

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

Michael L. Blinov*, James C. Schaff, Oliver Ruebenacker, Xintao Wei, Dan Vasilescu, Fei Gao, Frank Morgan, Li Ye, Anuradha Lakshminarayana, Ion I. Moraru and Leslie M. Loew
Center for Cell Analysis and Modeling, University of Connecticut Health Center, Farmington, CT 06030, USA

Associate Editor: Martin Bishop

ABSTRACT

Summary: Pathway Commons is a resource permitting simultaneous queries of multiple pathway databases. However, there is no standard mechanism for using these 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 and implementation: 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. Tutorials and video are available at user guide page.

Contact: vcell_support@uchc.edu

Received on July 1, 2013; revised on October 16, 2013; accepted on November 8, 2013

1 INTRODUCTION

The Pathway Commons collection of databases (Cerami *et al.*, 2011) provides access to >442 000 interactions among 86 000 physical entities. The data are 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 (Cowan *et al.*, 2012; Moraru *et al.*, 2008), a platform designed for building and simulating compartmental and spatial models and analysis of simulation results. The VCell database contains >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 and so forth. 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 use 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) 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

The screenshot displays the Pathway Commons at VCell interface. It is divided into several panels:

- Panel A:** Search interface for Pathway Commons. A search bar contains 'egr'. Below it, a list of search results is shown, including 'EGFR interacts with phospholipase C-gan...', 'EGFR-dependent Endothelin signaling eve...', 'EGFR1 (Cancer Cell Map, Homo sapiens)', 'EGFR1 (Cancer Cell Map, Homo sapiens)', 'SHC1 events in EGR signaling [Reaction]', 'Signaling by EGFR [Reactions, Homo sap]', 'Signaling events mediated by EGFR1 an...', and 'VEGF and VEGFR signaling by work [ICI]'.
- Panel B:** Table of 'GRB2 events in EGR signaling'. The table has columns for Entity Name, Type, and Imported?. The rows are:

Entity Name	Type	Imported?
SOS1-mediated nucleotide exchange of RAS (EGF...	biochemical ...	<input type="checkbox"/>
GRB2 binds SOS1	biochemical ...	<input checked="" type="checkbox"/>
GRB2	protein	<input checked="" type="checkbox"/>
GRB2:SOS1	complex	<input type="checkbox"/>
EGF:p-6Y-EGFR:GRB2:SOS1	complex	<input type="checkbox"/>
GRB2:SOS1 complex binds to EGF:EGFR complex	biochemical ...	<input type="checkbox"/>
- Panel C:** Pathway diagram showing interactions between GRB2, SOS1, and EGFR. The diagram includes nodes for GRB2, SOS1, and EGFR, with arrows indicating interactions. A red 'C' is overlaid on the diagram.
- Panel D:** Object Properties panel for 'GRB2'. The table shows:

Property	Value
Type	protein
Name	GRB2 (double-click lookup)
Linked physiology ...	GRB2_c0
Unif	UNIPROT:P52993-1
xref	CPATH:504162
xref (related)	ENTREZ_GENE:2885
Comment	DOMAIN The SH3 domains mediate interaction with RALGP...
- Panel E:** Object Properties panel for 'EGFR'. The table shows:

Property	Value
Type	protein
Name	EGFR (double-click lookup)
Linked physiology ...	EGFR_c0
Unif	UNIPROT:P42292-1
xref	CPATH:504162
xref (related)	ENTREZ_GENE:2885
Comment	DOMAIN The SH3 domains mediate interaction with RALGP...
- Panel F:** Navigation tree on the left side of the interface, showing categories like Physiology, Reaction Diagram, Structures, Species, Applications, Parameters and Functions, and Pathway.

Fig. 1. Pathway Commons at VCell. The Pathway database panel (A) allows the user to search Pathway Commons and select pathways for bringing into VCell. The Pathway preview panel (B) allows the user to import elements of the selected pathway into VCell. The Pathway view panel (C) allows the user to see multiple views of imported pathway data, to select elements to inspect in the Object properties panel (D) and to convert into VCell model elements (E). Selecting an item under Pathway in the navigation panel (F) will open the corresponding tab in the view panel (C)

2.2 Creating new models from pathway data

Selected physical entities and physical interactions from pathways can be automatically converted into VCell species and reactions in a *Reaction Diagram*. The default location of newly created species is in a single compartment and the default stoichiometries in newly created reactions are set to 1, as these types of information are often missing in the pathway databases. Users can then define multiple compartments and change reaction properties. Reaction kinetics defaults to mass action, but any kinetic formalism and rate laws can be specified. The new VCell model is fully annotated providing all available information about this element coming from multiple databases (Fig. 1, bottom right panel). Once all the initial concentrations (or copy numbers for stochastic simulations), rate constants and cellular geometry parameters are specified, the model can be sent to the appropriate VCell simulator to provide time courses computation.

3 IMPLEMENTATION

The format for VCell model representation is the VCell Markup Language (VCML), an XML dialect that has some similarities to SBML. Pathway Commons provides data in BioPAX format (Demir *et al.*, 2010), which is an Resource Description Framework (RDF)-based ontology. As we pointed out earlier (Ruebenacker *et al.*, 2009), BioPAX data are not fully

compatible with XML-based modeling formats, and converting BioPAX data to annotated SBML or VCML will lead to data loss and distortion (Ruebenacker *et al.*, 2009). Previously, we developed an intermediate representation in the form of the SBPAX 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 standalone software tool, SyBil (Ruebenacker and Blinov, 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 these 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.

ACKNOWLEDGEMENT

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

Funding: National Institutes of Health (P41-GM103313, U54-RR022232, R01-GM095485).

Conflict of Interest: none declared.

REFERENCES

- Cerami, E.G. *et al.* (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.*, **39**, D685–D690.
- Cowan, A.E. *et al.* (2012) Spatial modeling of cell signaling networks. *Methods Cell Biol.*, **110**, 195–221.
- Demir, E. *et al.* (2010) The BioPAX community standard for pathway data sharing. *Nat. Biotechnol.*, **28**, 935–942.

- 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.
- Le Novère,N. *et al.* (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.*, **23**, 1509–1515.
- Le Novère,N. *et al.* (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.*, **34**, D689–D691.
- Moraru,I.I. *et al.* (2008) The Virtual Cell modeling and simulation software environment. *IET Syst. Biol.*, **2**, 352–362.
- Ruebenacker,O. and Blinov,M.L. (2011) Using views of Systems Biology Cloud: application for model building. *Theory Biosci.*, **130**, 45–54.
- Ruebenacker,O. *et al.* (2009) Integrating BioPAX pathway knowledge with SBML models. *IET Syst. Biol.*, **3**, 317–328.