3 Systems and Prototypes

PathCase-SB: Database-Enabled Tools for Regulatory Metabolic Networks

X. Qi, S.A. Coskun, A. Cakmak, E. Cheng, E. A.Cicek, M. Das, L. Yang, R. Jadeja, S. Syed, R.K. Dash, N. Lai, G. Özsoyoğlu, Z. M. Özsoyoğlu

ABSTRACT

Integration of metabolic pathways resources and metabolic network models, and deploying new tools on the integrated platform is useful for systems biology research on understanding the regulation of metabolic networks. PathCase-SB is such an integrative a web-based application, providing a database-enabled framework and tools towards effective and efficient systems biology model development for mechanistic simulations of biological systems. Current PathCase-SB user interfaces include browser, querying, visualization, provenance, simulation and model composition interfaces. PathCase-SB is built, released, and already being used by researchers across the globe.

INTRODUCTION

PathCase Systems Biology (PathCase-SB) is a web-based application that brings together (a) systems biology data sources (currently BioModels Database [1]), and (b) pathways data sources (currently, KEGG [2]), with the goal of providing additional and new capabilities and tools, made possible due to the integration. The premise is that the integrated use of regulatory metabolic network models and metabolic pathways resources allows for new capabilities and tools to be built that would not be possible otherwise, and thus can help systems biology researchers in understanding the regulation in metabolic networks.

PathCase-SB does not curate models or pathways. It provides (currently, six) user interfaces for users to (i) explore (browse), search and query both models and pathways, (ii) visualize both pathways and modeled networks, and view their mappings in multiple ways, (iii) comparatively simulate, possibly with users' own data, either BioModels Database models or users' own models uploaded to PathCase-SB, (iv) compose a new model from existing SBML models in a semi-automated manner.

In sections below, we summarize the capabilities of PathCase-SB, Version 2, with additional capabilities of model composition added recently. More details about the PathCase-architecture and database design, and PathCase Interfaces are available elsewhere [3, 4].

METHODS AND IMPLEMENTATION

PathCase-SB Database

The database (a Microsoft SQLServer database) is designed to contain data from different systems biology and biochemical network databases in separate tables for common entities of different (systems biology and/or biochemical network) data sources, e.g., currently, BioModels Database (and, in the future, CellML [5]) and KEGG (and, in the future, Reactome [6]) data sources. As an example, we maintain a species table occurring in BioModels Database models and molecular_entities table for KEGG molecular entities. Such an approach allows us to separate data from different data sources cleanly, and, to add new data sources seamlessly, without dealing with data cleansing, data integration, and data curation problems.

Strict data-separation-per-data-source approach of PathCase-SB also requires an additional "mapping" effort involving all pairs of data sources on their corresponding entities. For BioModels Database and KEGG, this involves three mappings: <species, molecular entities>, <reactions, process-entities>, <models, pathways>. PathCase-SB database also keeps separate tables for different systems biology data sources, e.g., currently, distinct sets of tables are maintained for BioModels Database and CellML (not yet open to public) data sources. Please see [3] for more details.

Currently, PathCase-SB database contains four classes of tables capturing the following information: (a) biochemical reaction network-related tables, (b) systems biology data source-related tables (e.g., quantitative kinetic information, dynamic behavior, involved species, reactions, etc.), (c) tables mapping data from different data sources, (d) tables annotating systems biology data with other ontologies or taxonomies, e.g., Gene Ontology [7].

PathCase-SB Browser Interface

Browser interface provides a variety of browsing-based mechanisms for users to access the PathCase-SB database, starting from a basic overview that lists the entities in the database to hierarchically drilled-down levels that include, among others, reaction, species, and compartments. The PathCase-SB browser interface presents a unique multi-faceted view of the database, which allows users to access and organize the biochemical information with distinct focus points. As an example, researchers have the option to browse models by their corresponding pathways, studied organisms, or relevant GO terms (e.g., for an enrichment pre-study). Each browser item is linked to an information-rich "details page" that organizes (i) lists of participants and their roles in each model entry, and the kinetic models of the corresponding biochemical reactions and their parameters, (ii) gateways to interactive graphical tools and interfaces (e.g., simulation and visualization engines), (iii) data provenance information for source tracking, and (iv) entry points to related parameterized queries for a customized and focused study of the underlying data. Fig.1, part A illustrates the browser expansions as a result of clicking (i) to "BioModels Database models by name" (leftmost pane), and, then (ii) to model *Akman2008_Circadian_Clock_Model2* (next leftmost pane), which results in the final expansion.

The PathCase-SB Browser interface provides to users

- An embedded in-place keyword search facility with paged result listings,
- Relationships between BioModels Database models and ontologies (e.g., Gene Ontology, EC (Enzyme Commission) number taxonomy [8], and System Biology Ontology [9]),
- Biological compartment-based relationships between different models. The idea is to allow

modelers to see the listings of models that capture biological networks within the same compartment (e.g., say, liver cytosol) to help them with larger-model composition and model merge operations.

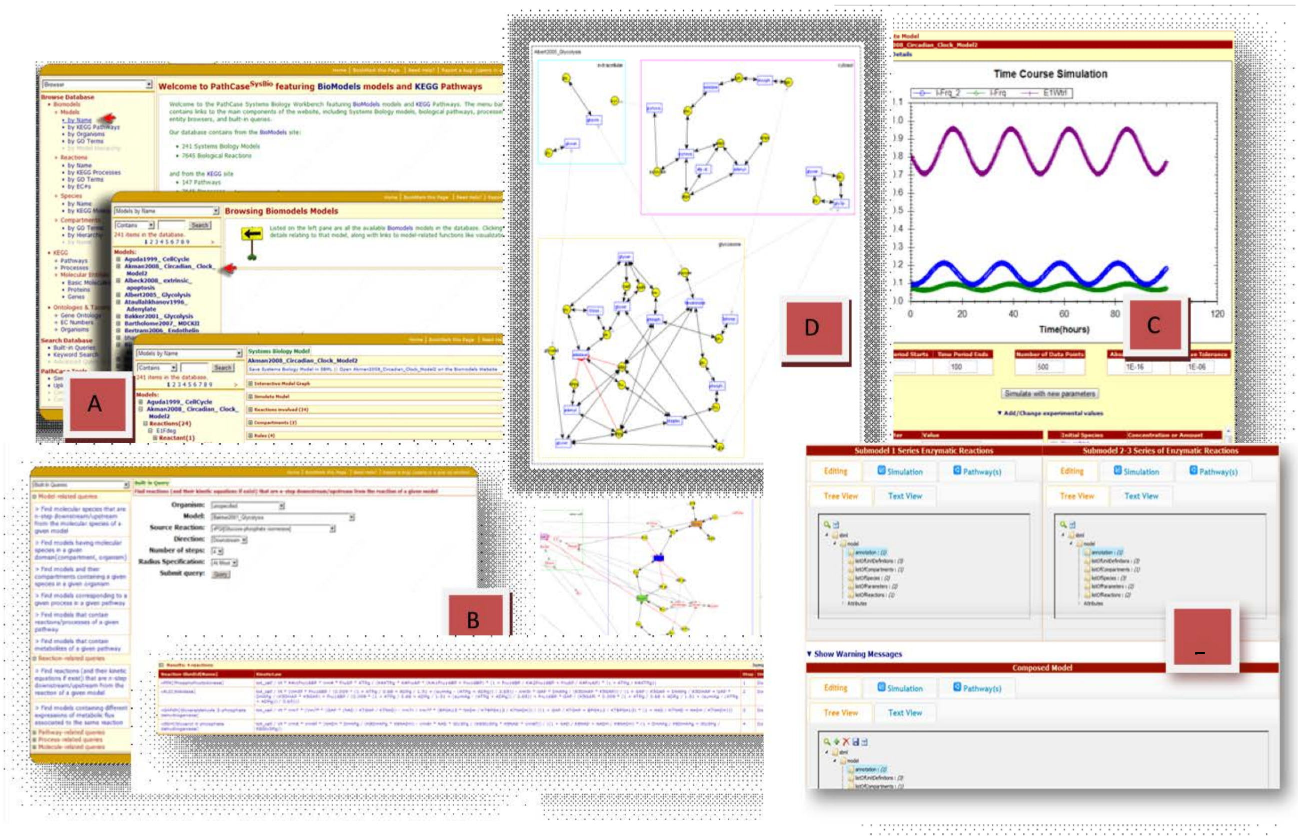

Figure 1. Sample views of the PathCase-SB System. Parts A, B, C, D, and E represent Browser Interface, Built-In Queries Interface, Simulation Interface, Visualization

PathCase-SB Querying Interface

PathCase-SB is designed for users to pose built-in (i.e., predefined) queries involving models and other database objects. Built-in queries can be characterized as a small set of popular queries that are provided through very simple user interfaces. Note that some of the built-in queries are pathways-only queries from PathCase[10], and, provided only as a convenience.

PathCase-SB built-in queries are grouped into

- *Model queries*. E.g., “Find models and their compartments containing a given species in a given organism”.
- *Pathway queries*. E.g., “Visualize a set of pathways”.

Reaction queries. E.g., “Find reactions (and their kinetic equations if exist) that are n-step downstream/upstream from the reaction of a given model”, which is illustrated in Fig. 1, part B: (i) query specification shown in the pane to the left of "B", (ii) visualized query result shown in the pane to the right of "B", and (iii) textual query result shown in the pane below "B".

- *Species/molecule queries*. E.g., “Find species that are n-step downstream/upstream from the species of a given model”).

In the spirit of the open communication framework, all built-in queries are made available through web services, so that third-party applications can directly send their query execution requests to the web services, execute the requested PathCase-SB built-in query, and receive the execution results as an SBML (for a model-based output) or as a BioPax (for a metabolic network graph output) document. Such an approach promotes open data exchange, and is beneficial to other tool builders. This is also consistent with the currently available web services for built-in queries of the existing PathCase system.

PathCase-SB Simulation Interface

This interface allows users to either directly use the SBML files of models in the PathCase-SB database or to upload their own SBML files of models (the *iModel Tool*), and simulate them. In addition, users can see simulation results side by side in a comparative manner (the *SimCom Tool*). For the simulation, a high performance cellular network simulation service, namely, RoadRunner [11, 12] is used through the REST web services provided by Systems Biology Workbench application[13] programming interface (API). A third party library called ZedGraph (available at http://zedgraph.org) is employed to render the simulation results as a graph.

Fig. 1, Part C illustrates the simulation of the model *Akman2008_Circadian_Clock_Model2* (also shown in Fig. 1, Part A). Note that the simulation interface lets users to modify: a) parameter values; b) initial conditions for time, concentrations or amounts of species and boundary species; c)number of data points to plot; d) absolute and relative tolerance. Also, it allows to select metabolic fluxes to plot, , add experimental values, and visualize results of new simulation within the same time-course simulation graph.

Within the simulation interface, input for experimental results is manually editable on the field specified for users. Users can directly use comma or space delimited input from commonly used environments such as MATLAB or Excel. Users can also access model details (e.g., model version) from the link above the simulation graph.

In general, there is more than one mathematical model for the same pathway. Therefore side-by-side comparisons of model simulations for the same pathway can allow researchers to observe similarities and differences between models. The *SimCom tool* provides the functionality to simulate up to four different models in the same pathway side by side (in new pop-up windows) from PathCase Systems Biology web site.

The *iModel tool* allows users to upload their own SBML models onto the PathCase Systems Biology web site to simulate. Uploaded models are parsed by the *PathCase SBML Parser* [14] which uses libSBML[15] library. After parsing, the model is stored in a separate temporary database (which is emptied regularly for privacy and copyright protection purposes; therefore the uploaded models are not kept in our real database), and input to the *iModel tool*.

In terms of import/export capabilities, PathCase-SB does not provide export capabilities for models and pathways since all of its models and pathways are available in the original data sources (e.g., BioModels Database, KEGG). Import capabilities for pathways is provided by PathCase[10], and import capabilities for models is provided by the *iModel tool*.

Finally, simulation interface is fully-integrated with model detail pages to simulate a particular model in-place, while allowing browsing model details.

PathCase-SB Visualization Interface

A client side java applet, called PathCase-SB Graph Viewer, produces interactive pathway graphs, model graphs, or both with various mappings from one to the other. The visualized model/pathway can be manually or automatically rearranged, zoomed in/out, panned, expanded/collapsed, queried from, saved locally as jpeg file, and studied in detail. Fig. 1, part D, illustrates the visualization of the BioModels Database systems biology model *Albert2005_Glycolysis* (reached by clicking to the "Interactive Model Graph" link of the model). Note that the modelers of this model have specified three compartments, namely, "extracellular", "cytosol", and "Glysome". Also note that Pathcase-SB system does not curate data, and simply displays the model/pathway as curated by the original data source (i.e., in this model's case, the BioModels Database).

All visualization features are provided on client-side, and without server-side communication, allowing for high scalability. Visualization is embedded into the corresponding model/pathway pages, and (i) does not require any separate installation effort as a manageability convenience for users, and (ii) provides a platform-independent access regardless of the client operating system or browser differences (presently optimized for IE, Firefox, Safari, and Chrome browsers).

PathCase-SB *Graph Viewer*, when accessed from different places within PathCase-SB, has many different legends, basic controls, and toolbar capabilities. The *Graph Viewer* is employed by

- *Browser Interface* (appears as a menu item at many places with name "Interactive Model/Pathway Graph") to visualize the full PathCase-SB metabolic network (in multiple condensed/expanded forms), individual pathways, metabolic subnetworks, and networks of systems biology models,

- *Built-In Queries*: For each query that produces a metabolic subnetwork, the results are visualized by the *Graph Viewer*,

- *The iModel Tool* (e.g., "Upload your own model"): Biochemical networks of user models are visualized by the *Graph Viewer*.

The functionality of the *Graph Viewer* will be expanded in the future, to be used in future tools "Compose-Models" and "Compare-Models".

PathCase-SB Provenance Tool

Provenance, also called lineage or pedigree, is defined as metadata that tracks the steps of data derivation, which adds value to the data itself [16]. With our use of other web-based data sources, providing provenance of data in PathCase-SB is a necessity. We provide three levels of provenance data:

- *Model related information:* Creation date, Modification date, Notes of author and publication id. This data is stored in our servers.

- *Authors and publications:* Author-related information is stored in our servers, whereas publication data is referred to the outer source (BioModels Database CiteXplore),

- *Papers* that are being cited by the publication of the model. This level is again hosted by BioModels Database CiteXplore.

All three layers of provenance are specified via a stand-alone panel: when a model in a panel is opened, provenance data about the selected model is displayed.

Finally, as a short comparison, we note that there are at least three web-based applications with model simulation and visualization features (http://cytosolve.mit.edu, http://jjj.biochem.sun.ac.za, and http://www.itb.cnr.it/cellcycle). However, PathCase-SB is a comprehensive web-based application that not only includes simulation and visualization (of model networks and pathways), but also has database, browser and querying interfaces, as well as a provenance tool.

Model Composition Interface

This interface is an AJAX[17] enabled advanced web-based biological model composition interface for SBML formatted mathematical models. The interface allows user to choose two standard SBML models, either from the user's computer or from PathCase-SB database, and merge them into a new model. Part of the merge process is completed automatically and then the modeler can manually alter combined models to finalize the merge results. Fig. 1, Part E illustrates an overview of this interface.

Within the model composition interface, four components, namely, Tree View, Text View, Simulation and the Pathway(s) visualization tabs are provided. The tree and text views enable the user to edit the SBML model. The Tree View contains an easy-to-use hierarchical XML representation of the SBML model. The Text View gives advanced users the ability to go over the original SBML file. The Simulation and Pathway(s) visualization tabs have the same capabilities as discussed before. Both source modes and new merged models can be simulated and visualized in the integrated model composition interface.

Currently, SBML versions 1 and 2 of Level 1, Version 1, 2, 3 and 4 of Level 2, and Version 1 of Level 3 are all supported. For more details about interfaces and/or architectural details about PathCase-SB model composition interfaces, please see [18].

Future Extensions

In the future, we are planning to

- add CellML models and Reactome pathways into PathCase-SB, and
- add to PathCase-SB visualization interface special features that allow users to provide "omics" related data analysis.

ACKNOWLEDGEMENT

We gratefully acknowledge the ideas and contributions of our deceased co-researcher Dr. Marco Cabrera.

REFERENCES

[1] N. Le Novère, et al, "BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems." *Nucleic Acids Res*. 34 (Database-Issue), 2006: 689-691.

[2] M. Kanehisa, et al, "From genomics to chemical genomics: new developments in KEGG." *Nucleic Acids Res.* 34, 2006, D354-357.

[3] A. Cakmak, et al, "PathCase-SB architecture and database design." *BMC Syst Biol.* Nov 2011 9;5:188.

[4] S. Coskun, et al , "PathCase-SB: Integrating Data Sources and Providing Tools for Systems Biology Research." *BMC Syst Biol.* , submitted for publication.

[5] C.M. Lloyd, et al. "CELLML: Its future, present, and past." *Progress in Biophysics and Molecular biology*. Vol. 85, Issues 2-3, 2004, pp. 433-450.

[6] L. Matthews, et al, "Reactome knowledgebase of human biological pathways and processes."

Nucleic Acids Res. Jan 2009;37(Database issue):D619-22.

[7] M. Ashburner, et al, "Gene Ontology: tool for the unification of biology." Nat. Gen. May 2000, 25(1):25-9.

[8] E. C. Webb, *Enzyme Nomenclature 1992*. Int. Union of Biochemistry and Molecular Biology, Academic Press.

[9] N. Le Novere, et al, "Adding semantics in kinetic models of biochemical pathways." *ESCEC Conf.*, 2006.

[10] B. Elliott, et al, "PathCase: Pathways database System. Metabolic Pathways System." *Bioinformatics*. Vol. 24, 2008, No. 21, pp.2526-2533.

[11] F.T. Bergmann, and H.M. Sauro, " SBW – a modular framework for systems biology." in *Proc., 38th Conf. on Winter Simulation*, 2006, pp. 1637–1645.

[12] F.T. Bergmann, and H.M. Sauro, "Comparing simulation results of SBML capable simulators." *Bioinformatics*, Vol. 24, 2008 Pp. 1963 – 1965.

[13] H. M. Sauro, et al. "Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration." *OMICS*. Winter, 2003; 7(4):355-72.

[14] M. Das, "PathCase: SBML Parser and CellML Parser." MS project. CWRU 2010.

[15] J. B. Benjamin, et al, "LibSBML: an API Library for SBML." *Bioinformatics* 24(6), 2008: 880-881.

[16] R. Bose, and J. Frew, "Lineage retrieval for Scientific Data processing: a survey." *ACM Comp*. 2005, 37(1).

[17] http://en.wikipedia.org/wiki/AJAX

[18] A. Coskun, "PathCase-SB Model Composition Interface: Biological Model Composition Tool for SBML Models", manuscript under preparation, 2012.