Modeling using BioPax standard

Oliver Ruebenacker[1] and Michael L. Blinov[1]

The key issue that should be resolved when making reusable modeling modules is species recognition: are species S1 of model 1 and species S2 of model 2 identical and should they be mapped onto the same species in a merged model? When merging models, species recognition should be automated. We address this issue by providing a modeling process compatible with Biological Pathways Exchange (BioPax) standard [1].

***Keywords* — Model, Pathway data, BioPax, SBML, the Virtual Cell.**

The main format currently used for encoding of pathway models is Systems Biology Markup Language (SBML) [2], which is designed mainly to enable the exchange of biochemical networks between different software packages with little or no human intervention. SBML model contains data necessary for simulation, such as species, interactions among species, and kinetic laws for these interactions. The essential feature of such simulation-centric XML standards (SBML, CellML [3], etc) is no hierarchy of different types of species or different types of reactions. There is a list of things of the same kind, called species, and a list of things of the same kind, called reactions. As long as the data is from a single source and small enough to be tweaked by hand, flat and simple formats are most welcome. The user knows what each symbol means and therefore software do not need to. But this is changing: as projects grow, the need is growing to combine data from different sources and to process them by software sophisticated enough to know that there is some sort of difference between a complex of proteins and a small molecule. BioPax standard provides a pathway exchange format that aims to facilitate sharing of pathway information between databases and users. BioPaX is based on OWL (Web Ontology Language [4]) that is designed for use by applications that need to process the content of information instead of just presenting information to humans. OWL provides vocabulary along with a formal semantics. BioPax concepts, unlike XML concepts, have relationships to each other that can be processed automatically. An automatic reasoner can infer that if B is a kind of A, then B has all properties that A has. These relationships between different concepts are the key to merging or linking different sets of information from different sources. Each element of BioPax file is linked to an originating biological database. A unique biological identification for each element of the model makes BioPax standard a recipe for reusable modeling modules.

Currently, several online resources provide information about pathways in BioPax format: NetPath – Signal Transduction Pathways [5], Reactome - a curated knowledgebase of biological pathways [6], and some others, with more resources aiming at BioPax representation. These resources store complete pathways, pathways participants, and individual interactions. Due to principal differences between SBML and BioPax, there are no one-to-one translators from BioPax to SBML. We have implemented import of pathway data in BioPax standard into the Virtual Cell modeling framework [7], using Jena (Java application programming interface that provides support for Resource Description Framework). The imported data forms a model skeleton where each species being automatically related to some database entry. This model skeleton may lack simulation-related information (such as concentrations, kinetic laws etc) and may have a lot of abundant information (organisms, different names, linking species to variety of databases, etc). We call it a meta-model – by analogy with meta-tags in HTML that provide invisible to a reader but searchable information. A meta-model can be easily visualized, as different BioPax objects (proteins, small molecules, complexes etc) have different representations (e.g. colors). Each object is linked to biological information from public databases. A modeler may need to add some information to convert a meta-model into a computable Virtual Cell model, but this additional information can be stored within BioPax format. Meta-models are stored in a Virtual Cell database, with possibility to export into other BioPax-supporting databases. Several meta-models can be automatically merged into a larger meta-model. The merged model can be compactly visualized as a set of modules, where all elements of the same meta-model are compressed into a single node.

Acknowledgements: This work was funded by NIH grant for Technology Center for Network and Pathways.

[1]Center for Cell Analysis and Modeling, University of Connecticut Health Center. E-mail: oruebenacker@uchc.edu , blinov@uchc.edu.

REFERENCES

[1] BioPax website, http://biopax.org
[2] Hucka M, et al. (2003) The Systems Biology Markup Language (SBML): A Medium for Representation and Exchange of Biochemical Network Models. *Bioinformatics* **19(4)**, 524-531, http://www.sbml.org
[3] Lloyd CM et al. (2004) CellML: its future, present and past. *Progress in Biophys. and Mol. Biol.* **85(2-3)**, 433-450, http://cellml.org
[4] Jiang K, Nash C. (2005) Ontology-based aggregation of biological pathway datasets. *Conf Proc IEEE Eng Med Biol Soc*. **7**, 7742-5.
[5] Netpath website, http://netpath.org
[6] Vastrik I, et al. (2007) Reactome: a knowledgebase of biological pathways and processes. *Genome Biol*. Mar 16; **8(3).**
[7] Slepchenko BM, et al. (2003) Quantitative cell biology with the Virtual Cell. *Trends Cell Biol*. **13(11)**, 570-6.