

SBML Reaction Finder: Retrieve and extract specific reactions from the BioModels database

Maxwell Lewis Neal^{1*} and Herbert M. Sauro²

¹ Division of Biomedical and Health Informatics

² Department of Bioengineering
University of Washington, Seattle, WA, 98195-7240

* Corresponding author

Address:

Maxwell Lewis Neal
University of Washington
Division of Biomedical and Health Informatics
1959 NE Pacific Street
Box 357240
Seattle, WA 98195-7240

Phone:

MLN: (206) 543-8769

Email:

MLN: mneal@uw.edu

HMS: <mailto:hsauro@uw.edu>

Abstract

Summary: The SBML Reaction Finder (SRF) application leverages the deep semantic annotations in the BioModels database to provide efficient retrieval and extraction of individual reactions from SBML models. We hope that the SRF will be useful to quantitative modelers who seek to accelerate their modeling efforts by reusing previously published representations of specific chemical reactions.

Availability and Implementation: The SRF is open source, coded in Java, and distributed under the Mozilla Public License Version 1.1. Windows, Macintosh and Linux distributions are available for download at

<http://sourceforge.net/projects/sbmlrxnfinder>.

Contact: mneal@uw.edu

Overview

The BioModels database of Systems Biology Markup Language (SBML) models represents one of the richest collections of reusable, quantitative chemical network models (Le Novère, et al., 2006), and the BioModels website has many features for discovering, exploring and downloading these models. Together, the several hundred curated models in BioModels represent thousands of annotated biochemical reactions,

and the repository website provides tools for investigating any given model's individual reactions. However, discovering the different formulations for one particular kind of reaction (phosphofruktokinase activity, for example) and extracting an instance for reuse is currently a cumbersome task, requiring navigation through a number of webpages, and direct exploration of model organization. We have developed a tool called the SBML Reaction Finder (SRF) to address this limitation and to provide modelers with the means to quickly retrieve and reuse quantitative representations of individual chemical reactions found in the BioModels database. We built the SRF to simplify access to these reactions, and to promote a more modular approach to chemical network modeling by making model components as accessible as the models themselves (Sauro and Bergmann, 2009).

Search and extraction features

The SRF provides a simple interface for retrieving specific reactions of interest (Figure 1). Users enter a search phrase and the SRF looks for pattern matches within a custom knowledge base that represents each individual reaction in the BioModels database. As the user types in a search phrase, the SRF automatically suggests standard reaction names collected from the reactions' Gene Ontology (GO) annotations to assist users in finding reactions of interest. (The BioModels curation team uses other knowledge sources for annotating reactions as well, such as KEGG, but currently the SRF only utilizes knowledge associated with GO annotations.) Users may also limit their search results to specific types of organisms using the "Limit search by taxon" drop-down box.

After performing a search, the SRF lists the individual reactions within BioModels that match the search criteria, including the preferred name of each reaction's GO annotation (if present), the name of its parent model, and the taxon represented. Hovering over an individual search result will also display the BioModels ID number for the reaction's parent model. Clicking a search result in the top panel displays the quantitative details of the reaction in the bottom panel, including the names of the reactants, products, and modifiers, the kinetic law that sets the reaction rate, and the values of the reaction-specific and global model parameters. For example, Figure 1 shows a search on "pfk," with the details for the phosphofruktokinase reaction in the carbon metabolism model by Chassagnole et al. (Chassagnole, et al., 2002) listed in the bottom panel.

To reuse single reactions found by the SRF, users can extract a selected reaction as a new, standalone SBML model by clicking the "Save reaction as SBML" button. Alternatively, users can simply copy and paste the reaction formula and parameter values into their simulation environment of choice (MATLAB, etc.). Users can also choose to store each individual reaction from the parent model separately by clicking the "Save all *n* model reactions" button. The SRF also provides a simple method for copying the reaction data directly into the Jarnac simulation tool that is part of the Systems Biology Workbench (Bergmann and Sauro, 2006; Sauro, et al., 2003). Right- or control-clicking the bottom panel of the SRF and selecting "Copy as SBML" will copy an SBML representation of the reaction to the clipboard that can be directly added to a Jarnac model by pasting the clipboard contents into Jarnac's model editor interface.

The SBML Reaction Finder knowledge base

To retrieve individual reactions of interest, the SRF searches a custom knowledge base, coded in the Web Ontology Language (OWL), that stores the biological annotations, unique identifiers, and source model affiliation for each reaction in BioModels. Although RDF would be expressive enough to capture the information in our knowledge base, we chose to use OWL so that, as the tool grows in sophistication, we will have the opportunity to explore automated reasoning as a means of improving model retrieval and extraction. At the time of manuscript preparation, the SRF's knowledge base contained a total of 6,533 reactions, 5,648 (86%) of which are associated with a GO term, either through direct annotation against GO or through Enzyme Commission number cross-references. Each time the user performs a search with the SRF, the program looks for pattern matches in the names and IDs of the reactions in this knowledge base, the preferred names and synonyms of any GO annotations, the names of their parent models, and the names and synonyms of the taxon in which they occur. Users can update this knowledge base at any time to ensure that they are searching over the most up-to date version of the BioModels database. The SRF update tool makes use of BioModels, UniProt, BioPortal and BRENDA web services to collect all the curated SBML models in the repository, classify their reactions according to their GO annotations, and store the human-readable names of these annotations for easier retrieval of individual reactions.

When extracting a single reaction, the SRF uses the knowledge base to find the reaction's source model and unique ID. The program then uses functions provided by the libSBML package (<http://sourceforge.net/projects/sbml/files/libsbml/>) version 5.0.0 to load the source model from a local store of the BioModels corpus, collect all of the information associated with the reaction, and then write out a new SBML model.

Discussion and Availability

The functionality of the SRF demonstrates the important role that semantics play in a modular approach to modeling. Because the models in the BioModels database include rich semantic annotations against reference ontologies, modelers interested in reusing reactions from the repository are not required to know the code-level, idiosyncratic identifiers of these reactions to discover them. Instead, they can search for and extract components at a biological, rather than syntactic, level of conceptualization (Sauro and Bergmann, 2008). With such semantics-based extraction methods, modelers can easily build collections of interoperable model components and ultimately apply them in the modular construction of more complex systems. It is our hope that the SRF will be useful to quantitative modelers who seek to accelerate their own modeling efforts by utilizing previously published representations of specific chemical reactions.

The SRF is open source, coded in Java, and distributed under the Mozilla Public License Version 1.1. Windows, Macintosh and Linux distributions are available for download at <http://sourceforge.net/projects/sbmlrxnfinder/>.

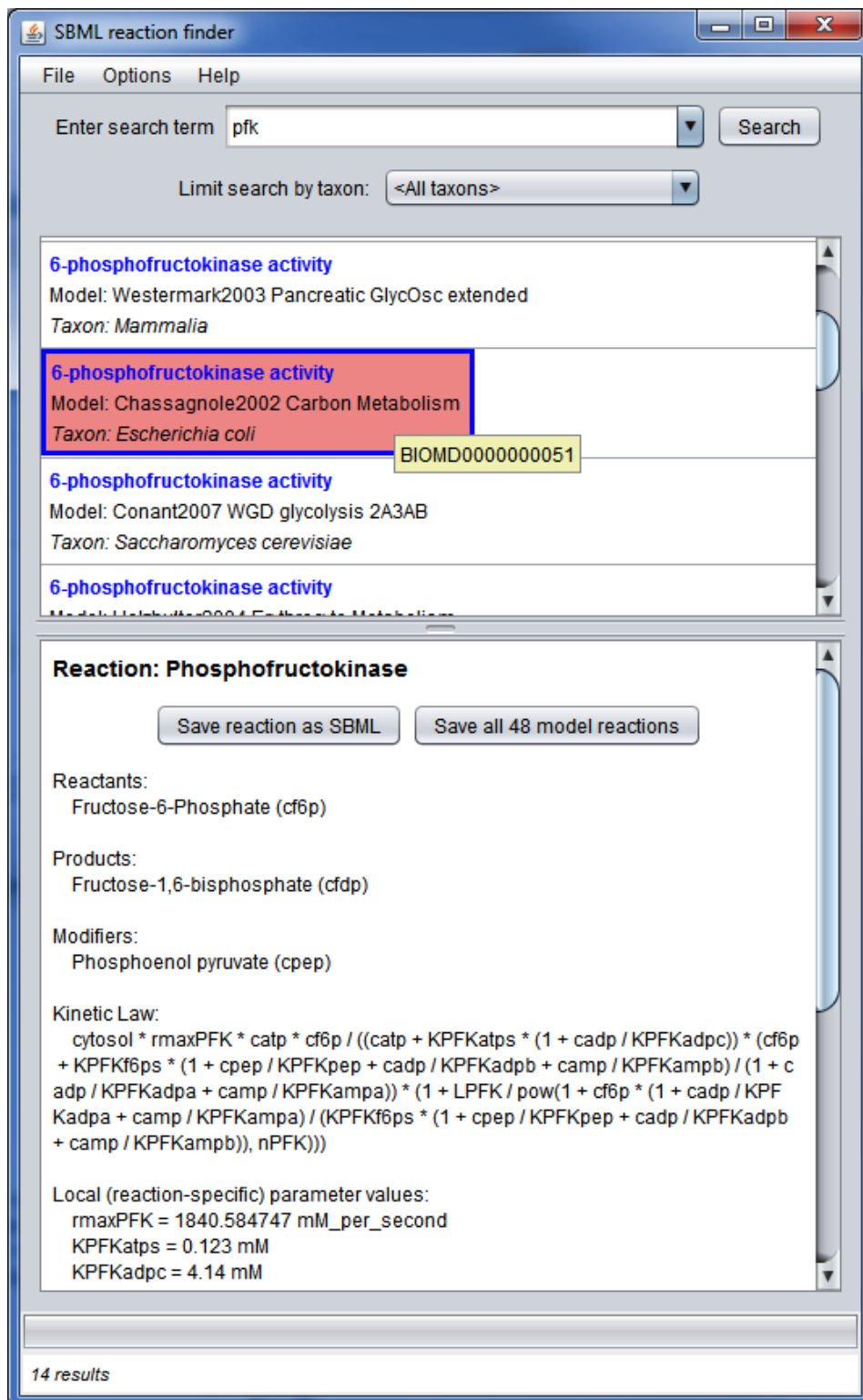


Figure 1. Windows screen shot of the SBML Reaction Finder showing results for a search on "pfk." The bottom panel shows the data for the selected reaction: phosphofruktokinase enzyme activity in the 2002 Chassagnole et al. model of *Escherichia coli* carbon metabolism.

References

Bergmann, F.T. and Sauro, H.M. (2006) SBW-a modular framework for systems biology. *Proceedings of the 37th Winter Simulation Conference*. 1637-1645.

Chassagnole, C., Noisommit-Rizzi, N., Schmid, J.W., Mauch, K. and Reuss, M. (2002) Dynamic modeling of the central carbon metabolism of *Escherichia coli*, *Biotechnology and Bioengineering*, **79**, 53-73.

Le Novere, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., Sauro, H., Schilstra, M. and Shapiro, B. (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems, *Nucleic Acids Research*, **34**, D689-D691.

Sauro, H.M. and Bergmann, F. (2008) Standards and ontologies in computational systems biology, *Essays in biochemistry*, **45**, 211.

Sauro, H.M. and Bergmann, F.T. (2009) Software Tools for Systems Biology. In Liu, E.T. and Lauffenburger, D.A. (eds), *Systems biomedicine*. Elsevier, 289-310.

Sauro, H.M., Hucka, M., Finney, A., Wellock, C., Bolouri, H., Doyle, J. and Kitano, H. (2003) Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration, *Omics A Journal of Integrative Biology*, **7**, 355-372.