BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 18 2010, pages 2352–2353
doi:10.1093/bioinformatics/btq425

Systems biology

Advance Access publication July 29, 2010

MetNetMaker: a free and open-source tool for the creation of novel metabolic networks in SBML format

Thomas Forth*, Glenn A. McConkey and David R. Westhead
Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK

Associate Editor: John Quackenbush

ABSTRACT

Summary: An application has been developed to help with the creation and editing of Systems Biology Markup Language (SBML) format metabolic networks up to the organism scale. Networks are defined as a collection of Kyoto Encyclopedia of Genes and Genomes (KEGG) LIGAND reactions with an optional associated Enzyme Classification (EC) number for each reaction. Additional custom reactions can be defined by the user. Reactions within the network can be assigned flux constraints and compartmentalization is supported for each reaction in addition to the support for reactions that occur across compartment boundaries. Exported networks are fully SBML L2V4 compatible with an optional L2V1 export for compatibility with old versions of the COBRA toolbox.

Availability and implementation: The software runs in the free Microsoft Access 2007™ Runtime (Microsoft Inc.), which is included with the installer and works on Windows XP SP2 or better. Full source code is viewable in the full version of Access 2007 or 2010. Users must have a license to use the KEGG LIGAND database (free academic licensing is available). Please go to www.bioinformatics.leeds.ac.uk/~pytf/metnetmaker for software download, help and tutorials.

Contact: pytf@leeds.ac.uk

Received on February 18, 2010; revised on June 18, 2010; accepted on July 15, 2010

1 INTRODUCTION

Flux-balance analysis (FBA), as described in Varma and Palsson (1994), with an updated description in Kauffman *et al.* (2003) and clarified in a primer by Orth *et al.* (2010), seeks flux solutions corresponding to steady-state metabolite concentrations, $S \cdot v = 0$, where S is the stoichiometric matrix of a reconstructed metabolic network and v is a vector of the flux through each reaction in the network. Stoichiometric matrices typically have more reactions (columns of S) than metabolites (rows of S) meaning that a solution v can be any linear combination of many vectors in the null space of S. The linear combination of these vectors that gives a single optimal vector can be found by linear programming; often by maximizing the creation, in the correct proportions, of the metabolites defined in an experimentally measured biomass function, or more generally maximizing or minimizing any other chosen objective function. Additional constraints on the flux through certain reactions can be added—often with reference to known uptake rates of carbon and nitrogen sources—providing a further link between the computational model and experimental results. The COBRA toolbox (Becker *et al.*, 2007), which applies techniques set out in Price *et al.* (2003), is a commonly used tool for FBA. In addition to many other features, this software can perform constraints-based FBA on Systems Biology Markup Language (SBML; Hucka *et al.*, 2008) format metabolic networks containing additional information on reaction flux contraints.

Many existing whole-organism metabolic network reconstructions have been created using the Pathway Tools (Karp *et al.*, 2002) software based on the BioCyc pathway genome database collection and the Metacyc compound/reaction/pathway naming convention (Caspi, 2006). These reconstructions are often unsuitable for FBA, since they contain holes due to a lack sufficient manual curation. Manipulation of these networks, such as adding new transport reactions or adding/removing pathways, and quickly exporting SBML format models is cumbersome and limits the usefulness of the software for rapid network design and testing.

FBA-ready metabolic network models are generally recreated, as outlined in Thiele and Palsson (2010). Recent examples of this approach are for the pathogens: *Leishmania major* (Chavali *et al.*, 2008) and *Salmonella typhimurium* (Raghunathan *et al.*, 2009) both of which were built in Simpheny™ (Genomatica Inc.), a well developed but closed source and expensive software package that uses a custom compound/reaction naming convention.

2 THE METNETMAKER APPLICATION

MetNetMaker uses the Kyoto Encyclopedia of Genes and Genomes (KEGG) LIGAND compound/reaction naming convention (Goto *et al.*, 2002) and provides a graphical interface dedicated to FBA-ready metabolic network creation. The software is written in Microsoft Access 2007 and uses Structured Query Language style queries for database lookup and Visual Basic for Applications code for the user interaction. The application structure and source code is editable within the full version of Access, allowing users to customize the software if necessary. Excel™ (Microsoft Inc.) files (*.xls) are used to save pathways during network development and can be edited, and added to, outside of MetNetMaker, while still retaining compatibility with the software.

The software has two main tabs for the user to interact with: the Reaction Creator and the Reaction Picker. The Reaction Picker contains a third key element of the application, the Selected Reactions table, which can be opened in a separate window. A diagram of the application structure, showing the role of each application component, is shown in Figure 1.

The Reaction Creator helps to create custom reactions, either those not defined in KEGG or special reactions such as biomass/objective reactions, transport reactions (movement of a metabolite between compartments) or boundary reactions (movement of a metabolite between the model and an external metabolite pool), and adds them to the reaction database. Custom reactions are defined in terms

*To whom correspondence should be addressed.

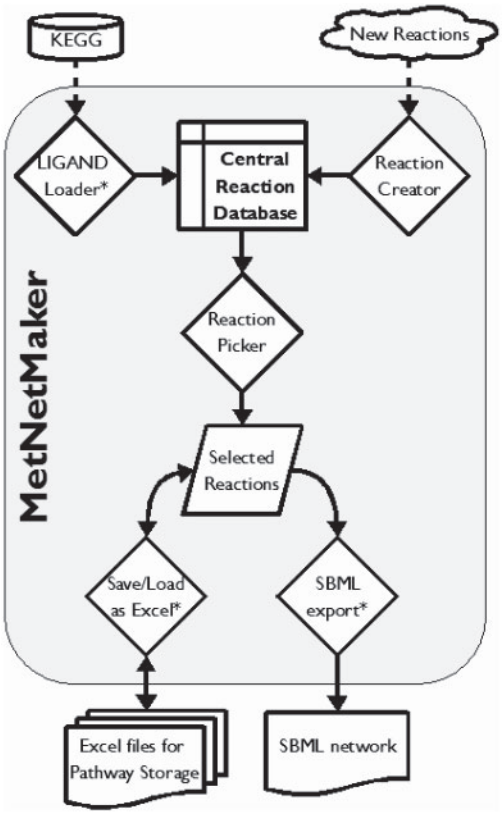

Fig. 1. The application structure of MetNetMaker. Processes marked with an asterisk are accessible from the application's toolbar.

of KEGG compound IDs and are checked for similarity with the existing reactions to avoid duplication. Special reactions have the compartment of each metabolite defined during creation, so that they can occur across compartments within the final model.

The Reaction Picker allows users to browse the reaction database and select reactions using filters for Enzyme Classification (EC) number, pathway, reaction name or compounds involved in the reaction. Filters can be applied in parallel to find a precise reaction, which is then added to the Selected Reactions table. A compartment must be chosen for each standard reaction, and a single reaction can exist in multiple compartments if required. Special reactions are marked so that each metabolite occurs in the compartment it was defined in when the reaction was created. Optionally, the user may specify flux constraints through any reaction and mark any number of reactions as objective functions. An EC number can also be defined for each reaction: either chosen from a suggested list of EC numbers known to be associated with the reaction or taken from the user input. Where a reaction is catalysed by an unknown enzyme, the EC number can be marked as unknown and reactions can be defined as occurring spontaneously in a similar way.

The Selected Reactions table holds all the reactions in the current network and reactions in this table can be edited or deleted until the reactions are ready to be saved. Editable collections of reactions are saved as Excel™ (*.xls) format files and can later be reloaded into the Selected Reactions table and joined to create a partial or whole-organism metabolic network ready for export as an SBML file. A further feature of the Selected Reactions table is that the preferred direction of each reaction can be set by the user. Although unimportant computationally (reactions are considered reversible unless their flux is constrained to zero in either direction), this allows for easier visualization of networks in third-party SBML viewers and allows MetNetMaker to create a list of dead-end compounds. Dead-end compounds are those that are created but not used—or used but not created—within a given compartment and strongly suggest holes in the network or the lack of adequate special reactions.

In addition to model import/export, showing dead-end compounds and application navigation, the MetNetMaker toolbar has two further roles. First, it allows the import and export of custom reactions and compounds that have been defined using the Reaction Creator. Second, it contains the controls to download and parse the KEGG LIGAND database and to populate MetNetMaker's central reaction database with this parsed information. Help and video tutorials that illustrate these uses are available on the MetNetMaker website.

3 SUMMARY

MetNetMaker helps in the creation and rapid manipulation of SBML format metabolic networks with reactions and compounds defined as in the KEGG LIGAND database. It allows comprehensive flux-rate constraints to be added to individual reactions and helps the user add the biomass/objective, transport and boundary reactions required for network analysis using the COBRA toolbox. A built-in tool warns of dead-end compounds within the network and helps the user create custom reactions such as those required to join metabolism across internal compartment divisions. The simplicity and narrow aim of the software may make it more capable and easier to use for the creation and manipulation of partial or whole-organism metabolic networks than Pathway Tools. Because it uses the well-documented compound/reaction definitions of the KEGG LIGAND database it creates metabolic networks that are more portable and accessible than those created in Simpheny™.

ACKNOWLEDGEMENTS

The authors wish to thank Marcus Herrgard and Liz Webb for their advice and encouragement when starting the project.

Conflict of Interest: none declared.

REFERENCES

Becker,S.A. *et al*. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat. Protoc*., **2**, 727–738.

Caspi,R. (2006) MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.*, **34**, D511–D516.

Chavali,A.K. *et al*. (2008) Systems analysis of metabolism in the pathogenic trypanosomatid Leishmania major. *Mol. Syst. Biol.*, **4**, 1–19.

Goto,S. *et al*. (2002) LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.,* **30**, 402–404.

Hucka,M. *et al*. (2008) *Systems Biology Markup Language (SBML) Level* 2: *Structures and Facilities for Model Definitions*. Nature Precedings. Available at http://dx.doi.org/10.1038/npre.2008.2715.1.

Karp,P.D. *et al*. (2002) The Pathway Tools software. *Bioinformatics*, **18**, 225–232.

Kauffman,K.J. *et al*. (2003) Advances in flux balance analysis. *Curr. Opin. Biotechnol.*, **14**, 491–496.

Orth,J.D. *et al*. (2010) What is flux balance analysis? *Nat. Biotechnol.*, **28**, 245–248.

Price,N.D. *et al*. (2003) Genome-scale microbial in silico models: the constraints-based approach. *Trends Biotechnol*., **21**, 162–169.

Raghunathan,A. *et al*. (2009) Constraint-based analysis of metabolic capacity of Salmonella typhimurium during host-pathogen interaction. *BMC Syst. Biol.,* **3**, 38.

Thiele,I. and Palsson,B.Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc*., **5**, 93–121.

Varma,A. and Palsson,B. (1994) Metabolic flux balancing: basic concepts, scientific and practical use. *Nat. Biotechnol*., **12**, 994–998.