BIOINFORMATICS APPLICATIONS NOTE Vol. 27 no. 14 2011, pages 2009–2010
doi:10.1093/bioinformatics/btr308

Systems biology

Advance Access publication May 19, 2011

rBioNet: A COBRA toolbox extension for *r*econstructing high-quality *bio*chemical *net*works

Stefan Gretar Thorleifsson[1] and Ines Thiele[1,2,*]

[1]Center for Systems Biology and [2]Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Reykjavik, Iceland

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: Genome-scale metabolic networks are widely used in systems biology. However, to date no freely available tool exists that ensures quality control during the reconstruction process.

Results: Here, we present a COBRA toolbox extension, rBioNet, enabling the construction of publication-level biochemical networks while enforcing necessary quality control measures. rBioNet has an intuitive user interface facilitating the reconstruction process for novices and experts.

Availability: The rBioNet is freely available from http://opencobra.sourceforge.net.

Contact: ithiele@hi.is

Received on January 7, 2011; revised on April 7, 2011; accepted on May 7, 2011

1 INTRODUCTION

Genome-scale metabolic reconstructions play an important role in molecular systems biology as they provide a structured format for genomic, genetic and biochemical information available for a target organism (Palsson, 2006). A reconstruction often has detailed gene information, literature references and reaction/metabolite information associated in addition to an organism-specific reaction list (Thiele and Palsson, 2010). The reconstruction process is well established for metabolic networks and has been applied to >45 organisms on a genome scale. While this approach is widely used, no comprehensive, easy-to-use and free-of-charge software package exists that ensures quality of the mostly manually curated reconstruction content. Existing tools for metabolic reconstructions include MetAnnoGen (Gille *et al.*, 2007) and MEMOSys (Pabinger *et al.*, 2011). However, many researchers still use spreadsheets for the reconstruction process to assemble and monitor the reconstruction content, which have the disadvantage of missing quality control modules. A quality check-as-you-go would make the reconstruction effort more efficient and traceable.

Here, we present a freely available software package that implements important quality control and assurance measures (QC/QA), which are crucial for the construction of high-quality genome-scale biochemical networks (Thiele and Palsson, 2010).

rBioNet is embedded in the Matlab programming environment relying on many QC/QA measures already present in the COBRA toolbox (Becker *et al.*, 2007; Schellenberger *et al.*, 2011). Biochemical reconstructions, assembled with rBioNet, can be readily converted into mathematical models, and they can then be interrogated using constraint-based methods (Palsson, 2006).

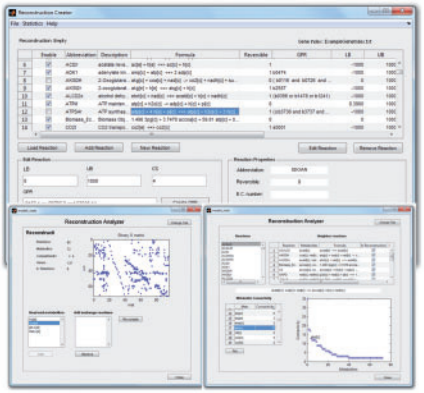

Fig. 1. Snapshot of rBioNet. Reconstruction creator and analyzer are shown.

2 METHODS

rBioNet consists of three parts: (i) a metabolite creator with associated metabolite database (MetDB); (ii) a reaction creator with reaction database (RxnDB); and (iii) a reconstruction creator (Fig. 1). To ensure consistency and quality, reactions can only contain metabolites present in MetDB and they can only be added to a reconstruction when they are present in RxnDB.

MetDB and RxnDB are stored in metab.mat and rxn.mat, respectively. The provided files within this package contain only sample metabolites and reactions, which may be expanded prior to or during any reconstruction effort.

Metabolite creator: there are three ways to create metabolites: (i) manual addition of metabolite information; (ii) loading of a tab-delimited file containing all necessary information; or (iii) loading metabolite information from other COBRA reconstructions (model structure). Required and optional information is described in the manual. When saving a new metabolite, entered information will be checked against MetDB for potential duplicate entries. This check will be done based on metabolite abbreviaton and charged formulae. No entry can have the same metabolite abbreviation but duplicated formulae may exist (e.g. for isoforms). Metabolites within MetDB are organism- and compartment independent. A new metabolite can also be created by opening, modifying and saving an existing entry as a new metabolite (under a new abbreviation).

Reaction creator: there are three ways of populating the RxnDB: (i) manually entering corresponding information; (ii) loading a tab-delimited file; or (iii) loading another COBRA reconstruction (model structure). Also, a new reaction can be added to the database by copying and modifying

*To whom correspondence should be addressed.

an existing entry. Note that in all cases the reaction(s) can only contain metabolites that exist in the MetDB. Reactions are organism independent but compartment specific. When adding a new reaction to RxnDB, the reaction will be checked for mass and charge balance. An unbalanced reaction will provoke a warning and a detailed report of unbalanced elements/charge. However, the user may choose to proceed as certain reactions are unbalanced even in high-quality reconstructions (e.g. biomass reaction, exchange reactions) (Thiele and Palsson, 2010). Subsequently, the new reaction is checked for uniqueness. This step may require some time depending on the size of RxnDB. The check includes uniqueness of reaction abbreviation and of reaction formulae. No duplicated reactions with different abbreviations are permitted. However, a reaction may re-occur in a different cellular compartment. New compartments can be readily created.

Reconstruction creator: a user can choose to create a reconstruction from scratch or to load an existing reconstruction for expansion and/or modification. Again, only reactions that are present in the RxnDB can be added to the reconstruction. All reaction information present in the RxnDB will be associated with the model when a reaction is added to the working model. The user can edit and add further information (e.g. references, notes, subsystems) to a reconstruction reaction. Furthermore, global information, such as reconstructor, reconstructed organism, version, and source of gene index, can be added to the reconstruction.

GPR creator: each reconstruction reaction can be associated with genes via gene–protein–reaction–associations (GPRs). Therefore, a gene index needs to be imported into the reconstruction creator, a tab-delimited file describing key properties of the organism's gene annotations. A GPR is created using Boolean logic, 'AND', 'OR'.

Reconstruction analyzer: basic information about the reconstruction is provided and dead-end metabolites are highlighted. This analyzer provides real-time feedback on reconstruction progress and potential missing reactions. A detailed description can be found in the manual.

Export a reconstruction: a reconstruction can be saved in Matlab format as a model structure. All associated information will be stored. This model structure can be used for further analysis within the COBRA toolbox or exported to other formats, including spreadsheet and Systems Biology Markup Language (SBML) using the COBRA toolbox (writeCBmodel).

Import a reconstruction: a reconstruction can be imported into rBioNet using the reconstruction creator from Matlab format (model structure), spreadsheet or SBML (via COBRA toolbox). If a reconstruction had been constructed with rBioNet, all associated information will be loaded. When loading an existing reconstruction into the reconstruction creator, the user is asked to provide a file location of the gene index. This file is optional but if provided, information is accessible through the reconstruction creator.

Reconstruction versus model: a reconstruction provides a detailed list of biochemical functions encoded by an organism's genome and serves as a basis for condition-specific models. rBioNet is an environment to assemble a reconstruction. The constraints on the exported model structure have to be adjusted as desired using the COBRA toolbox to obtain a condition-specific model. We do not consider reaction directionality, which are contained in the exported structure, as condition-specific constraints for this purpose.

3 IMPLEMENTATION

rBioNet was encoded in Matlab programming environment (MathWorks, Inc.) requiring the COBRA toolbox, version 1.3 or higher (Becker *et al.*, 2007; Schellenberger *et al.*, 2011), to be installed. A user-friendly interface faciliates the use of this tool by novices in programming and/or metabolic network reconstruction. A comprehensive manual is provided including installation information and frequently asked questions.

4 DISCUSSION

rBioNet is freely available permitting the reconstruction of high-quality, genome-scale biochemical networks consistent with established standard procedures (Thiele and Palsson, 2010). In particular, it ensures QC/QA measures required for publication-level reconstructions. rBioNet is embeded within the COBRA toolbox enabling iterative approach of reconstruction, validation and debugging. In particular, we put emphasis on an intuitive interface that allows novices to Matlab to perform high-quality metabolic reconstructions. The database allows customized inclusion of metabolites and reactions in the corresponding databases but thanks to its import function, online databases, e.g. KEGG ligand database (Kanehisa *et al.*, 2010), may be incorporated into MetDB and RxnDB.

A current shortcoming of rBioNet includes missing database binding for the metabolite, reaction and reconstruction creator. Nevertheless, the current setup enables simultaneous use and modification by multiple users. Future extensions may include the availability of a globally available and accessible database for metabolites and reactions, which could be directly used by any rBioNet user and thus, would permit ultimatively compability between biochemical models. The creation of such database would require significant work and would need to be coordinated with the existing biochemical databases and efforts.

ACKNOWLEDGEMENT

The authors would like to thank M.K. Aurich, R.M.T. Fleming, A. Heinken, J. Nogales, O. Rolfsson and S. Sahoo for testing and valuable discussion.

Funding: Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery through Advanced Computing program (grant DE-SC0002009); European Research Council (grant proposal no. 232816); Rannis Nyskopunarsjodur namsmanna.

Conflict of Interest: none declared.

REFERENCES

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

Gille,C. *et al*. (2007) METANNOGEN: compiling features of biochemical reactions needed for the reconstruction of metabolic networks. *BMC Syst. Biol.*, **1**, 5.

Kanehisa,M. *et al*. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355.

Pabinger,S. *et al*. (2011) MEMOSys: Bioinformatics platform for genome-scale metabolic models. *BMC Syst. Biol.*, **5**, 20.

Palsson,B.Ø. (2006) *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, New York, NY, USA, pp. 10011–4211.

Schellenberger,J. *et al*. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox v2.0. *Nat. Protoc.*, in press.

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