BIOINFORMATICS APPLICATIONS NOTE Vol. 31 no. 4 2015, pages 610–611
doi:10.1093/bioinformatics/btu685

Systems biology

Advance Access publication October 17, 2014

MOST: a software environment for constraint-based metabolic modeling and strain design

James J. Kelley[1], Anatoliy Lane[1], Xiaowei Li[1], Brahmaji Mutthoju[1], Shay Maor[1], Dennis Egen[1] and Desmond S. Lun[1,2,*]

[1]Center for Computational and Integrative Biology and Department of Computer Science, Rutgers University, Camden, NJ 08102, USA and [2]Phenomics and Bioinformatics Research Centre and School of Mathematics and Statistics, University of South Australia, Mawson Lakes, SA 5095, Australia

Associate editor: Jonathan Wren

ABSTRACT

Summary: MOST (metabolic optimization and simulation tool) is a software package that implements GDBB (genetic design through branch and bound) in an intuitive user-friendly interface with excel-like editing functionality, as well as implementing FBA (flux balance analysis), and supporting systems biology markup language and comma-separated values files. GDBB is currently the fastest algorithm for finding gene knockouts predicted by FBA to increase production of desired products, but GDBB has only been available on a command line interface, which is difficult to use for those without programming knowledge, until the release of MOST.

Availability and implementation: MOST is distributed for free on the GNU General Public License. The software and full documentation are available at http://most.ccib.rutgers.edu/.

Contact: dslun@rutgers.edu.

Received on May 29, 2014; revised on October 1, 2014; accepted on October 14, 2014

1 INTRODUCTION

Many software packages are currently available to create constraint-based models of metabolism, load existing models, export models and run analyses on these models (Lakshmanan *et al*., 2014) to predict the production of desired compounds by microbes under genetic manipulations (Egen and Lun, 2012). This software differs in data input type and display of results, where input type ranges from command line to spreadsheet, and results display as text or in spreadsheet form. Many packages can load the commonly used systems biology markup language (SBML) (Hucka *et al*., 2003) format, and some can load text files, microsoft excel files or other formats (Lakshmanan *et al*., 2014). Many packages also export SBML or other formats (Lakshmanan *et al*., 2014). In general all the software packages implement FBA (Orth *et al*., 2010), and some implement OptKnock (Burgard *et al*., 2003), evolutionary algorithms (Patil *et al*., 2005), GDLS (genetic design through local search) (Lun *et al*., 2009) or other analyses.

Much of this software is command line, which requires programming knowledge, deterring potential users. Because very few software packages exist that implement an editable spreadsheet type interface, Excel is commonly used to build models. Excel can be used to run analyses on models, but it requires programming knowledge to do so. In addition, the solver used in Excel may not be able to handle large models (Lakshmanan *et al*., 2014). There clearly exists a need for a program that allows models to be built in a user-friendly intuitive Excel-like interface, and that can run analyses on these models without requiring programming knowledge.

Genetic design through branch and bound (GDBB) is currently the fastest technique for finding a set of gene knockouts that gives the best production flux of a particular compound as predicted by FBA (Egen and Lun, 2012). A synthetic objective vector, the vector g in Equation 2 from Egen and Lun (2012) and Equation 3 from Lun *et al*., (2009), is used by GDBB to define the substance for which production is to be maximized. Because GDBB is only available as an implementation in MATLAB, which is command line and difficult to use for those without programming knowledge, there is a need for GDBB to be implemented in a more user-friendly software package. MOST (metabolic optimization and simulation tool) provides an intuitive, user-friendly interface that can be used for creating, loading and exporting models, and running analyses on these models. MOST implements GDBB, FBA, E-Flux2 and simplified Pearson correlation with transcriptomic data (SPOT), and supports SBML and CSV files. E-Flux2 and SPOT are methods for integration of transcriptomic data into constraint-based models (M. Kim *et al*., submitted for publication).

MOST stands out for its Excel-like spreadsheet editing functionality that includes functions that are specialized for model building and editing such as a Reaction Editor, which can be used to construct reaction equations from an existing table of metabolites.

2 DESCRIPTION OF TOOL

MOST is a standalone software package where constraint-based models of metabolism can be created, and existing models can be loaded, edited and saved. The main portion of the interface consists of two spreadsheets: the Reactions table and the Metabolites table. MOST has a tree where links to the original loaded model and analyses run on the model are located, and a console where results are displayed (see Fig. 1). The results of analyses in spreadsheets and in the console can be saved. MOST has been

*To whom correspondence should be addressed.

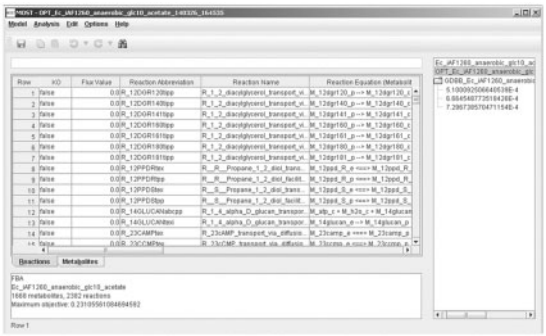

Fig. 1. MOST user interface with title bar, menu bar, toolbar, formula bar, tables, and console (from upper left to lower left), and tree on right

extensively tested on Windows XP, 7 and 8. Linux and Macintosh releases are also available. MOST requires a Mixed Integer Linear Program solver to run FBA, E-Flux2 and GDBB. MOST offers a choice of solvers, either Gurobi solver (Gurobi Optimization, Houston, TX, USA), which is free for academic use, or GLPK (GNU Linear Programming Kit; http://www.gnu.org/software/glpk/), a free solver. MOST also requires a Quadratic Solver for E-Flux2, either Gurobi or Ipopt (Interior Point OPTimizer; http://openopt.org/IPOPT), and a Nonlinear solver for SPOT, Ipopt.

2.1 Ease of use

MOST provides an intuitive user interface, and extensive documentation on its website describing software requirements, installation, model building and editing, loading and saving of models, and running analyses. The Reaction Editor makes it easy to create new reactions from a table of existing metabolites, since entering text generates selectable suggestions. Using the Reaction Editor ensures correct spacing and syntax. If a metabolite exists in a reaction equation but is not present in the metabolites table, it will be automatically added. It is easy to 'clean up' a model, i.e. remove metabolites that do not participate in any reactions, by using the 'delete all unused metabolites' menu item. Unused metabolites or participating reactions can be highlighted. MOST provides error checking and helps the user to fix errors.

2.2 Genetic design through branch and bound

GDBB can be run from the analysis menu by selecting the GDBB item. There must be a non-zero entry in the synthetic objective column to get meaningful results. A dialog will be displayed allowing a user to select number of knockouts, indefinite or finite time and start analysis. Each solution generated will add an entry to the GDBB folder in the tree, and a new entry to the console. MOST is distributed with the test files that run through the test cases mentioned by Egen and Lun (2012).

2.3 Model exchange

MOST has been extensively tested for loading SBML files, having successfully loaded approximately 100 different SBML models. MOST writes SBML files and these SBML files are read correctly by OptFlux (Rocha, 2010) and flux analysis and modeling environment (Boele, 2010). MOST can also export to CSV files, which are Excel readable. MOST has dialogs that help the user to match CSV file columns to the corresponding columns in reactions and metabolites tables. MOST auto-matches columns, but selected columns can be changed. MOST can also correct the syntax of a model so that it conforms to SBML specifications, and write the corrected model as an SBML file.

3 SUMMARY

MOST separates itself from other software packages for constraint-based modeling because it is that it is the only software package that implements GDBB, the fastest method for finding gene knockouts predicted by FBA to have high output flux of desired products, E-Flux2, and SPOT in an intuitive, easy to use interface with Excel-like editing functionality. MOST provides a Reaction Editor, which prevents syntax errors when editing reaction equations. In addition, MOST implements FBA and has the ability to load and save SBML and CSV files.

ACKNOWLEDGEMENTS

The authors would like to thank Trent Kroeger for his contributions in starting off the MOST project, and Zhiyuan Zhang for work on FBA.

Conflicts of interest: None declared.

REFERENCES

Boele,J. *et al.* (2012) FAME, the flux analysis and modeling environment. *BMC Syst. Biol.*, **6**, 8.

Burgard,A.P. *et al.* (2003) OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.*, **84**, 647–657.

Egen,D. and Lun,D.S. (2012) Truncated branch and bound achieves efficient constraint-based genetic design. *Bioinformatics*, **28**, 1619–1623.

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.

Lakshmanan,M. *et al.* (2014) Software applications for flux balance analysis. *Brief. Bioinform.*, **15**, 108–122.

Lun,D.S. *et al.* (2009) Large-scale identification of genetic design strategies using localsearch.Mol. *Syst. Biol.*, **5**, 296.

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

Patil,K.R. *et al.* (2005) Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics*, **6**, 308.

Rocha,I. *et al.* (2010) OptFlux: an open-source software platform for *in silico* metabolic engineering. *BMC Syst. Biol.*, **4**, 45.