A C library for retrieving specific reactions from the BioModels database

M. L. Neal[1,*], M. Galdzicki[1], J. T. Gallimore[2] and H. M. Sauro[1]

[1]Department of Bioengineering, University of Washington, Seattle, WA 98195, USA and [2]Department of Computer Science, Wofford College, Spartanburg, SC 29303, USA

ABSTRACT

Summary: We describe libSBMLReactionFinder, a C library for retrieving specific biochemical reactions from the curated systems biology markup language models contained in the BioModels database. The library leverages semantic annotations in the database to associate reactions with human-readable descriptions, making the reactions retrievable through simple string searches. Our goal is to provide a useful tool for quantitative modelers who seek to accelerate modeling efforts through the reuse of previously published representations of specific chemical reactions.

Availability and implementation: The library is open-source and dual licensed under the Mozilla Public License Version 2.0 and GNU General Public License Version 2.0. Project source code, downloads and documentation are available at http://code.google.com/p/lib-sbml-reaction-finder.

Contact: mneal@uw.edu

1 OVERVIEW

The BioModels database of systems biology markup language (SBML) models represents one of the richest collections of reusable quantitative chemical network models in existence (Hucka *et al.*, 2003; Li *et al.*, 2010), and the BioModels Web site 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 Web site provides tools for investigating any given model's individual reactions. However, discovering the different formulations for one particular kind of reaction (phosphofructokinase activity, for example) and extracting an instance for reuse is currently a cumbersome task, requiring navigation through a number of web pages and direct exploration of model organization. We have developed a C library called libSBMLReactionFinder (libSBMLRF) to address this limitation and to provide modelers with the means to programmatically retrieve and reuse quantitative representations of individual chemical reactions found in the BioModels database. We built the library 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.

*To whom correspondence should be addressed.

2 SEARCH AND RETRIEVAL FEATURES

libSBMLRF provides a simple application programming interface (API) for retrieving reactions of interest. Users enter a phrase into a search function, and the library looks for pattern matches within a custom knowledge base that represents each individual reaction in the BioModels database (see Section 3). After performing a search, the libSBMLRF returns data associated with each individual reaction in BioModels that match the search criteria, including the reaction's BioModels ID, name, the minimal information required in the annotation of models (MIRIAM) identifiers it is annotated against and the BioModels ID of the parent model. Users may limit their search results to specific organisms by specifying the scientific name of the organism in the search call. For example, a search for 'phosphofructokinase' limited to *Saccharomyces cerevisiae* models returns nine results, each representing a specific reaction in the BioModels database. Removing the organism limiter on this search returns 17 reactions.

The library also provides convenience functions for retrieving the IDs of the reactants, products and modifiers in the reaction, the kinetic law governing the reaction rate and the names and values of the parameters used in the kinetic law. These methods rely on functionality provided by the libSBML software library (Bornstein *et al.*, 2008) for accessing the original representation of a specific reaction in its parent SBML model. We include precompiled binaries of libSBML version 5.8.0 as part of our library distribution.

3 KNOWLEDGE BASE OF BIOCHEMICAL REACTIONS

To retrieve individual reactions of interest, libSBMLRF searches a custom knowledge base, coded in extensible mark-up language, that stores the semantic classification, unique identifiers and source model affiliation for each BioModels reaction. At the time of manuscript preparation, this knowledge base contained a total of 12 335 reactions. When users perform searches, the library looks for pattern matches in the names of the reactions in this knowledge base, their IDs and the preferred names and synonyms of any Gene Ontology, Kyoto Encyclopedia of Genes and Genomes or Enzyme Commission identifiers they are annotated against. The library includes a function for generating this knowledge base anew, ensuring that users can get the most up-to-date information about BioModels reactions. This function makes use of libSBML as well as BioModels, UniProt,

```
<SBMLreactionFinderKB>
  <listOfCuratedSBMLmodels>
    ...
    <curatedModel id="BIOMD0000000253" name="Teusink1998_Glycolysis_TurboDesign" taxonURI="http://identifiers.org/taxonomy/4932"/>
    ...
  </listOfCuratedSBMLmodels>
  <listOfTaxons>
    ...
    <taxon commonName="Baker's yeast" scientificName="Saccharomyces cerevisiae" miriamURI="http://identifiers.org/taxonomy/4932"/>
    ...
  </listOfTaxons>
  <listOfReactions>
    ...
    <SBMLreaction id="PFK" name="phosphofructokinase" sourceModelID="BIOMD0000000253">
      <miriamURI>http://identifiers.org/obo.go/GO:0008662</miriamURI>
    </SBMLreaction>
    ...
  </listOfReactions>
  <listOfReactionCVterms>
    ...
    <reactionCVterm name="1-phosphofructokinase activity" miriamURI="http://identifiers.org/obo.go/GO:0008662" ... >
      <synonym>fructose 1-phosphate kinase activity</synonym>
      <synonym>fructose-1-phosphate kinase activity</synonym>
      <synonym>1-phosphofructokinase (phosphorylating)</synonym>
      <synonym>ATP:D-fructose-phosphate 6-phosphotransferase activity</synonym>
      <synonym>D-fructose-1-phosphate kinase activity</synonym>
    </reactionCVterm>
    ...
  </listOfReactionCVterms>
</SBMLreactionFinderKB>
```

Fig. 1. A portion of the extensible mark-up language knowledge base that is automatically created by libSBMLRF and used to retrieve biochemical reactions of interest. Information for the phosphofructokinase reaction named 'PFK' in model BIOMD0000000253 is shown

BioPortal and Kyoto Encyclopedia of Genes and Genomes web services to collect all the curated SBML models in the repository, cache them locally, identify all reactions in the models and associate the reactions with the human-readable names and synonyms of the controlled vocabulary terms they are annotated against. The knowledge base has a simple schema that includes lists of curated SBML models, biological taxons associated with those models, individual reactions and controlled vocabulary terms used to annotate the reactions. Figure 1 illustrates the schema by showing the representation of the phosphofructokinase reaction from model BIOMD0000000253 (Teusink *et al.*, 1998).

The knowledge base associates reactions with searchable human-readable descriptions and acts as an index for all reactions in the BioModels corpus. When using the library's functions to retrieve a specific reaction's chemical participants, its rate law or its rate law parameters, the library looks up the reaction ID and the ID of its parent model in the knowledge base, then uses functions provided by libSBML to load the reaction's parent SBML model and return the additional information.

4 DISCUSSION AND AVAILABILITY

The functionality of libSBMLRF demonstrates the important role that semantics play in a modular approach to modeling (Sauro and Bergmann, 2008). Because the models in the BioModels repository 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. With such semantics-based retrieval methods, modelers can build collections of interoperable model components and apply them in the modular construction of more complex systems. It is our hope libSBMLRF will be useful to quantitative modelers who seek to accelerate their own modeling efforts by using previously published representations of specific chemical reactions.

libSBMLRF is an outgrowth of the SBML Reaction Finder project (Neal and Sauro, 2012), which is a stand-alone Java-based application for retrieving and extracting reactions of interest from the BioModels database. Although both projects aim to give modelers direct, efficient access to individual biochemical reactions, we created libSBMLRF to make this functionality available programmatically as a C API. We chose C primarily because we plan to integrate the library within our JDesigner software and a new software project currently under development, both of which are coded in C. libSBMLRF is open-source and dual licensed under the Mozilla Public License Version 2.0 and GNU General Public License Version 2.0. Source code, project downloads and documentation are available at http://code.google.com/p/lib-sbml-reaction-finder.

Conflict of Interest: none declared.

REFERENCES

Bornstein,B.J. *et al.* (2008) LibSBML: an API library for SBML. *Bioinformatics*, **24**, 880.

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.

Li,C. *et al.* (2010) BioModels database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst. Biol.*, **4**, 92.

Neal,M.L. and Sauro,H.M. (2012) SBML Reaction Finder: retrieve and extract specific reactions from the BioModels database. *Nat. Precedings*, [Epub ahead of print, doi:10.1038/npre.2012.7028.1, March 27, 2012].

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

Teusink,B. *et al.* (1998) The danger of metabolic pathways with turbo design. *Trends Biochem. Sci.*, **23**, 162.