

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

Oliver Ruebenacker¹, Ion I. Moraru¹, and Michael L. Blinov¹

Short Abstract —Thousands of biochemical interactions are available for free download 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. Here we describe such a framework that we are developing as a part of the Virtual Cell modeling and simulation environment.

Keywords — kinetic modeling; pathway databases; SBML; BioPAX; systems biology; protein-protein interactions.

Computational biology has two thriving communities each with a large pool of data and with their own established exchange standards. Quantitative models used for simulations are typically published or exchanged in Systems Biology Markup Language (SBML), [1] or CellML [2] format, while many molecular interaction and pathway databases use the Biological Pathways eXchange format (BioPAX), [3]. There is an obvious benefit of being able to exchange data between these two communities, which has motivated the creation of a number of conversion and annotation schemes [4]. However, integrating data from both communities is difficult, due in particular to semantic differences between the two formats. Information can not be ported whenever one format does not support its representation (most notably, BioPAX's lack of support for kinetic modeling, and SBML's lack of support for relationships not of direct concern for modeling). But even if the representation of certain information is supported by both formats (such as reaction participants, catalysis, products, etc.) we often lose information, or its meaning, or its correct interpretation. To prevent these kinds of losses, it is necessary to go beyond the current ad-hoc use of annotations for conversion between formats, and create a structured framework to store knowledge about the variety of relationships between SBML and BioPAX objects.

Here we introduce such a framework based on Systems Biology Pathway eXchange, called SBPAX, as a generic approach to integrate model-centric formats similar to SBML or CellML with pathway-centric formats similar to BioPAX. SBPAX is an OWL-based [5] schema that serves as a glue to integrate different data formats, despite semantic differences. SBPAX is based on an ontology which includes compatible terms for all relevant SBML or BioPAX terms, which means that from a suitable SBPAX repository, information can be exported to SBML or BioPAX without loss. SBPAX can be used as an extension or annotation scheme allowing the target languages to support the representation of all relevant information.

To implement SBPAX, we designed Sybil [6], a tool for analyzing and visualizing OWL data. Based on the Jena Semantic Web framework for Java, Sybil supports handling of generic RDF/OWL data (such as visualization, SPARQL queries and RDF/OWL reasoning) as well as functions specific to handling SBML, BioPAX and SBPAX data. Currently, Sybil fully supports BioPAX Level 2, but also offers some support for other ontologies, and is expected to fully support all upcoming versions of BioPAX. 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. Most of the work of building a SBPAX repository, imported SBML or BioPAX data, is done by SYBREAM, a custom reasoning engine designed to complement and validate SBPAX data.

REFERENCES

- [1] Hucka M, *et al.* (2003) *Bioinformatics* **19**, 524-531.
- [2] Lloyd CM, *et al.* (2004) *Prog Biophys Mol Biol* **85**, 433-50.
- [3] Luciano JS. (2005) *Drug Discov Today* **10**, 937-42.
- [4] Zinoviev A, *et al.* (2008) *Bioinformatics* **24**, 876-7.
- [5] Aitken S., *et al.* (2008) *Bioinformatics* **24**, 443-4.
- [6] Ruebenacker O, *et al.* (2007) *Proc. of the 2007 IEEE BIBE*: 339-348.

This work has been supported by NIH grants RR022232 and GM076570.

¹Center for Cell Analysis and Modeling, University of Connecticut Health Center, Farmington, CT. E-mail: ruebenacker@uchc.edu, moraru@uchc.edu, blinov@uchc.edu.