Towards unifying systems biology - using pathway data in BioPAX format for SBML simulators

Ion I. Moraru, Oliver Ruebenacker, Michael L. Blinov

Center for Cell Analysis and Modeling, University of Connecticut Health Center, Farmington, CT 06030, USA

moraru@exchange.uchc.edu, oruebenacker@exchange.uchc.edu, blinov@uchc.edu

Objective: Thousands of biochemical interactions are available from public sources in the Biological Pathways Exchange (BioPAX) format. However, the current standard for exchange of simulation-ready biological models is System Biology Markup Language (SBML). This markup language is structurally and semantically different from BioPAX. Some conversion schemes exist, using annotations and based on simple one-to-one mappings between SBML and BioPAX objects, which ignores semantic differences and therefore often leads to significant loss of information or meaning. A comprehensive modeling framework capable of representing the complex relationships between SBML and BioPAX data is needed to take full advantage of existing pathway data in kinetic modeling, thus integrating these two formats by gluing them together.

Results: Here we describe such a framework that we are developing as a part of the Virtual Cell (http://vcell.org) modeling and simulation environment. Systems Biology Linker (Sybil, http://vcell.org/biopax) is a tool for analyzing and visualizing BioPAX data, and converting them to SBML. Based on the Jena Semantic Web framework for Java, Sybil supports handling of generic RDF/OWL data (such as visualization and reasoning) as well as functions specific to handling SBML and BioPAX data. Sybil uses Systems Biology Pathway eXchange, called SBPAX, as a generic approach to integrate model-centric formats similar to SBML with pathway-centric formats similar to BioPAX. SBPAX is an OWL-based schema that serves as a glue to integrate different data formats, despite semantic differences. Sybil offers various visualization modes showing reaction networks to varying degrees of details, including displaying nodes for reactions only as well as displaying Petri nets consisting of reaction nodes and reaction participants and catalysts. Sybil also allows collapsing and exploding various parts of the network individually, for example exposing reaction participant nodes to show all their components.

From data to models

World of pathway databases

Data Source	Protein(SwissProt)	Coverage (SwissProt)	Interaction	Citation
Reactome	1229 (1194)	5% (8%)	21394	Vastrik *et al*, 2007
Panther	2997 (1670)	12% (12%)	75694	Mi *et al*, 2007
CellMap	567 (567)	2% (4%)	1195	cancer.cellmap.org
INOH	719 (711)	3% (5%)	11759	Kushida *et al*, 2006
NCI-Nature	593 (592)	2% (4%)	2900	pid.nci.nih.gov
NCI-BioCarta	936 (936)	4% (6%)	4752	pid.nci.nih.gov
KEGG	2033 (1947)	8% (13%)	11144	Kanehisa *et al*, 2004
Total	**5283 (3847)**	**21% (27%)**	**118867**	

Two worlds

Pathway data Community

- Knowledge
- Link information so that it can be found
- Ontologies
- Relationships
- Reason and query

Modeling Community

- Data
- Do the math so that it can be calculated
- Data models
- Quantities
- Simulate and fit

Two worlds – two descriptions

BioPAX

- "Biological Pathway Exchange"
- Qualitative, no kinetics
- RDF / OWL
- Hierarchy of Terms
- "Designed by librarians"

Physical Entity
  Protein, complex, small molecule, DNA, RNA
Interaction
  Conversion, control
    Transport, catalysis, biochemical reaction, modulation
  Physical entity participant
    Sequence participant

SBML

- "System Biological Markup Language"
- Quantitative information
- XML
- No term hierarchy
- "Designed by programmers"

Units
Compartments
Parameters
SpeciesTypes
Species
Reactions
  Reactants
    speciesReference
  Products
    speciesReference
  KineticLaw

RDF and OWL

- RDF stands for Resource Description Framework
- Resource = Something identifiable
- OWL (ontology) stands for usage of RDF, where restrictions on RDF properties are described by RDF statements

XML (Xtensible Markup Language)

```
<sbml>
  <ListOfReactions>
    <Reaction id="R1">
      ... stoichiometry="3" ...
    </Reaction>
    ...
  </ListOfReactions>
</sbml>
```

Notable Bio-Med ontologies

SBO = Systems Biology Ontology
BFO = Basic Formal Ontology
BioPAX = Biological Pathway Exchange
GO = Gene Ontology
OBO = Open Biomedical Ontology

Notable Markup standards

HTML = Hypertext Markup Language
XML = X-tensible Markup Language
SBML = Systems Biology Markup Language
CellML – Cellular MarkUp Language
VCML – Virtual Cell MarkUp Language

Majority of Pathway databases supports BioPAX export: Reactome, CellMap, INOH, NCI-Nature, NCI-BioCarta

Few support SBML export: Reactome, Panther. However, SBML export is limited.

What we have:

- Many pathway databases exporting BioPAX
- Many simulators importing SBML
- Automatic conversion tools rely on one-to-one mapping, which is not always correct.
- Lack of links between data in SBML and BioPAX

What we want:

- A smarter conversion from BioPAX to SBML
- A way to establish links between SBML and BioPAX data and store these links together with data
- Better visualization of pathways based on hierarchical structure of BioPAX.

Conversion problem

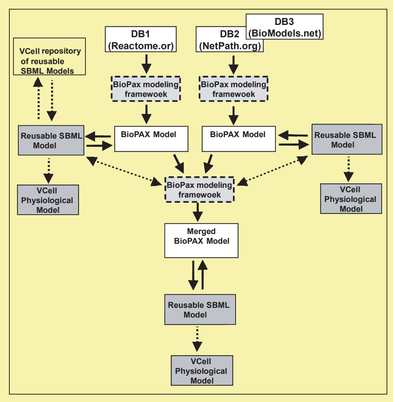
- Sometimes an entity in BioPAX corresponds to multiple species types in SBML (e.g. a protein in varying phosphorylation states)
- Sometimes a species type in SBML corresponds to multiple entities in BioPAX (e.g. SNIPs, or generic participants in transport, translation and transcription)

System Biology Linker (SyBiL)

Integrating SBML and BioPAX by:

- Flexible semi-automatic conversion
- Storing relationships between SBML and BioPAX data
- Building a joint repository (SBPAX) of SBML and BioPAX data

Sybil workflow

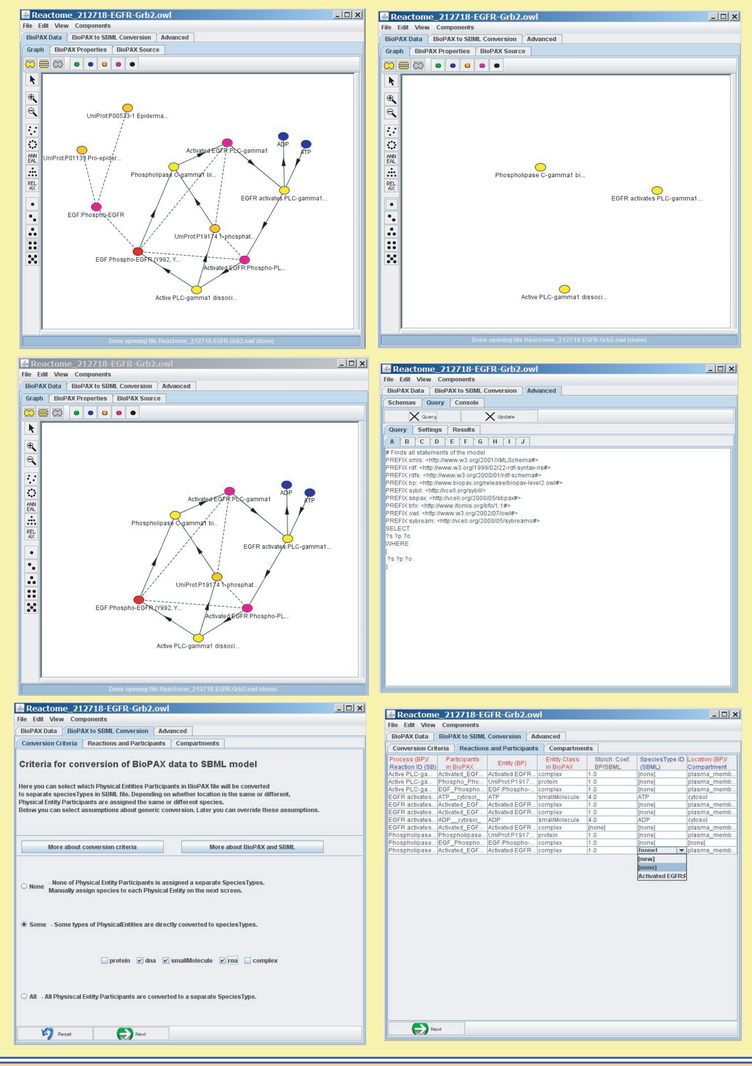


Systems Biology PathwayExchange (SBPAX)

SBPAX is an OWL-based ontology

Used for conversion and mapping between SBML and BioPAX

- Describes a common subset of SBML and BioPAX
- Flexible enough to allow import from both
- Distinction between an object and a model of the object



Sybil capabilities

Implemented:

- BioPAX visualization
- Conversion BioPAX -> SBPAX -> SBML

Planned:

- SBPAX and SBML visualization
- Conversion SBML -> SBPAX -> BioPAX

BioPAX@VCell is available at http://www.vcell.org/biopax