BIOCHEMICAL NETWORKS COMPARISON TOOL

ZINTIS ROVE, MARTINS MEDNIS, ILONA ODZINA

Faculty of Information Technology, Latvia University of Agriculture, Latvia
zintis.rove@gmail.com

Abstract: *In genome-size model (reconstruction) case, each model may contain several hundreds of reagents and reactions. To be semantically comparable and then merged, biochemical network models require strict standardization. Unfortunately, the standardization of biochemical substances is not solved, therefore comparison of models made by different authors is not a trivial task, because the names of reagents and reactions often have synonyms. Without a computerized algorithm, comparison of models is time consuming process. In this paper a model is viewed as two interrelated mathematical sets. An algorithm to compare two stoichiometric models and application example is presented.*

Keywords: stoichiometric models, comparison, tool

Introduction

The genome-size models are available in different formats. Mostly it is SBML (System Biology Markup Language) file format. The SBML enables the model use of multiple software tools without having to rewrite models to tools inner format. Also, these models can be easily published and shared to other researchers to different software environment (Andrew Finney, Michael Hucka, 2003). Still the comparison of models is complicated as the standardization of substances and reactions is not a completed task. It is possible to compare structures of models visually (Kostromins and Stalidzans, 2012) or parameters of the structure (Rubina and Stalidzans, 2010). Still that is not enough to compare the scope of models or join the models as equal metabolites and reactions have to be recognized for that purpose.
To read, write and manipulate with SBML models, the LibSBML application programming interface (API) library is available. The LibSBML API supports different SBML levels and versions. LibSBML is written in C++ and provides language bindings for Common Lisp, Java, Python, Perl, MATLAB and Octave, and includes many features that facilitates adoption and use of both SBML and the library. Developers can embed LibSBML in their applications, saving themselves the work of implementing their own SBML parsing, manipulation, and validation software (Benjamin et al., 2008).
The BiGG database is systematized biochemically, genetically and genomically structured knowledgebase of metabolic reconstructions. In BiGG database models do not contains gaps or incorrect annotations, preventing errors in model predictions. The BiGG knowledgebase is designed to work with COBRA toolbox (Schellenberger et al. 2011).
BiGG two main functions are content browsing and exporting whole reconstructions as SBML files, which are specifically designed to work with Matlab COBRA toolbox and System Biology Research Tool. Using COBRA toolbox, the SBML file exported from BiGG database may be imported as a network data structure into Matlab (Schellenberger et al., 2010).
The COBRA toolbox allows predictive computations of both steady-state and dynamic optimal growth behavior, the effects of gene deletions, comprehensive robustness analyses, sampling the range of possible cellular metabolic states and the determination of network modules. Functions enabling these calculations are included in the toolbox, allowing a user to input a genome-scale metabolic model distributed in SBML format (Beker et al., 2007).
COBRA toolbox can import models in various formats, including SBML, but the Excel spreadsheet with defined columns is preferred.
COBRA compatible model comparison using web based tool ModeRator has been described by Mednis et al, 2012.
To show application of the algorithm, the model comparison software prototype has been developed. The software comparison process overview is shown in Figure 1. Comparison software supports model file format (*.xls*) of COBRA toolbox. COBRA Excel file of model contains one sheet of information about all reactions and another sheet about all metabolites that are used in these reactions. Columns have specific order and each of them contains specific information. To compare reactions, the necessary columns are used – abbreviation, description, reaction and E.C number column. From metabolites sheet, abbreviation, descriptions and chemical formulas are used.

Materials and methods

Reactions and metabolites can be defined as lists, but to simplify mathematical notation, further they are viewed as sets.
Each model contains of 2 mathematical sets *R* and *M* (1):

$$Model = MO = \{R, M\} \quad (1)$$

Where set *R* contains all model reactions (2), but set *M* contains all metabolites (3):

$$R = \{r_i | i = 1..m\} \quad (2)$$

$$M = \{m_i | i = 1..n\} \quad (3)$$

Each reaction can contain from one to several reactants *S* and products *P*:

$$S = \{S_1, S_2, \dots, S_l\} \quad (4)$$

$$P = \{P_1, P_2, \dots, P_l\} \quad (5)$$

All model reactants (6) and products (7) are elements that includes in set *M*:

$$\forall S_i \in M \quad (6)$$

$$\forall P_i \in M \quad (7)$$

Two models can be described following:

$$Model\ 1 = MO_1 = \{R_1, M_1\} \quad (8)$$

$$Model\ 2 = MO_2 = \{R_2, M_2\} \quad (9)$$

Two models can be compared by following coherence:

$$\text{If } \forall r_i \in MO_1 \text{ and } \forall r_j \in MO_2 \text{ then } r_i \perp r_j \Longrightarrow MO_1 \perp MO_2$$

Formulation of the problem:

1. are all models equal
2. does all models have the same reactions
3. are two reactions equal

An algorithm has been developed for two models comparison. It has 4 steps:

1. In the first step each reaction metabolite abbreviations are replaced by its chemical formula. This process has been done to all model reactions. In this step H, H_2O and another user defined metabolites are removed.
2. Each model reaction is divided into two parts – left and right formula parts. The reaction parts are divided by "<=>" or "->" symbols.
3. All model M_1 reactions are compared with M_2 reactions:
   - If metabolites count between two reactions is different, automatically those are not equal.
   - Each M_1 reaction left side metabolites are compared with M_2 left side metabolites. The same is done with the right reaction side metabolites.
   - Reactions are compared by E.C numbers. If reaction R_a E.C numbers are equal to reaction R_b ones then E.C numbers are considered as equal. Also if one of those reactions doesn't have an E.C number, those are supposed to be equal.
4. Duplicate reactions in the same model are found by comparing reactions abbreviations – a reaction abbreviation is compared to all other reactions abbreviation.

To test algorithm application, two COBRA file format models of Zymomonas mobilis bacteria are used:

- test model A which is based on Singapore reconstruction (Widiastuti et al., 2011).
- test model B which is based on Korea models reconstruction (Lee et al., 2010).

Test model one contains 747 reactions and 601 metabolites, but test model two contains 600 reactions and 614 metabolites, although both models represent the same bacteria.

To compare these models, the developed model comparison software prototype ModelComparator is used.

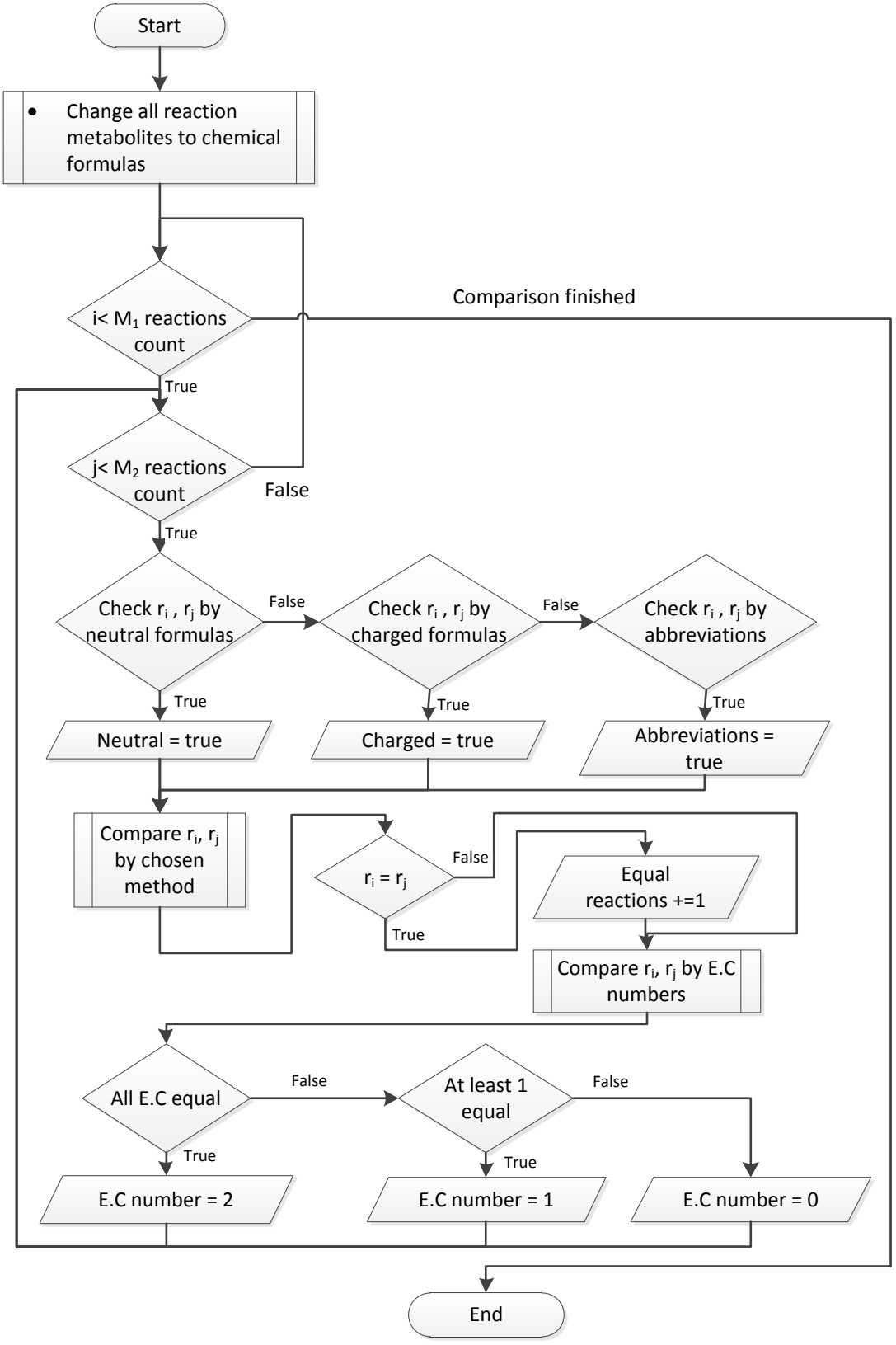

Fig.1. **Two model comparison process in developed comparison software prototype.**

Results and discussion

The comparison software prototype resulted that test model A has 33.20% (comparing by neutral formulas) equal reactions with test model A Table 2. Using default option to remove all H_2O and H metabolites from reactions the ModelComparator results more equal reactions Table 2 than performing comparison without that option Table 1:

Table 1

Test models comparison summary

	By charged formulas	By neutral formulas	By abbreviation
Equal reaction overriding E.C number	111	189	0
Equal reactions by E.C number	83	122	0
At least 1 equal E.C	3	10	0
Reversed reactions	36	79	0
Equality %	14,86 %	25,30 %	0 %

Table 2

Test models comparison summary – H_20, H ignored

	By charged formulas	By neutral formulas	By abbreviation
Equal reaction overriding E.C number	139	248	0
Equal reactions by E.C number	107	174	0
At least 1 equal E.C	3	11	0
Reversed reactions	36	79	0
Equality %	18,61 %	33,20 %	0 %

Comparing by abbreviation the result is 0 %, because one of models doesn't have correct abbreviations.
After comparison is accomplished, the Model comparator results various data about models (Fig.2):

- List of all equal reactions, where detailed information between two reactions is available. The reactions with E.C numbers and chemical formulas or abbreviations have highlighted row in green color.
- Additionally, the Model Comparator shows if reaction is reversed (in "=" column).
- List of all duplicate reactions in the same model.
- List of all model metabolites.
- Two models comparison summary is available.

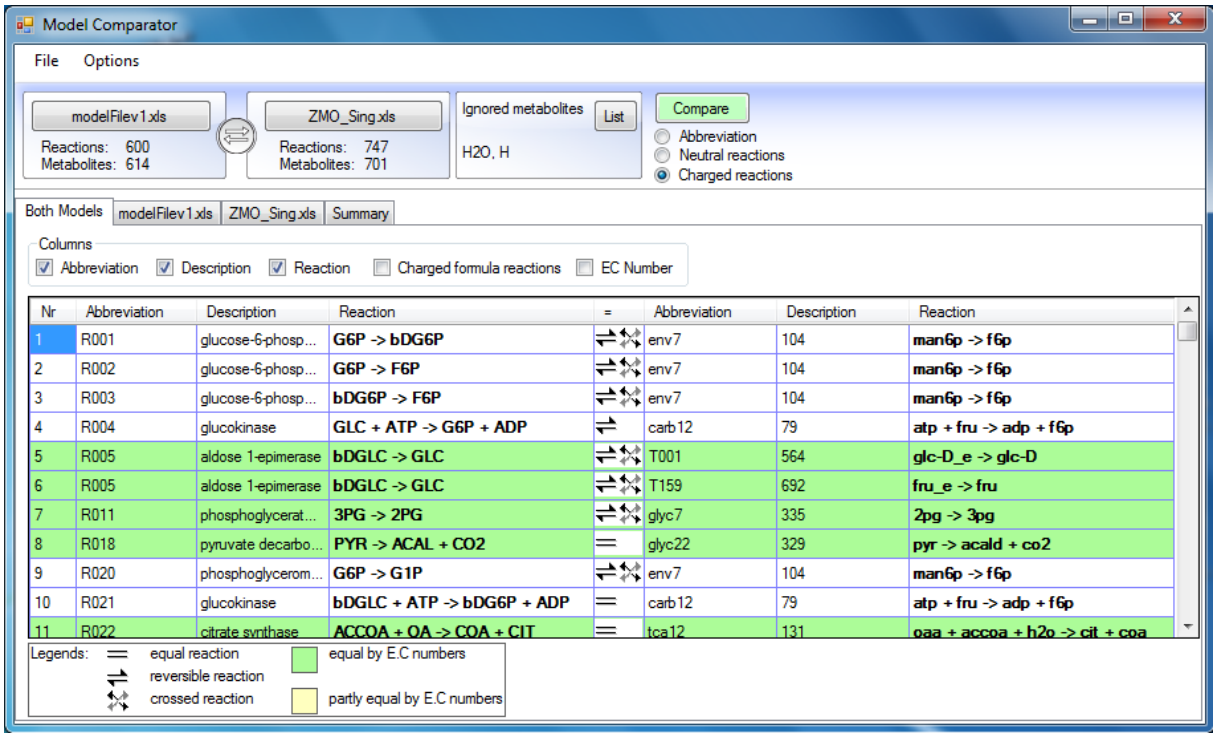

Fig.2. **The ModelComparator user interface.**

Information about two model comparison summary is available (Fig.3). The similarity between two models is shown in percents. The quick way to obtain similarity between second to first models, the switch button is available.

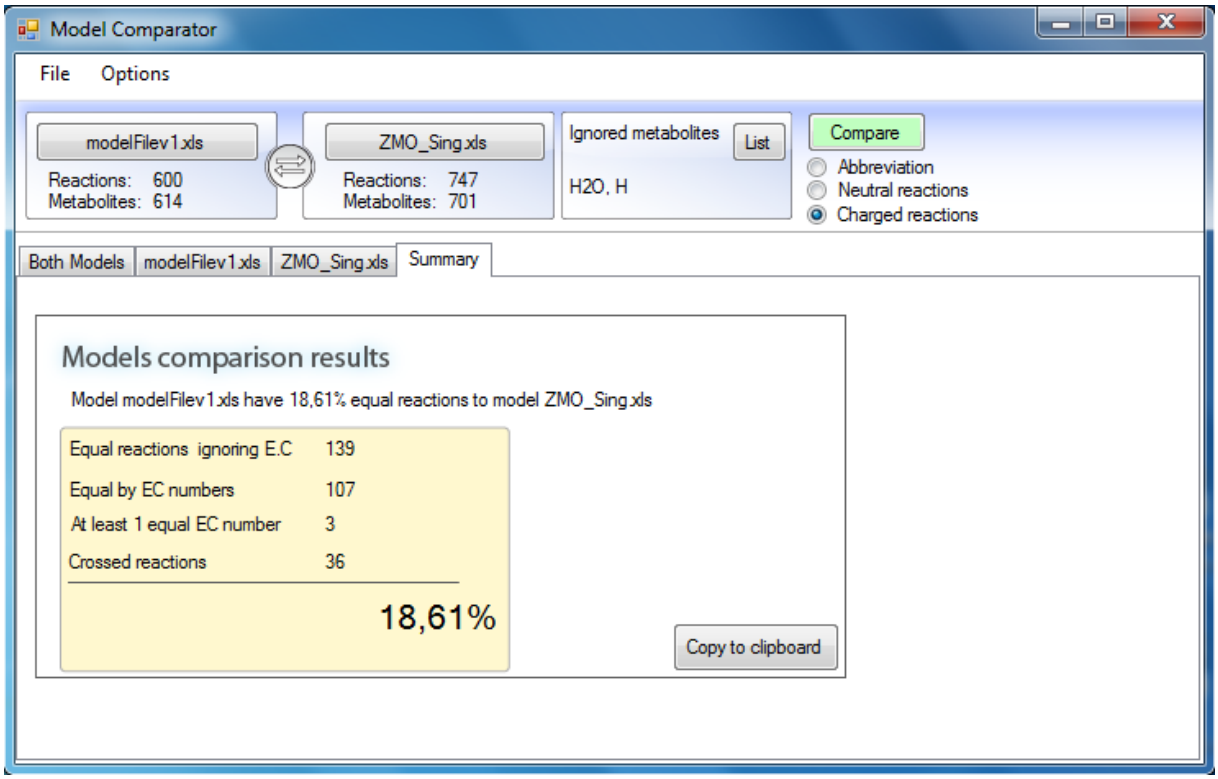

Fig.3. **Two models comparison summary results in ModelComparator.**

Conclusion

ModelComparator is desktop application that allows comparison of COBRA compatible models in *.xls* file format by comparing reactions and involved metabolites. ModelComparator has been tested on two representative genome-scale models of *Zymomonas Mobilis*. The obtained results shows that equalizing all reactions by removing water and hydrogen improves the comparison results. The use of reactants chemical formula as a mapping layer involves the risk of isomers, which can be overcome by additional E.C. numbers check.

References

Becker, S.A., Feist, A.M., Mo, M.L., Hannum, G., Palsson, B.Ø., Herrgard, M.J., 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Department of Bioengineering,University of California San Diego. Nature Protocols 2(3),727:38. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17406635

Bornstein, B.J., Keating, M.S., Jouraku, A., Hucka, M., 2008. LibSBML: An API Library for SBML. Bioinformatics 24(6): pp.880–881. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2517632/?tool=pmcentrez

Finney, A., Hucka, M., 2003. Systems Biology Markup Language (SBML) Level 2: Structures and Facilities for Model Definitions. California Institute of Technology, Pasadena. 68 p.

Kostromins, A., Stalidzans, E., 2012. Paint4Net: COBRA Toolbox extension for visualization of stoichiometric models of metabolism. BioSystems doi: 10.1016/j.biosystems.2012.03.002.

Lee, K.Y., Park, J.M., Kim, T.Y., Yun, H., Lee, S.Y., 2010. The genome-scale metabolic network analysis of Zymomonas mobilis ZM4 explains physiological features and suggests ethanol and succinic acid production strategies. Microbial Cell Factories 9(1): 94.p. Available at: http://ncbi.nlm.nih.gov/pubmed/21092328

Mednis, M., Rove, Z., Galvanauskas, V., 2012. ModeRator - a software tool for comparison of stoichiometric models, Proceedings of IEEE 7th International Symposium on Applied Computational Intelligence and Informatics, Timisoara, May 24-26, 2012. Accepted paper

Rubina, T., Stalidzans, E., 2010. Software tools for structure analysis of biochemical networks, 2010. In Proceedings of Applied Information and Communication Technologies 2010, 22.-23. April 2010. Jelgava, Latvia, pp. 33-49.

Schellenberger, J., Park, O. J., Conrad, T.M., Palsson, B.Ø., 2010. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinformatics pp.11:213

Schellenberger, J., Richard, Q., Ronan, Fleming, M.T., Thiele, I., Jeffrey, D.O., Feist, A.M., Zielinski, D.C., 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox

v2.0. Nature Protocols 6(9): pp.1290-1307. Available at: http://www.nature.com/doifinder/10.1038/nprot.2011.308

Widiastuti, H., Kim, J.Y., Selvarasu, S., Karimi, I.A., Kim, H, Seo, J.S., Lee, D.Y., 2011. Genome-scale modeling and in silico analysis of ethanologenic bacteria Zymomonas mobilis. Biotechnology and Bioengineering 108(3): pp.655-665. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20967753