ALGORITHM OF GENOME-SCALE METABOLIC ENGINEERING IMPLEMENTATION

ILONA ODZINA, AGRIS PENTJUSS

Biosystems group, Latvia University of Agriculture, Latvia
Ilona.odzina@llu.lv

Abstract: *Systems biology is a rapidly growing field that is based on building and validating in silico models of biological systems using a wealth of experimental data. These models can be applied to generate novel, testable and often quantitative predictions of cellular behavior (Becker et al., 2007). Building these models is required stoichiometric model analysis, followed by the addition of dynamic model parameters. To make it easier to resolve this task is necessary to develop an algorithm that will facilitate the metabolic engineering problems solution. The algorithm describes the metabolic engineering task progress from stoichiometric analysis of the dynamic parameters of the addition and analysis. In generating these models need stoichiometric model analysis, followed by the addition of dynamic model parameters. For an easier to resolve this task is necessary to develop an algorithm that will facilitate the metabolic engineering solutions to the task.*

Keywords: Stoichiometric analysis, model analysis, algorithm.

Introduction

To be able to study the system is the need to model the entire biological system dynamic behaviour, but the system is large and the whole system create a dynamic model is a complex and long process. Therefore, one of the options are: create a system stoihiometric model and then, part of a system that describes the problem of the creation a dynamic model. Optimized dynamic model results of a full model stoihiometric to verify whether such a solution is possible and analyze system behavior with the new data. The exact models for the could provide opportunities for biologists find out how one or more system components and variable changes may affect the entire system as a whole. Algorithm's describes the genome scale compjuter modelling the methodology of the analysis and improvement of metabolic engineering solutions without using the modified organism experimental data. Algorithm's describes the genome scale structural model and process scale models of the dynamic interaction.

Description of the algorithm

Algorithm's describes the genome scale compjutermodeling the methodology of the analysis and improvement of metabolic engineering solutions without using the modified organism experimental data. Algorithm describes the genome scale structural model and process scale models of the dynamic interaction (Fig. 2).

Step 1. Building Stoichiometric model. Create stoichiometric model of biological systems that describe salient problem (Palsson, 2009; Odzina et al., 2010; Odzina et al, 2011). Stoichiometric model create using a protocol for generating a high-quality genome-scale metabolic reconstruction. The protocol describes the procedure necessary to reconstruct metabolic networks intended to be used for computational modeling, including the constraint-based reconstruction and analysis (COBRA). This protocol describes in detail the process to generate metabolic reconstructions applicable for representatives of all domains of life (Thiele and Palson, 2010). To create stoichiometric model need using a protocol two first step (Fig.1): Draft reconstruction and Refinement of reconstruction.

Step 2. Analysis of Stoichiometric model (Matlab COBRA toolbox). A limited number of software tools and packages are available (freely and commercially), which aim at assisting and facilitating the reconstruction process. Stoichiometric analysis in this case, you are asked to use the *Matlab COBRA toolbox. Matlab COBRA toolbox* is software that silico metabolic network reconstruction of the creation and use of various methods of analysis. All metabolites and reactions data must be with every specific requirement with regard to the software methods of analysis.A variety of tools have been developed to facilitate simulations and to handle the models like COBRA toolbox *(http://gcrg.ucsd.edu/Downloads/Cobra_Toolbox)*. Tool called COBRA toolbox (Kauffman et al., 2003; Reed and Palsson, 2003) include implementations of many of the commonly used forms of constraint-based analysis such as flux balance analysis (FBA) (Orth et al, 2010), gene deletions, flux variability analysis, sampling, and batch simulations together with tools to read in and manipulate constraint-based models (Rahmanian et al., 2009*)*. Authors find it useful to perform model integrity check before importing models into COBRA toolbox. Model Comparator (Rove et al., 2012) and ModeRator (Mednis et al., 2012) allow quick identification of duplicate reactions and missing metabolites. Analysing models stoichiometric, we are interested in possible stationery positions, flow values, limiting the reaction of flow, if you are the target function. Prior to model's analysis, it lay so that it can handle the *Matlab COBRA tollbox*, this process description "A protocol for generating a high-quality genome-scale metabolic reconstruction" the third step (Thiele and Palsson, 2010). After the preparation of the analysis shall be carried out by the Step 4 of the protocol (Fig. 1). After the step 2. of the checks carried out, shall ascertain whether systems stoichiometric model is

possible steady state *(Is obtained corresponding stady state of the experimental data (reference sources))*. If it not steady state, then need perform corrections of stoichiometric model (*Step 3. Additing / correction the stoichiometric model*). Check reactions in data bases, for example: KEGG (http://www.genome.jp/kegg/), BRENDA (http://www.brenda-enzymes.info/), Transport DB (http://www.membranetransport.org/), PubChem (http://pubchem.ncbi.nlm.nih.gov/), Transport Classification Database (http://www.tcdb.org/), the model seed (http://blog.theseed.org/model_seed/). After additing and corrections execute algorithm's step 2 and repeated all analysis.
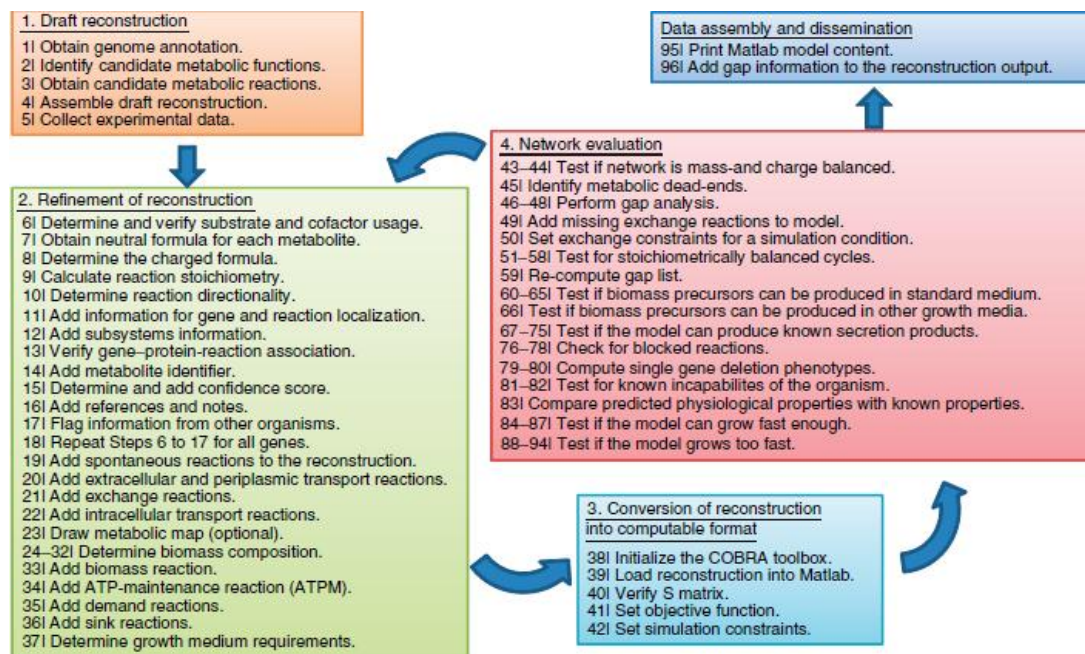

Fig.1. **Steps of protocols to creating stoichiometric model** (Thiele and Palsson, 2010)

Step 4. Addition complementary reactions to stoichiometric model. Where, after checks are carried out in Step 2 is reached steady state, to the satisfaction of the experimental data, which is taken from the literature sources, then you can add system stoichiometrijas the model with the reactions, which wants to add to achieve a goal (Modify). Stoichiometrijas model is added to the reactions and reagents.

Step 5. Analysis of Stoichiometric model (COBRA). A modified stoichiometric models analysis (Step 2) how shall also analyse the constraining the reactions of the target function (shadow price) (Palsson, 2009). The analysis of shadow prices and reduced costs, which are associated with the LP solution, can be of great help to identify metabolites or reactions that limit the rate of biomass (http://www.copasi.org/tiki-index.php?page=UserManual&structure=DocumentationNew). IF steady state after the analysis is not possible, then returns to the stoichiometric model addition, because you might need to add additional events added to the already reactions that the model could work.

To analysis can also use the programs Paint4Net. The „*Paint4Net*" is an open source extension developed for the „*COBRA Toolbox*" in „*MATLAB*" scripting language for visualization of stoichiometric models of metabolism. Both flexible visualization and information on reaction and metabolite data can be reached using „*Paint4Net*". „*Paint4Net*" functionality can be used to: (1) get lists of involved metabolites and dead end metabolites of the visualized part of the network, (2) exclude (filter) particular metabolites from representation, (3) find isolated parts of a network and (4) find running cycles when all the substrates are cut down (Kostromins and Stalidzans, 2012).

Step 6. *Create dynamic model of variable process.* If after analysis model has reached steady state, only could not be verified, in the absence of experimental data modified system, you can create a dynamic process model of intresting process. Dynamic model van creating several software. One of software is COPASI. COPASI is a software application for simulation and analysis of biochemical networks and their dynamics. COPASI is a stand-alone program that supports models in the SBML standard and can simulate their behavior using ODEs or Gillespie's stochastic simulation algorithm; arbitrary discrete events can be included in such simulations. COPASI carries out several analyses of the network and its dynamics and has extensive support for parameter estimation and optimization. (Hoops et al., 2006) COPASI provides means to visualize data in customizable plots, histograms and animations of network diagrams. Systems Biology Markup Language (SBML) is a modular language, with a core comprising a complete format that can be used alone. SBML is oriented towards representing biochemical networks common in research on a number of topics, including cell signaling pathways, metabolic pathways, biochemical reactions, gene regulation, and many others (Hucka, 2003).

The meaning of each component: *Unit definition*: A name for a unit used in the expression of quantities in a model. *Compartment:* A container of finite volume for substances. In SBML Level 1, a compartment is primarily a topological structure with a volume but no geometric qualities. *Species:* A substance or entity that takes part in a reaction. *Reaction*: A statement describing some transformation, transport or binding process that can change the amount of one or more species. For example, a reaction may describe how certain entities (reactants) are transformed into certain other entities (products). Reactions have associated rate laws describing how quickly they take place. *Parameters*: A quantity that has a symbolic name. Dynamic model have initial concentrations and kinetic parameters: the equation contains rate of reaction (Vmax, Vmaxf, Vmaxr), affinity constant (KmS, KmP), inhibition constant (Ki), concentrations of metabolites (for example: ATP, ADP). This information search of databases from references source. *Rule*: In SBML, a mathematical expression that is added to the differential equations constructed from the set of reactions and can be used to set parameter values, establish constraints between quantities, etc. (http://www.copasi.org/tiki-index.php?page=UserManual&structure=DocumentationNew) (Hoops et al., 2006)

Step 7. Finding the best stationary state. Time Course Calculation. With the time course simulation, you can calculate the trajectory for the species in your model over a given time interval. There are different methods to calculate such trajectories and depending on your model, one or several of them may be appropriate to do a time course simulation of your model. COPASI supports three different methodologies to calculate a trajectory. The first method is to do a deterministic time course simulation of your model using the LSODA (Petzold, 1983) algorithm. For systems with small particle numbers, it is sometimes better to do a stochastic simulation rather than a deterministic one. COPASI supports a method for the stochastic calculation of time series, which is called stochastic and uses the next reaction method described by Gibson and Bruck (Gibson and Bruck, 2000) Since the deterministic simulation is inappropriate for some systems but on the other hand, the stochastic simulation is too time consuming, there are some methods that try to combine the advantages of both deterministic and stochastic simulation. Most of those methods are termed hybrid methods. COPASI also includes such a hybrid method which in some systems where deterministic simulation would lead to incorrect results will give the correct time series but is still computationally less demanding than a pure stochastic simulation (http://www.copasi.org/tiki-index.php?page=UserManual&structure=DocumentationNew).

Step 8. Analysis of stability. If model have stationary state, then make Steady State Calculation**.** The best technically feasible stationary state should be found taking into account the dynamic parameters of both the cellular dynamics and industrial control system. The steady state is the state in which the state variables of the model, e.g. the species concentrations do not change in time. Mathematically this is expressed by setting the differential equations that describe the time evolution of the metabolic system to zero. This forms a system of algebraic non-linear equations. To solve them, COPASI can use a series of strategies using more than one numerical method. All calculations are done based on particle numbers and particle number rates rather than concentrations internally. The reduced model is used. The Jacobian (which is used in the Newton method and when eigenvalues of the Jacobian are requested) is calculated using finite differences. The Eigenvalues of the Jacobian are calculated using LAPACK (http://www.copasi.org/tiki-index.php?page=UserManual&structure=DocumentationNew) (Hoops, 2004).

Step 9. Prevention of Instability. Necessary to prevent instability. Are to be checked for all parameters, the reactions equations. Added to the reaction, to complement the process and ensure stability.

Step 10. Additing dynamic model results in to stoichiometric model and make analysis of Stoichiometric model. Add modified stoichiometric model the flow of values of dynamic models. Or with these values can be to ensure of modified stoichiometric model stable stationary process adjuvants. If modification stoichiometric model have steady state with this new parameters then can make modification of organisms in laboratory. But if modification stoichiometric model does not produce the results you want then return to step 7 and search the other best steady-state.
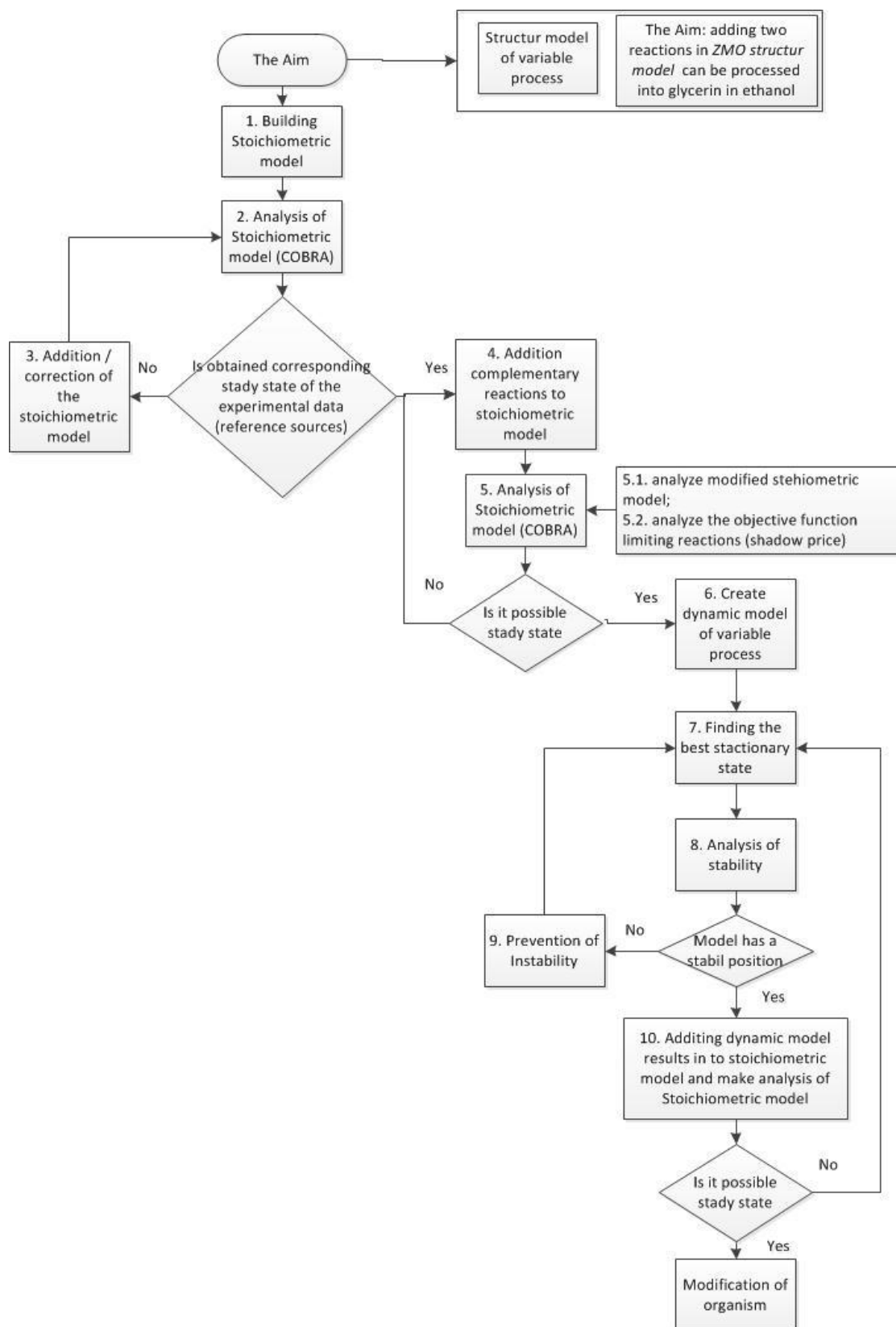
Fig. 2. Genome-scale metabolic engineering analysis solution algorithm

Conclusion

Algorithm's describes the genome scale computer modelling the methodology of the analysis and improvement of metabolic engineering solutions without using the modified organism experimental data. Algorithm's describes the genome scale structural model and process scale models of the dynamic interaction. Algorithm's consists of 10 steps.

One of systems biology industry's the problem is that it can not directly know the way of the biological systems in the molecular level. Consequently, it is necessary to carry out the modelling of biological systems. The application of the method is able to address the biological tasks in the molecular level, allows you to make modifications and verify or may be the solution. Modelling to simulate modified organism behaviour. Developed the method can be used in medicine, metabolic engineering tasks ecological problems, environmental problems, etc.

Acknowledgements

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. *Nature Protocols,* 2, pp. 727 – 738.

COPASI Development Team, 2010. COPASI Documentation Version 4.6 (Build 32) on August 3, 2010 pp 131 http://www.copasi.org/tiki-index.php?page=UserManual&structure=DocumentationNew

Gibson, M.A. and Bruck, J., 2000. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. J. Phys. Chem.. A104(9). pp.1876-1889.

Hoops, S., 2004. Stability Analysis in COPASI, Virginia Bioinformatics Institute, USA pp 1-5.

Hoops, S., Sahle,S., Gauges, R., Lee1, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., Kummer, U., 2006. COPASI—a Complex Pathway Simulator, Bioinformatics, Vol. 22 No. 24, pp 3067–3074.

Hucka,M., 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19, pp.524–531.

Kauffman,J.L., Prakash,P., Edwards,J.S, 2003. Advances in flux balance analysis. Current Opinion in Biotechnology, 14, pp. 491-496.

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

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, pp.306-311.

Odzina, I., Rubina, T., Rutkis, R., Kalnenieks, U., Stalidzans, E., 2010. Structural Model of biochemical network of Zymomonas mobilis adaptation for glycerol conversion into bioethanol. 4-th international scientific conference "Applied Information and Communication Technologies 2010", Jelgava, LLU, CD – pp.50-54.

Odzina, I., Rutkis, R., Kalnenieks, U., 2011. In silico engineering of biochemical network of Zymomonas mobilis adaption for Glycerol conversion into Bioethanol. 10th International Scientific Conference „Engineering for rural development, proceedings, Volume 10, Jelgava, LLU, pp.272-275.

Orth,J.D., Thiele,I., Palsson,B.Ø., 2010. What is flux balance analysis? Nature biotechnology, pp. 245-248.

Palsson, B.Ø, 2009. Systems Biology: Properties of Reconstructed Networks. University of California, San Diego, Cambrige university press, pp 322.

Petzold, L., 1983. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. SIAM J. Sci. Stat. Comput.. 4., pp.136 - 148.

Rahmanian,S., Schellenberger,J., Palsson,B.Ø, 2009. Customized Visualization of Metabolic Maps in the Cobra Toolbox. Available at: http://ucsdstudents.calit2.net/files/posters/2009/Sorena%20Rahmaniyan.pdf, 15. February 2011.

Reed,J.L., Palsson,B.Ø, 2003. Thirteen years of building constraint-based in silico models of Escherichia coli. Bacteriology Journal, 185, pp. 2692-2699.

Rove, Z., Mednis, M., Odzina, I., 2012. Biochemical networks comparison tool, Applied information and communication technologies, Jelgava, LLU, April 24-26, 2012, Accepted paper.

Thiele,I. and Palsson,B.Ø, 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction, Nature Publishing Group, Vol.5 No.1 pp. 93‑121.