Application Note

Pathway Commons at Virtual Cell: use of pathway data for mathematical modeling

Michael L. Blinov1,*, James C. Schaff1, Oliver Ruebenacker1, Xintao Wei1, Dan Vasilescu1, Fei Gao1, Frank Morgan1, Li Ye1, Anuradha Lakshminarayana1, Ion I. Moraru1 and Leslie M. Loew1
1Center for Cell Analysis and Modeling, University of Connecticut Health Center, Farmington, CT 06030, USA

ABSTRACT
Summary: Pathway Commons is a resource permitting simultaneous queries of multiple pathway databases. However, there is no standard mechanism for using this data (stored in BioPAX format) to annotate and build quantitative mathematical models. Therefore, we developed a new module within the Virtual Cell modeling and simulation software. It provides pathway data retrieval and visualization and enables automatic creation of executable network models directly from qualitative connections between pathway nodes.
Availability: Available at Virtual Cell (http://vcell.org/). Application runs on all major platforms and does not require registration for use on the user’s computer. Web-start and installers are available. Tutorials and video are available at User Guide page.
Contact: vcell_support@uchc.edu

1 INTRODUCTION

The Pathway Commons collection of databases (Cerami et al., 2011) provides access to more than 442,000 interactions among 86,000 physical entities. The data is retrieved in the form of Biological Pathway Exchange, BioPAX (Demir et al., 2010) that can be converted to the Systems Biology Markup Language, SBML (Hucka et al., 2003) – the gold standard for exchanging mathematical models, and the native standard for the Biomodels model database (Le Novère et al., 2006). However, while converting to SBML, the wealth of information in annotations is often lost, because the semantics of general biological knowledge differs from that of mathematical models (Ruebenacker et al., 2009).

Here we describe a tool that allows users to access the rich trove of pathway data available through Pathway Commons and use it to build well-annotated mathematical models. It was implemented in the Virtual Cell (VCell) modeling and simulation framework (Moraru et al., 2008; Cowan et al., 2012), a platform designed for building and simulating compartmental and spatial models and analysis of simulation results. The VCell database contains over 54,000 user models of which ~500 are public. However, most of these models were created “ad-hoc”, using multiple iterations. In doing so, users rarely used detailed names to identify variables and functions in the mathematical model, but most often used simplified abbreviations like R, ATP, or even arbitrary symbols such as S1, P1, etc. As a result, in the absence of extensive accompanying documentation, a model is often difficult or even impossible to understand by anyone except for the model author(s). As models increase in complexity, sharing and reusing models and components become a necessity, and creating a well-annotated model is essential. For this purpose, Minimal Information Required In the Annotation of Models (MIRIAM) community standard was developed (Le Novère et al., 2005). Our software facilitates this by linking every model element to extensive documentation coming from pathway databases, helping researchers to produce models that are easier to understand and shared with the community.

The tool stores connectivity information derived from pathway interactions and can convert it into a reaction network in VCell, which can then be populated with rate constants and initial concentrations to generate kinetic simulations, which is a useful feature for researchers learning how to model. Online tutorials describe simple ways to create a simulation-ready qualitative model of a complete pathway. We also expect the tool to be appealing to bioinformaticians as a browser and editor for data in Pathway Commons. We hope that the ease of creating computational models to utilize pathway data can bring together researchers doing mathematical modeling, biologists working with pathway data, and bioinformaticians querying and visualizing it.

2 LINKING PATHWAYS AND MODELS

2.1 Accessing and linking pathway data to existing models

The new Pathway Commons at VCell interface allows users to search pathway data from inside VCell (Fig. 1, panel A). The entries of interest can be imported and stored alongside a VCell model as a Pathway Model to be available whenever necessary (panel B). The pathway data can be searched, filtered, and viewed as a diagram or as a list of objects (panel C). Each entry in a Pathway Model includes a list of identifiers from different databases with clickable web links that lead to relevant entries in any of these databases (panel D). While working on an existing VCell model, a user can create a relevant Pathway Model and link it to entities in the VCell Reaction Diagram (panel F) in order to fully annotate the appropriate elements of the mathematical system (panel E). All such linked elements are marked with the letter L in both the Pathway Diagram and the Reaction Diagram (panel A).

*To whom correspondence should be addressed.
ontology (Ruebenacker et al., 2009), which is an RDF/OWL schema that provides mapping of pathway entities to model elements. The first version of a BioPAX to model converter was implemented as a stand-alone software tool, SyBil (Ruebenacker et al., 2011). This was then ported and integrated into core functionality of the client-server VCell platform.

We introduced two new components to VCML - the Pathway Model and the Relationship Model. These permit unambiguous storage of the complete extracted BioPAX data, without any loss that may be caused by translation, and the mapping of the pathway elements to VCell Model entities.

The Pathway Model contains all data that was extracted from Pathway Commons during model generation. Some of it can be linked to BioModel elements, but the user may keep unlinked data in the Pathway Model for future use. Pathway Commons currently supplies the data in BioPAX Level 2 format, but we already support importing data in Level 3 format. Sesame open-source framework is used for querying and analyzing this RDF data. It allows for generating Java source files from ontologies and use of SPARQL to query RDF. Java objects were created for all BioPAX Level 3 classes, with conversion from Level 2 done internally. The Pathway model is stored inside VCML as RDF annotations under the top level element. Use of RDF allows for seamless incorporation of pathway data coming from multiple sources. To operate on Pathway Model components, we implemented BioPAX I/O operations as new Java classes in VCell.

The Relationship Model is a new element of the VCML schema that links elements of VCell BioModel (species and reactions) to elements in the Pathway Model (physical entities and interactions). The mapping is many-to-many, so several pathway objects can be linked to a single model element, and vice versa. Linked pathway entities have unique identities through the CPATH ID (the identifier assigned by Pathway Commons) and we store and display all UnificationXref IDs for reuse in different models.

ACKNOWLEDGEMENTS

We thank Emek Demir, Igor Rodchenkov, Garry Bader and the BioPAX team for their valuable help.

Funding: This work was funded by the National Institutes of Health [P41-GM103313, U54-RR022232, R01-GM095485].

Conflicts of Interest: none declared.

REFERENCES


