THE CONCEPT OF METABOLIC ENGINEERING SOFTWARE TOOL

JURIJS MEITALOVS

Biosystems Group, Department of Computer Systems, Latvia University of Agriculture, Latvia
jurijs.meitalovs@gmail.com

Abstract: *Computational tools are indispensable for any modern engineering disciplines. With synthetic biology as a new scientific field, it clearly needs computational tools to help in design and modeling process. Many algorithms and tools were developed for synthetic organisms engineering. The generation of novel metabolic pathways involves compounds that exist in biological databases and compounds that are suggested by special algorithms. As a result a model with a list of all necessary changes will be obtained. There is presented a concept of data flows and a workflow of a system that can realize such functionality as stoichiometric and elementary modes analysis, optimization and novel biochemical network dynamic analysis.*

Keywords: software tool, metabolic engineering

Introduction

Over the last few years a number of methods and algorithms have been proposed for the analysis and generation of possible pathways for the production of a product from given set of starting substrates using a set of known biochemical reactions and compounds (Blum and Kohlbacher 2008; Heath et al. 2010; Finley et al. 2009; Wang et al., 2006) They are used to discover orphan reactions and enzymes, for gap-filling in metabolic pathways, new pathways generation and other functions. The generation of pathways involves compounds that exist in biological databases, as well as compounds, suggesting novel biochemical routes for these compounds. But such algorithms usually do not allow implementing the result into existing models of organism and making stoichiometric and dynamic analysis of whole organism model. Also not always it is shown, genes from which organism are used to create the metabolic pathway.

Therefore it is necessary to develop the methodology and the software package that allows creating necessary metabolic reaction list to produce a product from a substrate. It should show which minimal set of reactions should be implemented into the organisms to achieve the necessary result based on different constrains and rules, i.e. present substrates, substrate expenses, reaction rates, potential revenue etc. (Carbonell et al., 2012). Important item is the production rate of a substrate which determines the economic feasibility of industrial process (Mozga and Stalidzans, 2011a; Lee et al., 2008). The aim is to develop concept of a system that will have wide functionality, e. g. specified format support for metabolic networks definition, searching of specific chemical reactions to convert a substrate into a product, possible reaction implementation in the metabolic network, constraint based analysis (economic factors, available substrates, specific organism etc.), possibility to knockout specific reactions, conversation options to different model notations, stoichiometric and dynamic analysis, using of common notation for reactions, automatic parameter estimation from online data bases.

At this moment there are many tools that have partly similar functionality to the described tool concept. But usually they have only some specific functionality and returned data sometimes is hard to use in other tools. The problem is that not all tools use one specific data format, for example SBML (Hucka et al. 2003) or SBOL (Rodriguez et al. 2009). In addition data processing can take long time in case of using of different software, e.g., Metaroute (Blum and Kohlbacher, 2008), Copasi (Hoops et al., 2006), Cobra (Becker et al., 2007) etc. The main goal of described tool is to allow user easily work with data and models, make corrections and search data to receive necessary result.

Architecture of the system

In order to perform given tasks, it is important to choose the correct architecture and functional modules of the system. The choice of the optimal architecture in the future will allow making changes and expanding the system with additional functional modules. Also, the module type system allows changing of the structure of the system. The system will consist of different modules or plugins (Figure 1). Each of them will respond for additional functionality. It will be possible to add or delete plugins or modules to expand the functionality of the system.

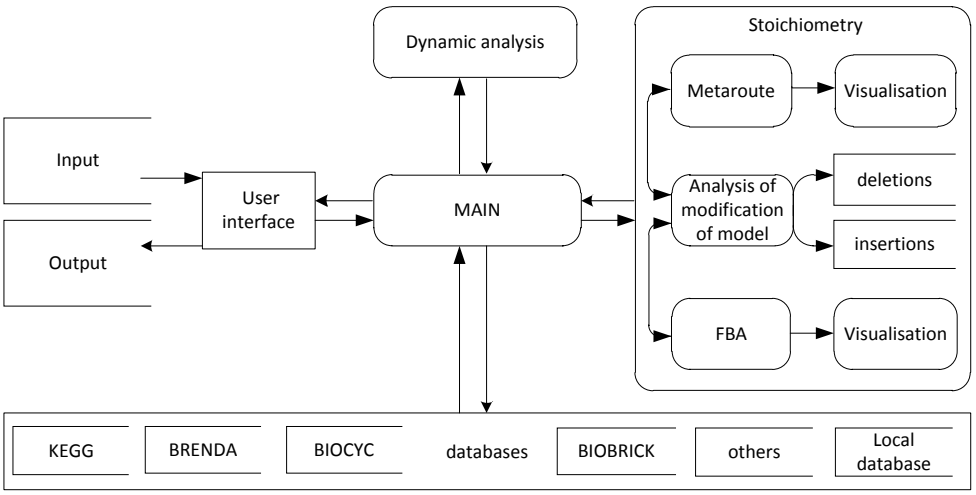

Fig. 1. **Structure of the system.**

The main notations for metabolic network model description are SBML (Hucka et al., 2003), SBOL (Rodriguez et al., 2009). It is necessary to develop separate algorithms for automatic model converting to one notation. For the user it is possible to input metabolic networks in different formats. It is possible for the user to choose the output format of metabolic network.

The main system module is responsible for all processes in the system – it provides connections with databases, reads and process data, and returns it in optimized format. It controls metabolic pathway search, stoichiometric analysis, modifications of the metabolic network, calculates flux balance analysis depending on parameters of the system, provides dynamic analysis and returns output result to the user. Main module can launch different other applications, for example Matlab to launch Cobra toolbox for FBA (Becker et al., 2007) or Paint4Net for visualization of the metabolic network.

It is possible to use separate software modules. Each component of the system should be like a separate module that can be installed or uninstalled from the system. It allows changing software components easily, it will not be necessary to change all source code. It will be possible to develop other necessary modules and add them to the system, i.e. integration with other software tools. For integration with other tool additional modules can be developed.

User interface will allow users to input data and parameters of metabolic network. Also it will allow choosing output parameters of modeling results. It will visualize modeling process and its results and show all the pathways in the system and control parameters. It will be WEB UI with client – server architecture. Users can open web interface and work with the system via Internet. It will be also possible to get a local version of the system, to install it on a local server. It should be configured for local database and Internet connection.

Communication with databases

Main functionality is separated to different parts. The first part is connection to databases using web services and APIs from databases (e.g. WSDL or SOAP). It is possible to receive information about reactions, metabolites, pathways and others. Connection to each database should be made separately to provide system safety in case of error in connection. Each database stores different type of information and it is useful to use different databases to get complete information, e.g. KEGG is a bioinformatics database containing information on genes, proteins, reactions, and pathways (Ogata et al., 1999), BRENDA is a comprehensive enzyme database (I. Schomburg et al., 2002), BioCyc is a collection of pathways and genomes databases, MetaCyc is an encyclopedia of experimentally defined metabolic pathways and enzymes that contains pathways from different organisms and metabolic reactions (Caspi et al., 2008).

To work with each database additional data class should be used to convert information from different database to one data format. That can allow to separate data reading and processing logic from other functionality of the system. To reduce data from data bases reading time it will be possible to cache data in local database, if this data is not being used for some time. Some information can be stored in the local database for a long time if this data is used often. That can allow reducing count of requests into database. The local database can be developed using different engines, e.g. the most popular free engines MySql, Postgres, Firebird, MS SQL Express.

It is necessary to go thought many steps to analyze metabolic network – find pathways from substrate metabolites to products depending on input parameters, model optimization, flux balance analysis and dynamic analysis (Palsson et al., 2003). Metabolic pathways analyzing module searches metabolic pathways from substrate to product based on input parameters included specific reaction and search results. There it is possible to choose specific microorganisms where algorithm should search for pathways (Blum and Kohlbacher, 2008b).

Performance of stoichiometric and dynamic analysis

During previous step the best pathway between product and substrate is chosen. The next step is optimization of the metabolic network. In this step new pathways are inserted into the metabolic network. It is possible to use additional algorithms to optimize metabolic network, i.e. to identify the optimal metabolic and regulatory gene deletions as well as gene over expressions that maximize biochemical production at the maximum cellular growth under transcriptional regulatory constraints (Burgard et al., 2003; Kim et al., 2010; Ranganathan et al., 2010).

The next step is FBA. One of these steps is flux balance analysis for analyzing the flow of metabolites through a metabolic network. Flux-balance analysis based on linear optimization is widely used to compute metabolic fluxes in large metabolic networks and gains increasingly importance in network curation and structural analysis. There are many algorithms and tools for flux balance analysis (e.g**.** Cobra, OptFlux, FluxAnalyzer and others (Becker et al., 2007; I. Rocha et al., 2010; Klamt et al., 2003).

Different types of visualization can be used for metabolic network visualization – all metabolic network, selected pathway, selected reaction or metabolite depending on user selection. Also it will be possible to export graphical information for external software, e.g. Graphviz or Paint4Net (Kostromins et al., 2012b). Selecting each element of metabolic network will make additional windows open with detailed information about the selected element. In the future it is planned to enable the deletion of reactions or pathways in graphical way during visualization of all metabolic networks or their part.

For a dynamic analysis software tool Copasi can be used – ready metabolic model after stoichiometric analysis in SBML format can be exported to Copasi. It is possible to use Copasi in two ways – using user interface and in batch mode (Hoops et al., 2006). In this system batch mode can be used to launch analysis in background mode and receive only the results. Using the UI of system it will be possible to change parameters of the dynamic analysis. COPASI can be used both for parameter estimation and optimization of dynamic model. Software ConvAn (Kostromins et al., 2012a) can be used to rationalize the time resources needed for optimization (Mozga and Stalidzans, 2011b). After the analysis of dynamic properties the output result can be returned to the user.

The output is an ideal metabolic network model in the format that chooses the user and the list of reactions and genes that should be added and deleted from the microorganism. It will be possible to save imported metabolic networks into the local database. It allows to use preceding results in the future work.

Conclusions

The concept of a software tool for biochemical network analysis and modeling is presented. This software tool can be useful in many research fields and it allows achieving the necessary results using limited number of tools. There are many different tools but usually they use different data format and cannot exchange results without adaptation. The described tool should implement possibilities to convert different data formats.

The concept of the described system is developed and possible realization components are discussed. Were shown functionality and algorithms that should be implemented in this system. In this paper were shown possible fields of use for such system.

The system is in a development stage and different models of the system are developed. Next steps are to connect all functionality in a common system. The novelty of such system is that in input parameters economic factors can be used, e.g. expenses of substrates and products, organism modification costs etc. That will allow researcher to find possible options for modification of microorganism and to predict results in the form of a model. It allows to decide in the early stage of modification project if it can bring in a profit and will this project be successful or not.

Acknowledgments

The author is acknowledging the financial support received from the State Education Development Agency of Latvia and the European Social Fund within the scientific research project "Establishment of Latvian interdisciplinary interuniversity scientific group of systems biology" (ref.no. 2009/0207/1DP/1.1.1.2.0/09/APIA/VIAA/128).

References

Becker, SA., 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(3), pp.727-38.

Blum, T., Kohlbacher, O., 2008. MetaRoute: fast search for relevant metabolic routes for interactive network navigation and visualization. *Bioinformatics (Oxford, England,)*, 24(18), pp.2108-9.

Burgard, A.P., Pharkya, P., Maranas, C.D., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and bioengineering*, 84(6), pp.647-57.

Carbonell, P., Fichera, D., Pandit, S.B., 2012. Enumerating metabolic pathways for the production of heterologous target chemicals in chassis organisms. *BMC Systems Biology*, 6(10).

Caspi, R., Foerster, H., Fulcher, C. a, Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., et al., 2008. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic acids research*, 36, pp.D623-31.

Finley, S.D., Broadbelt, L.J., Hatzimanikatis, V., 2009. Computational framework for predictive biodegradation. *Biotechnology and bioengineering*, 104(6), pp.1086-97.

Heath, A.P., Bennett, G.N., Kavraki, L.E., 2010. Finding metabolic pathways using atom tracking. *Bioinformatics (Oxford, England,)* 26(12), pp.1548-55.

Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., et al., 2006. COPASI-a COmplex PAthway SImulator. *Bioinformatics (Oxford, England)* 22(24), pp.3067-74.

Hucka, M., Finney, a., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Arkin, a. P., et al., 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4), pp.524-531.

Kim, J., Reed, J.L., 2010. OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC systems biolog,y* 4, p.53.

Klamt, S., Stelling, J., Ginkel, Martin, Gilles, E.D., 2003. FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics*, 19(2), p.261.

Kostromins, A., Mozga, I., Stalidzans, E., 2012a. ConvAn: a convergence analyzing tool for optimization of biochemical networks. *Biosystems*, 108 (1–3), pp. 73-77

Kostromins, A., Stalidzans, E., 2012b. Paint4Net: COBRA Toolbox extension for visualization of stoichiometric models of metabolism, *BioSystems*

Lee, S.K., Chou, H., Ham, T.S., Lee, T.S., Keasling, J.D., 2008. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current opinion in biotechnology*, 19(6), pp.556-63.

Misirli, G., Hallinan, J. S., Yu, T., Lawson, J. R., Wimalaratne, S. M., Cooling, M. T., Wipat, A., 2011. Model annotation for synthetic biology: automating model to nucleotide sequence conversion. *Bioinformatics (Oxford, England)*, 27(7), pp.973-9.

Mozga, I., Stalidzans, E., 2011a. Optimization protocol of biochemical networks for effective collaboration between industry representatives, biologists and modellers. In *9-th Annual Industrial Simulation Conference*, 91-96.

Mozga, I., & Stalidzans, E., 2011b. Convergence dynamics of biochemical pathway steady state stochastic global optimization. *In 2011 IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI)*, pp.231-235.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M., 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic acids research* 27(1): p.29-34.

Palsson, B.O., Papin, J.A. 2003. Development of network-based pathway definitions: the need to analyze real metabolic networks. *Trends in Biotechnology*, 21(5), pp.195-198.

Ranganathan, S., Suthers, P.F., Maranas, C.D., 2010. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS computational biology*, 6(4), p.e1000744.

Rocha, I., Maia, P., Evangelista, P., Vilaça, P., Soares, S., Pinto, J. P., Nielsen, J., et al., 2010. OptFlux: an open-source software platform for in silico metabolic engineering. *BMC systems biology*, 4, p.45.

Rodriguez, C., Bartram, S., Ramasubramanian, A., Endy, D. 2009. Synthetic Biology Open Language Visual (SBOLv) Specification. *DNA Sequence*.

Schomburg, I., Chang, A., Schomburg, D., 2002. BRENDA, enzyme data and metabolic information. *Nucleic acids research*, 30(1), pp.47-9.

Wang, L., Hatzimanikatis, V., 2006. Metabolic engineering under uncertainty. I: framework development. *Metabolic engineering*, 8(2) pp.133-41.