BIOINFORMATICS APPLICATIONS NOTE Vol. 27 no. 24 2011, pages 3437–3438
doi:10.1093/bioinformatics/btr586

Systems biology

Advance Access publication October 21, 2011

BioPAX support in CellDesigner

Huaiyu Mi[1,*], Anushya Muruganujan[1], Emek Demir[2], Yukiko Matsuoka[3], Akira Funahashi[4], Hiroaki Kitano[3] and Paul D. Thomas[1]

[1]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, [2]Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA, [3]Systems Biology Institute, Tokyo and [4]Department of Biosciences and Informatics, Keio University, Hiyoshi, Kouhoku-ku, Yokohama, Japan

Associate Editor: Martin Bishop

ABSTRACT

Motivation: BioPAX is a standard language for representing and exchanging models of biological processes at the molecular and cellular levels. It is widely used by different pathway databases and genomics data analysis software. Currently, the primary source of BioPAX data is direct exports from the curated pathway databases. It is still uncommon for wet-lab biologists to share and exchange pathway knowledge using BioPAX. Instead, pathways are usually represented as informal diagrams in the literature. In order to encourage formal representation of pathways, we describe a software package that allows users to create pathway diagrams using CellDesigner, a user-friendly graphical pathway-editing tool and save the pathway data in BioPAX Level 3 format.

Availability: The plug-in is freely available and can be downloaded at ftp://ftp.pantherdb.org/CellDesigner/plugins/BioPAX/

Contact: huaiyumi@usc.edu

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Received on May 13, 2011; revised on October 14, 2011; accepted on October 17, 2011

1 INTRODUCTION

Recent efforts to standardize pathway data have greatly facilitated the interpretation, exchange and integration of pathway knowledge in a rigorous and unambiguous way. The efforts have led to the creation of several widely accepted complementary community standards—Systems Biology Markup Language (SBML) (Hucka *et al.*, 2003), Biological Pathway Exchange (BioPAX) (Demir *et al.*, 2010) and Systems Biology Graphical Notation (SBGN) (Le Novère *et al.*, 2009). Both SBML and BioPAX are machine-readable formats for representing cellular processes. The former is concentrated on mathematical modeling, while the latter is focused on qualitative pathway knowledge. SBGN is a standard for graphical representation of pathways and can be used for visualizing both SBML and BioPAX models. BioPAX Level 3 covers a large spectrum of qualitative information in the literature including metabolic and signaling pathways as well as molecular and genetic interactions. As a result, BioPAX is currently supported by >40 different pathway database resources (http://www.pathguide.org/), including MetaCyc (Caspi *et al.*, 2010), PANTHER (Mi *et al.*, 2010), PID (Schaefer *et al.*, 2009) and Reactome (Croft *et al.*, 2011). Most of these resources generate BioPAX-compliant exchange files from the pathway data that are curated by highly trained curators using specialized curation software. It is still rare for wet-lab biologists to create and exchange pathway data using such standards. This creates a bottleneck for pathway knowledge accumulation as the database groups can only curate a fraction of the enormous amount of knowledge that is being generated. Extending the usage of formal machine-readable models, such as BioPAX, to the wider biological community would significantly alleviate this problem and help pathway databases to cope up with the load (Mi and Thomas, 2011). Furthermore, wet-lab biologists are the end-users of the pathway databases, and thus the enrichment of pathway database content will ultimately benefit their research. The major obstacle so far is the lack of software tools that can allow biologists to easily transcribe the pathway knowledge to BioPAX format.

Cytoscape BioPAX plug-in allows users to load and graphically render BioPAX files for visualization. It has been used by a number of software to import or export BioPAX files. The plug-in does not support BioPAX Level 3 yet. CellDesigner (Funahashi *et al.*, 2008) is a graphical, intuitive pathway-editing software, which uses controlled graphical notations for visual representation of the pathway and supports both SBML and SBGN standard (Hucka *et al.*, 2003; Le Novère *et al.*, 2009). It is used in the research community for generating pathway diagrams with controlled graphical notations for both wet-lab and systems biologists. Both CellDesigner and BioPAX describes pathway in terms of biochemical reaction and process, which is equivalent to SBGN Process Description (SBGN-PD) standard (Le Novère *et al.*, 2009). The data structures in CellDesigner and BioPAX are similar but not identical, and we have developed a mapping between the two. Here, we describe a software package that allows users to draw or open a pathway diagram in CellDesigner, and save the data accurately in BioPAX format. Thus, the software provides the connection between SBML, SBGN-PD and BioPAX.

2 MAPPING CELLDESIGNER TO BIOPAX LEVEL 3

In this work, the mapping is done between CellDesigner 4.1 and BioPAX Level 3, and refer to them as CellDesigner and BioPAX, respectively, throughout the rest of the article. Since CellDesigner supports SBML, our mapping is greatly facilitated by the existing SBML to BioPAX mappings developed by other groups (Ruebenacker *et al.*, 2009). The detailed mapping is described in

*To whom correspondence should be addressed.

the Supplementary Data 1 and Table S1. Here we will only discuss briefly our general approach of the mapping and our solutions to some of the issues that we encountered during the mapping. For the convenience of describing the mapping, all the CellDesigner terms are in *italic*, and the BioPAX terms are in *italic* and underline.

The mapping described here is designed to support the current development to export CellDesigner symbols to BioPAX ontology classes. One of our long-term goals is to import BioPAX pathways to CellDesigner. Although the general principle of the mapping is to ensure accurate translation of CellDesigner symbols to BioPAX ontology terms, we are also mindful about the accuracy of the reverse mapping.

The three main categories of symbols in CellDeigner, *Model*, *State node symbol* (also known as *species*) and *Arcs* can be mapped to *Pathway*, *PhysicialEntity* and *Interaction* in BioPAX, respectively. The symbols in each category in CellDesigner are mapped to the corresponding BioPAX subclasses in the following three approaches. First, a CellDesigner symbol has an exact (or one-to-one) match to a corresponding BioPAX subclass. Although each system uses different names, their underlying context and concept may be identical, e.g. *Heteromultimer* and *State Transition* in CellDeisgner versus *Complex* and *BiochemicalReaction* in BioPAX, respectively. Therefore, a one-to-one mapping can be created. Second, multiple CellDesigner symbols are mapped to a single BioPAX subclass, e.g. *Ion channel* and *Generic protein* symbols in CellDesigner are both mapped to *Protein* in BioPAX. Although there is loss of data in this case, most of the information can still be captured through the names (for entities) and the reactions involved (for transitions arcs). Third, the CellDesigner symbols cannot be mapped to any corresponding BioPAX subclasses, and therefore they have to be mapped to the parent class term. For example, *Inhibition* and *Physical Stimulation* in CellDesigner are mapped to *Control* in BioPAX, with values such as *Inhibition* or *Activation* associated to *ControlType*. Loss of data could be an issue in this type of mapping. In many cases, we tried to find ways within BioPAX to capture the information to minimize the information loss. The current implementation in the mapping has minimum loss of data going from CellDesigner to BioPAX. However, we do realize that the reverse map from BioPAX to CellDesigner is still ambiguous. The developers of this project are both involved in CellDesigner and BioPAX development. We are aware of the issues and are actively working toward a solution to resolve this problem (see Supplementary Data 1 for more detailed discussions on each case).

3 IMPLEMENTATION

The preliminary work has been implemented in CellDesigner 4.1 release in 2010. Due to the different release cycles of CellDesigner, BioPAX and our mapping updates, we have decided to implement the BioPAX translator as a CellDesigner plug-in, so that timely updates can be released. The existing implementation will be obsoleted in the next CellDesigner release.

CellDesigner includes an extensible plug-in system that allows third-party software to register as plug-ins for additional functionality. The interface allows any plug-in to retrieve information such as model, compartments, components and relationships between components as well as the relationships between the components and the compartments.

The BioPAX plug-in first reads the CellDesigner components via the Java plug-in API. It then maps them to the corresponding BioPAX Level 3 objects defined in the Paxtools, a Java library that can be used to create BioPAX objects and output the corresponding model in the OWL/RDF-XML format. The mapping between CellDesigner and BioPAX is not always one-to-one and the translator employs some rules and heuristics to resolve ambiguous mappings. See Supplementary Data 1 for details.

Importantly, the translator also adds cross-references to facilitate mapping to the external data sources. The translator uses web services provided by the Ontology Lookup Service to search for terms such as cellular component (GO database) (Gene Ontology Consortium, 2010), protein modification (PSI-MI) (Kerrien *et al.*, 2007) and small molecule description (ChEBI) (Degtyarenko *et al.*, 2008).

The BioPAX files produced by the plug-in are valid and error-free (Supplementary Data 1). The plug-in supports CellDesigner 4.1 and the recently released 4.2. It is freely available and can be downloaded at: ftp://ftp.pantherdb.org/CellDesigner/plugins/BioPAX/.

See Supplement Data 2 and Figure S3 for details about installation and use of the tool.

ACKNOWLEDGEMENTS

We thank Susan Paley and Peter Karp for their input in the initial map between BioPAX and CellDesigner, and Igor Rodchenkov for his continuous help to our understanding of BioPAX and Paxtools.

Funding: National Institute of General Medical Sciences (GM081084).

Conflict of Interest: none declared.

REFERENCES

Caspi,R. *et al.* (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **38**, D473–D479.

Croft,D. *et al.* (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, **39**, D691–D697.

Degtyarenko,K. *et al.* (2008) ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res.*, **36**, D344–D350.

Demir,E. *et al.* (2010) The BioPAX community standard for pathway data sharing. *Nat. Biotechnol.*, **28**, 935–942.

Funahashi,A. *et al.* (2008) CellDesigner 3.5: a versatile modeling tool for biochemical networks. *Proceedings of the IEEE 96*, 1254–1265.

Gene Ontology Consortium (2010) The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res.*, **38**, D331–D335.

Hucka,M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.

Kerrien,S. *et al.* (2007) Broadening the horizon–level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.*, **5**, 44.

Le Novère,N. *et al.* (2009) The Systems Biology Graphical Notation. *Nat. Biotechnol.*, **27**, 735–741.

Mi,H. and Thomas,P.D. (2011) Ontologies and standards in bioscience research: for machine or for human. *Front. Physio.*, **2**, 5.

Mi,H. *et al.* (2010) PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.*, **38**, D204–D210.

Ruebenacker,O. *et al.* (2009) Integrating BioPAX pathway knowledge with SBML models. *IET Syst. Biol.*, **3**, 317–328.

Schaefer,C.F. *et al.* (2009) PID: the Pathway Interaction Database. *Nucleic Acids Res.*, **37**, D674–D679.