>Local</td>
</tr>
<tr>
<td></td>
<td>Labelled inactive MFP affected by Wee1</td>
<td></td>
</tr>
<tr>
<td>kc</td>
<td>Species</td>
<td>vcp*C_m</td>
</tr>
<tr>
<td>kw</td>
<td>Species</td>
<td>vwp*Wm+</td>
</tr>
<tr>
<td>vcp_</td>
<td>Species</td>
<td>vcp*Cdc</td>
</tr>
<tr>
<td>vcpq_</td>
<td>Species</td>
<td>vcp*Cdc</td>
</tr>
<tr>
<td>vcpp_</td>
<td>Species</td>
<td>vcpp*Cdc</td>
</tr>
<tr>
<td>vcppp_</td>
<td>Species</td>
<td>vcpp*Cdc</td>
</tr>
<tr>
<td>vwp_</td>
<td>Species</td>
<td>vwp*Cm</td>
</tr>
<tr>
<td>vwpw_</td>
<td>Species</td>
<td>vwp*Cdc</td>
</tr>
<tr>
<td>kmc_</td>
<td>Species</td>
<td>kmc*Cdc</td>
</tr>
<tr>
<td>kmr_</td>
<td>Species</td>
<td>kmr*Cdc</td>
</tr>
<tr>
<td>kmw_</td>
<td>Species</td>
<td>kmw*Cdc</td>
</tr>
<tr>
<td>kmwr_</td>
<td>Species</td>
<td>kmwr*Cdc</td>
</tr>
<tr>
<td>wc_</td>
<td>Species</td>
<td>wc*Cdc2</td>
</tr>
<tr>
<td>ww_</td>
<td>Species</td>
<td>wc*Cdc2</td>
</tr>
<tr>
<td>Cdc25Total_</td>
<td></td>
<td>Species Cdc25Ti</td>
</tr>
</tbody>
</table>
Specialized software tools for computational modeling in biology

- > 100 available
- Range of capabilities
  - Editing/creating models
  - Simulating/analyzing
  - Visualizing
  - Databasing

SBML Model Integration Server

A web interface to the SBML_odeSolver pro

This server will integrate a valid SBML model. At this stage the web service is experimental!!!

Instructions

- Please upload a valid SBML Model
- Please provide a valid email address (you will be notified by email)
- Your model will be validated prior to integration using the function
- If validation errors occur, please correct them and resubmit your model

Choose File: no file selected

Your Email Address

<table>
<thead>
<tr>
<th>Simulation Time (use scientific notation e.g. 1e7 for 10000000)</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Print Step</td>
<td>100</td>
</tr>
<tr>
<td>Absolute Error</td>
<td>1e-09</td>
</tr>
<tr>
<td>Relative Error</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
Specialized software tools for computational modeling in biology

- > 100 available
- Range of capabilities
  - Editing/creating models
  - Simulating/analyzing
  - Visualizing
  - Databasing

Virtual Cell
Many common types of model representation frameworks

- (Continuous) nonlinear differential equations
- (Discrete) stochastic systems
- Boolean networks
- Bayesian networks
- Petri Nets
- others...
Ability to exchange models is critical
Ability to exchange models is critical

- Simply publishing equations is not enough
- Don’t want to transcribe equations from papers
- You want a common file format
Ability to exchange models is critical

- Simply publishing equations is not enough
  - Don’t want to transcribe equations from papers
  - You want a common file format
- Not a new idea—seems obvious
  - Still, a format hadn’t existed before year 2000
    - Each tool had its own unique proprietary format
    - (Fewer tools too)
JST ERATO
Kitano Project

One initial component: get 8–10 software systems interacting
SBML = Systems Biology Markup Language
SBML =
Systems Biology Markup Language

- Machine-readable format for computational models
SBML =
Systems Biology Markup Language

- Machine-readable format for computational models
- Suitable for reaction networks
- Arbitrary rate functions

\[
2 \text{ A} + \text{ B} \rightarrow \text{ C} \\
\text{ C} \leftrightarrow \text{ D} + \text{ F} \\
\ldots
\]
SBML = Systems Biology Markup Language

- Machine-readable format for computational models
- Suitable for reaction networks
  - Arbitrary rate functions
- Models can also include
  - Compartments
  - Mathematical “extras”

\[2 \text{A} + \text{B} \rightarrow \text{C} \]
\[
\text{C} \leftrightarrow \text{D} + \text{F}
\]
...
SBML = Systems Biology Markup Language

- Machine-readable format for computational models
- Suitable for reaction networks
  - Arbitrary rate functions
- Models can also include
  - Compartments
  - Mathematical “extras”
- Declarative, not procedural

\[
\begin{align*}
2A + B & \rightarrow C \\
C & \leftrightarrow D + F \\
& \ldots
\end{align*}
\]
SBML is an XML format

- SBML defined using UML and XML Schema
- Targeted at XML, but mostly independent of it
- A **lingua franca** for software, not humans
- Think HTML
Where is SBML today?
Now the *de facto* standard

The Systems Biology Markup Language (SBML) is a computer-readable format for representing models of biochemical reaction networks. SBML is applicable to metabolic networks, signaling pathways, regulatory networks, and many others.

**Internationally Supported and Widely Used**

SBML has been evolving since mid-2000 through the efforts of an international group of software developers and users. Today, SBML is supported by over 100 software systems, including:

- **Supported by >100 systems**
- **Accepted by journals**
- **Nature**
- **PLoS**
- **BMC**
- **Used in textbooks & courses**
A community of modelers and software developers

- **sbml-discuss** *(275+ people)*, **sbml-announce**
- **Annual SBML Forum** meeting *(at ICSB)*
- **Annual SBML Hackathon**
A community of modelers and software developers

- **sbml-discuss** (275+ people), **sbml-announce**
- Annual **SBML Forum meeting** (at ICSB)
- Annual **SBML Hackathon**
Support by SBML Team

- Writing grants for core development
- Writing infrastructure software
  - libSBML
  - MathSBML, SBMLToolbox
- Maintaining web & mailing list resources
- Organizing workshops & other events
Systems Biology Markup Language (SBML) Level 2: Structures and Facilities for Model Definitions

Andrew Finney
afinney@sbml.org
Physionomics PLC
Magdalen Centre
Oxford Science Park
Oxford, OX1 4GA, UK

Michael Hucka
mhucka@sbml.org
Biological Network Modeling Center
Beckman Institute, Mail Code 139-74
California Institute of Technology
Pasadena, CA 91125, USA

Nicolas Le Novère
lenov@ebi.ac.uk
European Bioinformatics Institute
Wellcome Trust Genome Campus, Hinxton
Cambridge, CB10 1SD, UK

SBML Level 2, Version 2, Revision 1
26 September 2006

Corrections and other revisions of this SBML language specification may appear over time.
Notifications of revisions are broadcast on the mailing list sbml-announce@caltech.edu

The latest revision of the SBML Level 2 Version 2 specification is available at

This revision of the SBML Level 2 Version 2 specification is available at
Examples of significant changes

- Many clarifications
- Simplification to the unit system
- Addition of species types, compartment types
- Addition of “constraints”
- Support for the Systems Biology Ontology (SBO)
- Recommended standard format for annotations
- “Revisions” process for handling errata
What lies ahead?
Revised SBML governance & development process

- Borrow ideas from W3C & other organizations
- Implement a better-defined, regimented process
  - Calls for proposals, etc.
  - Voting, etc.
- Have an architectural board to steer development
- Have more SBML Editors
  - Elect SBML Editors for limited terms
SBML Level 3
SBML Level 3

- Modular language extensions
- Core expected to be based “mostly” on Level 2 Version 2
SBML Level 3

- Modular language extensions
  - Core expected to be based “mostly” on Level 2 Version 2
- Layered on top of core: feature sets for—
  - Diagram storage
  - Multicomponent species
  - Models composed of submodels
  - Arrays and/or sets of components
  - Spatial geometry
  - Other capabilities
Full SBML Test Suite

- Allows developers to test implementation of SBML support
- Critical for improving software interoperability
- Currently have a partial “SBML semantic test suite”
- Needs further work to—
  - Complete coverage of SBML features
  - Improve ease of use
  - Update for Level 2 Version 2 and Level 3
  - Add web system for reporting results, comparisons, etc.
Got models?
Got models?

- Early FAQ in SBML: is there a database of models?
Got models?

- Early FAQ in SBML: is there a database of models?
- Raw model isn’t very “searchable”
Got models?

- Early FAQ in SBML: is there a database of models?
- Raw model isn’t very “searchable”
- SBML provides syntax
- But SBML doesn’t—
  - encode meaning
  - regulate use of names
Got models?

- Early FAQ in SBML: is there a database of models?
- Raw model isn’t very “searchable”
- SBML provides syntax
- But SBML doesn’t—
  - encode meaning
  - regulate use of names
Got models?

- Early FAQ in SBML: is there a database of models?
- Raw model isn’t very “searchable”
- SBML provides syntax
- But SBML doesn’t—
  - encode meaning
  - regulate use of names
Got models?

- Early FAQ in SBML: is there a database of models?
- Raw model isn’t very “searchable”
- SBML provides syntax
- But SBML doesn’t—
  - encode meaning
  - regulate use of names
Got models?

- Early FAQ in SBML: is there a database of models?
- Raw model isn’t very “searchable”
- SBML provides syntax
- But SBML doesn’t—
  - encode meaning
  - regulate use of names

```xml
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level3/version1/core"
xmlns:math="http://www.w3.org/1998/Math/MathML">
  <model>
    <listOfCompartment>
      <compartment id="cell"/>
    </listOfCompartment>
    <listOfSpecies>
      <species id="X0" compartment="cell" initialConcentration="0.01"/>
      <species id="X1" compartment="cell" initialConcentration="0.01"/>
      <species id="T" compartment="cell" initialConcentration="0.01"/>
      <species id="S1" compartment="cell" initialConcentration="0.01"/>
      <species id="S2" compartment="cell" initialConcentration="0.01"/>
    </listOfSpecies>
    <listOfParameters>
      <parameter id="Keq" value="2.5"/>
    </listOfParameters>
    <listOfRules>
      <assignmentRule variable="S1">
        <math xmlns="http://www.w3.org/1998/Math/MathML">
          <apply>
            <divide/>
            <ci> T </ci>
            <apply>
              <plus/>
              <cn> 1 </cn>
              <ci> Keq </ci>
            </apply>
          </apply>
        </math>
      </assignmentRule>
    </listOfRules>
  </model>
</sbml>
```
BioModels.net consortium

- International collaboration to develop:
  1. A public database of **curated, annotated** models: **BioModels Database**
  2. Guidelines for curation and annotation of models: **MIRIAM**
  3. Ontology suited to computational models: **SBO**

- Main collaborators
  - Nicolas Le Novère’s Computational Neurobiology group at EBI (UK)
  - Hucka and SBML Team at Caltech (USA) and U. Hertfordshire (UK)
  - Herbert Sauro’s group at Keck Graduate Institute
  - Hans Westerhoff & Jacky Snoep’s JWS Online (ZA and UK)
  - Hiroaki Kitano’s Systems Biology Institute
BioModels Database
http://www.ebi.ac.uk/biomodels

- Stores & serves quantitative models of bio. interest
- Free, public resource
- Models must be described in peer-reviewed publication(s)
- Imports models in SBML & CellML formats
- Exports in SBML, CellML, SciLab, XPP and BioPAX
Models are curated
Models are curated

- Human curators check correspondence to publication
Models are curated

- Human curators check correspondence to publication
- Human curators add annotations
Models are curated

- Human curators check correspondence to publication
- Human curators add annotations
- General info about model, author, publication, etc.
Models are curated

- Human curators check correspondence to publication
- Human curators add annotations
- General info about model, author, publication, etc.
- References to external data resources: UniProt, KEGG, NCBI, Gene Ontology, ChEBI, BIND, Reactome
Features of BioModels Database

- **True database**: can search model content & annotations
  - SBML XML stored in XINDICE; annotations in an SQL database
  - 70+ curated, dynamical models
    - E.g.: Tyson yeast cell cycle models, Elowitz E. coli repressilator, Teusink et al. yeast glycolysis, Rohwer et al. E. coli glucose transport
  - 26 non-curated models (e.g., models lacking kinetics, such as FBA)
- Model sources: us, Nature/EMBO *Molecular Systems Biology*, repositories such as JWS Online and CellML, individual researchers
“Minimal info. requested in the annotation of biochemical models”

Proposed guidelines for basic annotation of models

- Reference correspondence, e.g.,
  - Be encoded in a public, standardized format (SBML, CellML, etc.)
  - Must be instantiated in a simulation & all quantitative attributes defined

- Attribution annotation (info about model creators, source reference, etc.)

- External source annotations (linking model element to data source)

Goal: minimal common standards enabling sharing of curation effort
SBO = Systems Biology Ontology

- Occupies a space not filled by other ontologies
  - Primarily for describing rate laws and constituents
    - Classification of rate laws
      - Each term includes a mathematical function definition
    - Controlled vocabulary for the roles of reaction participants
      - E.g.: “substrate”, “catalyst”, “competitive inhibitor”, etc.
    - Controlled vocabulary for the roles of parameters in quantitative models
BioModels.net portal
Background

- No current standard for network diagrams in biology
  - No consistency —
    - Between authors
    - Between papers
    - Between publications
Background

- No current standard for network diagrams in biology
- No consistency—
  - Between authors
  - Between papers
  - Between publications
Standardization would bring benefits

- Notations standardized in electrical/electronics, software engineering, etc.
- Taught in textbooks
- Supported by software
- Automated verification
- Consistency makes it easier to read new diagrams
SBGN = Systems Biology Graphical Notation

- Goal: bring simulation/modeling community together and develop a proposal for a standard notation for some types of diagrams
  - Starting with metabolic & signaling networks
- Begun late '05 by Kitano/Le Novère/Hucka thanks to NEDO funding
- 2 SBGN workshops held so far (Feb and Oct 2006)
  - Many groups participating: Goryanin group (U. Edinburgh), BioPAX (MSKCC New York), SRI, CellML, EML (Germany), many others
- Join if you’re interested!
- Currently working towards a first proposal
Closing
The funding

- National Institute of General Medical Sciences (USA)
- JST ERATO Kitano Symbiotic Systems Project (Japan) (to 2003)
- National Science Foundation (USA)
- International Joint Research Program of NEDO (Japan)
- JST ERATO-SORST Program (Japan)
- Japanese Ministry of Agriculture
- BBSRC e-Science Initiative (UK)
- DARPA IPTO Bio-SPICE Bio-Computation Program (USA)
- Air Force Office of Scientific Research (USA)
- STRI, University of Hertfordshire (UK)
- Beckman Institute, Caltech (USA)
<table>
<thead>
<tr>
<th>SBML Team</th>
<th>BioModels DB Team</th>
<th>SBGN Team</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michael Hucka</td>
<td>Nicolas Le Novère</td>
<td>Hiroaki Kitano</td>
</tr>
<tr>
<td>Andrew Finney</td>
<td>Chen Li</td>
<td>Nicolas Le Novère</td>
</tr>
<tr>
<td>Ben Bornstein</td>
<td>Mélanie Courtot</td>
<td>Michael Hucka</td>
</tr>
<tr>
<td>Sarah Keating</td>
<td>Lu Li</td>
<td>Akira Funahashi</td>
</tr>
<tr>
<td>Bruce Shapiro</td>
<td>Camille Laibe</td>
<td></td>
</tr>
<tr>
<td>Ben Kovitz</td>
<td>Nicolas Rodriguez</td>
<td></td>
</tr>
<tr>
<td>Hamid Bolouri</td>
<td>Harish Dharuri</td>
<td></td>
</tr>
<tr>
<td>Herbert Sauro</td>
<td>Marco Donizelli</td>
<td></td>
</tr>
<tr>
<td>Maria Schilstra</td>
<td>Alexander Broicher</td>
<td></td>
</tr>
<tr>
<td>Jo Matthews</td>
<td>Arnaud Henry</td>
<td></td>
</tr>
</tbody>
</table>

A million thanks to the SBML Community too
Where to learn more

- http://sbml.org
- http://biomodels.net
- http://sbgn.org

Upcoming:
- SBML Hackathon 2007 in June at U. Newcastle, UK
- SBGN Workshop in 2007 (probably March 2007, Gosau, Austria)
- SBML Forum 2007 in Long Beach, CA, USA, Oct. 5-6 (ICSB 2007)

Thank you!