Baltic J. Modern Computing, Vol. 2 (2014), No. 3, 150-159

Reduction of Combinatorial Space of Adjustable Kinetic Parameters of Biochemical Network Models in Optimisation Task

Ivars MOZGA[1,2], Egils STALIDZANS[1,2,3]

[1]Biosystems Group, Department of Computer Systems, Latvia University of Agriculture,
Liela iela 2, LV3001, Jelgava, Latvia
[2]SIA TIBIT, Dobeles iela 10-9, LV3001, Jelgava, Latvia
[3]Institute of Microbiology and Biotechnology, University of Latvia,
Kronvalda boulv. 4, Riga, LV-1586, Latvia

ivars.mozga@gmail.com, egils.stalidzans@gmail.com

Abstract. The search for minimal set of adjustable parameters through optimising a kinetic model of biochemical networks is needed in industrial biotechnology to increase the productivity of industrial organism strains while keeping low the chance of causing unwanted side effects of implemented changes. As the search for minimal set of adjustable parameters is of combinatorial nature, the search space becomes very large even at relatively small number of parameters.
The presented approach of search space reduction is demonstrated on the example of kinetic model of yeast glycolysis. In parallel to the estimation of remaining range of optimisation potential the full search of combinations was combined with forward selection that allows reaching 91.4% of potential after optimising 625 parameter combinations. This result was reached by involving just seven out of fifteen adjustable parameters.

Keywords: Adjustable parameters, biochemical networks, design task, kinetic models, dynamic simulations, minimal set, optimisation potential.

1. Introduction

Information technologies have become one of the preconditions for the development of biotechnology; their application is determined by several reasons: storage and processing of large amounts of data, similarity analysis, data mining tasks and a variety of modelling applications. Modelling task is primarily to gather knowledge and convert them into such form it could be compared with experimental data, confirming or excluding any prior assumptions (Mauch, Buziol, Schmid, and Reuss, 2001). Models are also used for forecasting the impact of various changes in a system on the modelled process. This approach is widely used in metabolic engineering to search for changes in the system that are needed to improve its characteristics (Keasling, 2010; Sendín, Exler, and Banga, 2010).

Cellular biochemical network optimisation is a complex and time consuming task, especially with an increase in model size and number of variable parameters (Kostromins, Mozga, and Stalidzans, 2012; Mozga and Stalidzans, 2011a, 2011b; Stalidzans, Kostromins, and Sulins, 2012; Nikolaev, 2010; Xu and Wang, 2014).

However, a number of regularities are valid for the cellular biochemical networks, enabling simplification of the task to be solved. The space of task solutions can also be limited by the information available in databases on different types of experiments. Therefore, in the course of optimisation a set of varied measures must be carried out, during which the numerical optimisation methods are just one of the instruments to be applied at the right time and amount needed.

Optimisation procedure in general case must answer the following questions: what is the minimum number of variable parameters, which should be adjusted in biochemical system in order to obtain a viable organism that produces the maximum possible quantity of the product or provides the best ratio of product vs. quantity of consumed substrate (Nikolaev, 2010; Rodríguez-Prados et al., 2009; Rodríguez-Acosta, Regalado, and Torres, 1999; Sendín et al., 2010; Vital-Lopez, Armaou, Nikolaev, and Maranas, 2006)? What is the best sequence of actions that could lead us to this solution (Rodríguez-Acosta et al., 1999; Sendín et al., 2010)? Creation of such procedure would save a lot of time and resources, as well as improve the productivity of biotechnology companies (Rodríguez-Prados et al., 2009; Rodríguez-Acosta et al., 1999).

While the number of biochemical models and their size increases (Le Novère et al., 2006; Smallbone, Simeonidis, Swainston, and Mendes, 2010), the model optimisation procedures continue to develop. Mostly the stochastic global optimisation methods are applied due to their universality as they can be applied independently on the peculiarities of a model in contrast to deterministic methods (Banga, 2008). The main drawbacks of global stochastic methods are the peculiarities of their convergence to the optimal solution: 1) reaching of optimal solution is not guaranteed (Banga, 2008; Mendes and Kell, 1998) and 2) the duration of optimisation is hard to estimate (Kostromins et al., 2012; Mozga and Stalidzans, 2011a, 2011b; Nikolaev, 2010). Both mentioned drawbacks are addressed by performing parallel optimisation runs and analysing their consensus and stagnation states (Stalidzans et al., 2012; Sulins and Stalidzans, 2012; Sulins and Mednis, 2012). Parallel optimisation runs enable automation of optimisation (Bulipopa and Sulins, 2013; Sulins and Mednis, 2012) but it still requires intensive computation even if manual operations are minimised.

The procedure covering the full cycle of biochemical network design from the choice of criterion through to the suggestions for industrial tests has been proposed earlier (Mozga and Stalidzans, 2011c). This study aims to improve the proposed procedure and demonstrate its application efficiency for yeast glycolysis model as a case study.

2. Approach

2.1. Optimisation procedure actions and their sequence

The steady state optimisation procedure of biochemical network computer models enables effective use of resources during optimisation avoiding unnecessary activities or terminating the ongoing optimisation process if it is not useful. The basic principle of procedure is as follows: to obtain the maximum possible amount of information with every next step of the procedure with the least possible number of actions.

The sequence of effective optimisation actions in various biochemical processes can be described with steady state optimisation procedure of biochemical network computer

models (Mozga and Stalidzans, 2011c). The procedure generally consists of four stages: 1) choice of criterion and model, 2) determination of the optimisation potential, 3) ranking variable parameter combinations by efficiency, and 4) analysing the application of solutions.

Optimisation procedure describes actions that initially include choice of criterion and model; namely, choosing the reactions and conditions under which to complete optimisation. It results in modified numerical values of the optimisation parameters, which, through analysing the application of solutions, may be suggested for industrial testing. In the case when prolongation of optimisation actions could bear the loss because of too small improvement of the biotechnological process profitability, the ongoing optimisation is suspended and a different biochemical process is studied.

Choice of criterion and model

The first stage of the procedure intended for gaining overall knowledge of the process begins with the choice of criterion and model – what is the business task, what kind of biochemical process is chosen, which criteria have been set. Then the models are analysed for possible implementation with the set criteria by a modeller. This stage is based on the criterion set by industrialist according to the product they want to produce, and by making economic calculations establish what the minimum criterion increase is to achieve the desired business results (see Figure 1).

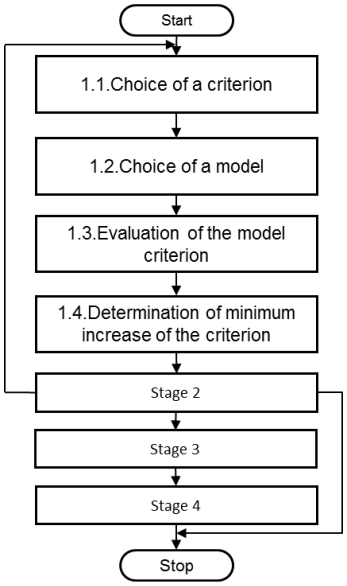

Figure 1. Choice of criterion and model

Determination of the optimisation potential

The second stage of the procedure provides for the detailed knowledge of the selected model behaviour (see Figure 2). At the beginning of the stage in collaboration with biologists, all model parameters that can be modified at the current microbiological technology development stage are selected. Further optimisations will be carried out with the respective parameters. Then the optimisation process follows, during which a

number of tests are conducted to see if the chosen model with the specific modified reactions can achieve the minimum criterion increase established at the previous stage. In the case of positive response optimisation is continued, otherwise the model or criterion must be changed.

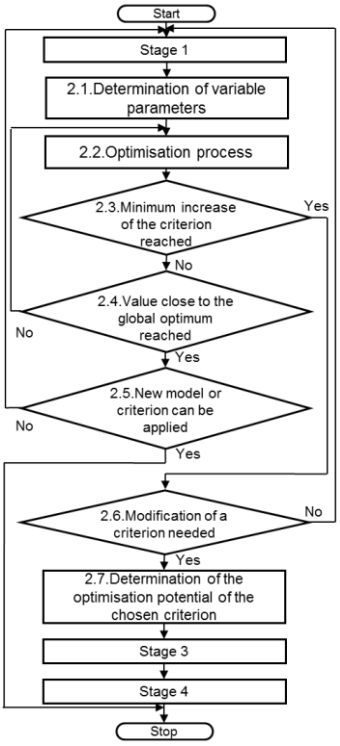

Figure 2. Determination of the optimisation potential

Ranking variable parameter combinations by efficiency

The task of the third stage of the procedure is to find the most effective variable parameter combinations in a case of specific number of variable parameters. The efficiency is understood as the ability of variable parameters or their combinations to improve the optimisation criterion value. Greater increase in the optimisation criterion corresponds to a higher efficiency. It is assumed that implementation of variable parameter concentration changes in every reaction at genome level costs equally (in the procedure more precise sums can also be taken into account). At the same time it is clear that the effect of various reactions and their combinations on the optimisation criterion is very different. Therefore, the combination of variable parameters with the minimum possible number of variable parameters must be found, which could use most of the total potential of the researched process already set in step 2.7. “Determination of the optimisation potential of the chosen criterion” (see Figure 3).

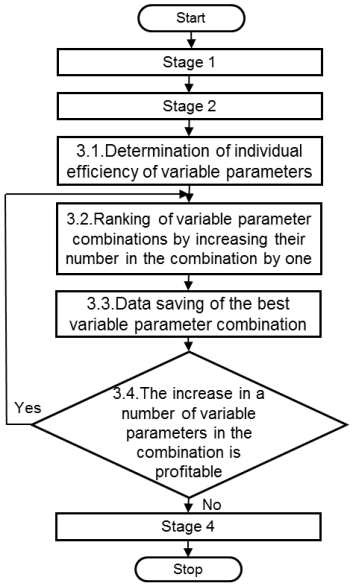

Figure 3. Ranking variable parameter combinations by efficiency

Analysing the application of solutions

Analysing the application of the found solutions and experimental testing of the final obtained parameter results of the optimisation process into production is carried out in Stage 4 of the procedure (see Figure 4). During Stage 4 it is verified if the found solution not only shows the existence of a steady state, but also is sufficiently stable for implementation with the production facilities at a reasonable cost. In addition to this, it must be verified if the steady state parameters (concentration of substances, temperature, *pH* values, etc.) are feasible.

Numerical optimisation during solving of the application example appears as the most time-consuming process with ambiguous results due to the fact that the numerical methods only converge to the global optimum, the value of which is unknown. Duration of the optimisation experiment can range from a few minutes up to 60-90 hours for a model with a few tens of reactions (Nikolaev, 2010). In practice it means that during optimisation it is not clear at what point to stop the optimisation process considering that the substantial improvement of the criterion value is not expected. Since there are many optimisation methods available, and each of them tend to have parameters which, if changed, affect their efficiency, the application of methods, the rate and stability of their convergence are very important factors influencing efficiency of the optimisation procedure.

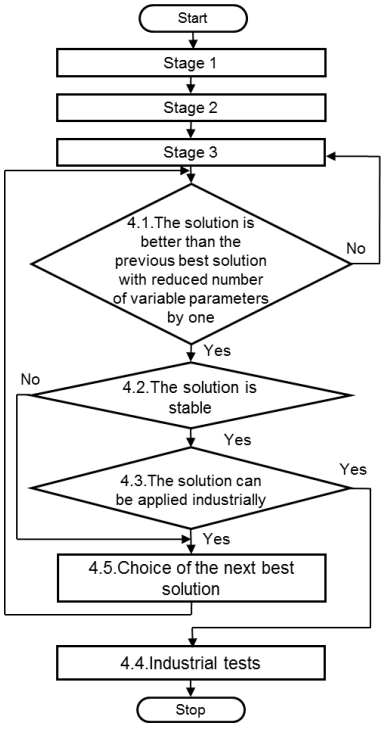

Figure 4. Analysing the application of the found solutions

2.2. Kinetic model of yeast glycolysis and optimisation settings

Yeast glycolysis model (Hynne, Danø, and Sørensen, 2001) downloaded from Biomodels database (Le Novère et al., 2006) is used as a test model for optimisation. The model contains 2 compartments, 24 reactions and 25 metabolites. Objective function in all optimisation runs was

$$K = \frac{Ethanol\ flow}{Glucose\,uptake} + 5 * Ethanol\ flow \tag{1}$$

Concentrations of enzymes catalysing 15 reactions ((*Hexokinase*), (*Alcohol dehydrogenase*), (*ATP consumption*), (*Glycerol synthesis*), (*Phospho- fructokinase*), (*Glyceraldehyde 3-phosphate dehydrogenase*), (*Storage*), (*Triosephosphate isomerase*), (*Pyruvate kinase*), (*Glucose uptake*), (*Phosphoglucoisomerase*), (*Phosphoenolpyruvate synthesis*), (*Pyruvate decarboxylase*), (*Aldolase*), and (*Adenylate kinase*)) were chosen as adjustable parameters. The initial values of variable enzyme concentration coefficients (VECC) were set at "1" to normalise the changes of enzyme concentrations.
COPASI (Hoops et al., 2006), build 30, is used as optimisation tool. Global stochastic optimization method "particle swarm" was applied with following method parameters:

Iteration Limit: 2000; Swarm Size: 50; Std. Deviation: 1e-06; Random Number Generator: 1; Seed: 0. The values of adjustable parameters were allowed to change within a wide range from -99% (VECC=0.01) up to 1,000% (VECC=10) from their initial values. "Steady state" subtask of optimisation was chosen.

3. Results and discussion

After changing all 15 enzyme concentrations by using the optimisation procedure, there is an increase in ethanol productivity from 0.80 mmol/min to 2.27 mmol/min and an improvement of product yield of substrate unit from 0.97 to 1.97, and a reduction of the following side products: glycerol from 0.085 mmol/min to 0.007 mmol/min, acetaldehyde from 0.065 mmol/min to 0.005 mmol/min, and cyanide from 0.020 mmol/min to 0.002 mmol/min.

The result of applying method for evaluating the remaining range of the optimisation potential can be assessed also in Figure 5: optimisation was stopped at seven variable enzyme concentration coefficients (VECC) in a combination as the remaining range of optimisation potential was not sufficient for gaining economic benefit in biotechnological production process.

The ranking of combinations was started from one VECC per combination. All 575 combinations of parameters of up to three in combination were optimised to find out the best combinations per one, two and three parameters in combination. That gave about 35% of possible objective function value increase (see Figure 5). Due to the large increase of the number of combinations for four VECC in combination the full search (optimisation of all combinations) was replaced by forward selection looking just for the best VECC among remaining VECC to be added to the combination. This way the procedure was continued until 91.4% of the optimisation potential of all VECC was reached with only seven parameters per combination (see Figure 5). Only 50 combinations of VECC were optimised giving 55% increase of total utilised potential from four to seven VECC in the combination of adjustable parameters.

That is reached by optimising just 625 combinations out of 16,383 combinations if full search of up to seven VECC per combination would be applied (see Figure 5).

As a result, it was confirmed that it is not worth analysing 16,384 combinations (a case of searching all combinations), thus considerably reducing the amount of necessary computational resources and time.

The coefficients for changing enzyme concentrations in seven reactions allowing for changes in the range 0.01 to 11.0 are developed for the baker's yeast (*Saccharomyces cerevisiae*) glycolysis model created by Hynne after the execution of the optimisation procedure (Hynne et al., 2001). Adjustable reactions and their coefficients (in parentheses) are as follows: Glucose Uptake (11.0), Hexokinase (2.73), Phosphofructokinase (2.02), Pyruvate Decarboxylase (5.93), Alcohol Dehydrogenase (11.0), Storage (0.01), and ATP consumption (11.0). The model predicts that after introducing changes the Ethanol Flow increases from 0.80 to 2.05 mmol/min, but Glucose Uptake increases from 0.83 to 1.12 mmol/min, and the following side products

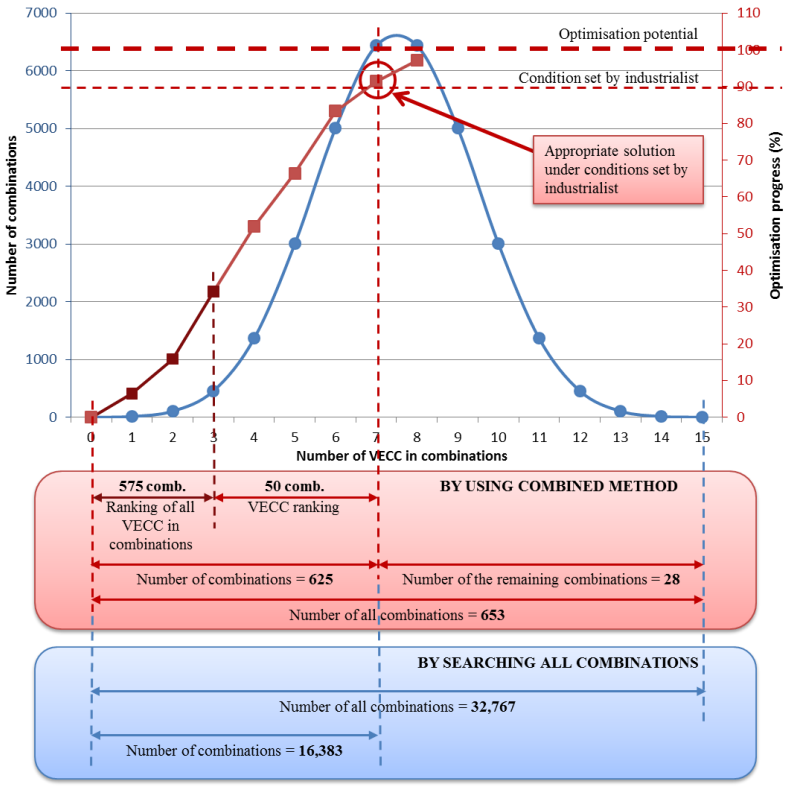

Figure 5. Application of method for evaluating the remaining range of the optimisation potential for reducing the number of combinations to be analysed

are reduced: glycerol from 0.085 mmol/min to 0.028 mmol/min, acetaldehyde from 0.065 mmol/min to 0.021 mmol/min, and cyanide from 0.020 mmol/min to 0.007 mmol/min. In the result, both the product yield of substrate unit is improved from 0.97 to 1.83, and ethanol yield is increased 2.6 times.

4. Conclusions

The proposed procedure is formalising the optimisation process of steady state of kinetic biochemical network models avoiding limitations of the intuitively suggested manipulations by different specialists. The proposed approach avoids scanning all the solution space keeping overview of the limitation of best possible solution determined in optimisation run when changes of all the adjustable parameters are enabled.

Combination of full parameter search with forward selection gives great reduction of the searched solution space and may be applied in case of time constraints. Otherwise, full search of parameter space is advised until the predetermined objective function value is reached. In both cases the reduction of solution space in the case study is at least by half.

The application of procedure suggests that after introducing the proposed changes in enzyme concentrations the Ethanol Flow increases from 0.80 to 2.05 mmol/min, but

Glucose Uptake increases from 0.83 to 1.12 mmol/min, and the following side products are reduced: glycerol from 0.085 mmol/min to 0.028 mmol/min, acetaldehyde from 0.065 mmol/min to 0.021 mmol/min, and cyanide from 0.020 mmol/min to 0.007 mmol/min. In the result, both the product yield of substrate unit is improved from 0.97 to 1.83, and ethanol yield is increased 2.6 times.

Acknowledgements

This work is funded by a project of European Structural Fund No. 2009/0207/1DP/1.1.1.2.0/09/APIA/VIAA/128 ”Latvian Interdisciplinary Interuniversity Scientific Group of Systems Biology” www.sysbio.lv.

References

Banga, J. R. (2008). Optimization in computational systems biology. *BMC Systems Biology*, *2*, 47. doi:10.1186/1752-0509-2-47

Bulipopa, N., and Sulins, J. (2013). Automatic termination of parallel optimization runs of global stochastic optimization methods by upper limit criterion. *Biosystems and Information Technology*, *2*(1), 15–18.

Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal M., Xu, L., Mendes, L., and Kummer, U. (2006). COPASI--a COmplex PAthway SImulator. *Bioinformatics (Oxford, England)*, *22*(24), 3067–74. doi:10.1093/bioinformatics/btl485

Hynne, F., Danø, S., and Sørensen, P. G. (2001). Full-scale model of glycolysis in Saccharomyces cerevisiae. *Biophysical Chemistry*, *94*(1-2), 121–163. doi:10.1016/S0301-4622(01)00229-0

Keasling, J. D. (2010). Manufacturing molecules through metabolic engineering. *Science (New York, N.Y.)*, *330*(6009), 1355–8. doi:10.1126/science.1193990

Kostromins, A., Mozga, I., and Stalidzans, E. (2012). ConvAn: a convergence analyzing tool for optimization of biochemical networks. *Biosystems*, *108*(1-3), 73–77. doi:10.1016/j.biosystems.2011.12.004

Le Novère, N., Bornstein, B. J., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., Sauro, H., Schilstra, M., Shapiro, B., Snoep, J.L., and Hucka, M. (2006). BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research*, *34*(Database issue), D689–91. doi:10.1093/nar/gkj092

Mauch, K., Buziol, S., Schmid, J., and Reuss, M. (2001). Computer-Aided Design of Metabolic Networks. In *AIChE Symposium Series* (pp. 82–91).

Mendes, P., and Kell, D. (1998). Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, *14*(10), 869–883. doi:10.1093/bioinformatics/14.10.869

Mozga, I., and Stalidzans, E. (2011a). Convergence Dynamics of Biochemical Models to the Global Optimum. In *Proceedings of the 3rd International Conference on E-Health and Bioengineering, 24-26 November 2011* (pp. 227–230). Iasi. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6150357&tag=1

Mozga, I., and Stalidzans, E. (2011b). Convergence dynamics of biochemical pathway steady state stochastic global optimization. In *Proceedings of IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI), 21-22 November 2011* (pp. 231–235). Budapest: IEEE. doi:10.1109/CINTI.2011.6108504

Mozga, I., and Stalidzans, E. (2011c). Optimization protocol of biochemical networks for effective collaboration between industry representatives, biologists and modellers. In *Proceedings of nternational Industrial Simulation Conference, 6.-8 June 2011* (pp. 91–96). Venice: EUROSIS.

Nikolaev, E. V. (2010). The elucidation of metabolic pathways and their improvements using stable optimization of large-scale kinetic models of cellular systems. *Metabolic Engineering*, *12*(1), 26–38. doi:10.1016/j.ymben.2009.08.010

Rodríguez-Prados, J.-C., de Atauri, P., Maury, J., Ortega, F., Portais, J.-C., Chassagnole, C., Acerenza, L., Lindley, N.D., and Cascante, M. (2009). In silico strategy to rationally engineer metabolite production: A case study for threonine in Escherichia coli. *Biotechnology and Bioengineering*, *103*(3), 609–20. doi:10.1002/bit.22271

Rodríguez-Acosta, F., Regalado, C. M., and Torres, N. V. (1999). Non-linear optimization of biotechnological processes by stochastic algorithms: Application to the maximization of the production rate of ethanol, glycerol and carbohydrates by Saccharomyces cerevisiae. *Journal of Biotechnology*, *68*(1), 15–28. doi:10.1016/S0168-1656(98)00178-3

Sendín, O. H., Exler, O., and Banga, J. R. (2010). Multi-objective mixed integer strategy for the optimisation of biological networks. *IET Systems Biology*, *4*(3), 236–48. doi:10.1049/iet-syb.2009.0045

Smallbone, K., Simeonidis, E., Swainston, N., and Mendes, P. (2010). Towards a genome-scale kinetic model of cellular metabolism. *BMC Systems Biology*, *4*(1), 6. Retrieved from http://eprints.ma.man.ac.uk/1424/

Stalidzans, E., Kostromins, A., and Sulins, J. (2012). Two stage optimization of biochemical pathways using parallel runs of global stochastic optimization methods. In *2012 IEEE 13th International Symposium on Computational Intelligence and Informatics (CINTI)* (pp. 365–369). IEEE. doi:10.1109/CINTI.2012.6496792

Sulins, J., and Mednis, M. (2012). Automatic termination of parallel optimization runs of stochastic global optimization methods in consensus or stagnation cases. *Biosystems and Information Technology*, *1*(1), 1–5. doi:10.11592/bit.120501

Sulins, J., and Stalidzans, E. (2012). Corunner: multiple optimization run manager for Copasi software. In *Proceedings of 5th International Scientific Conference on Applied Information and Communication Technologies, 26-27 April 2012* (pp. 312–316). Jelgava.

Vital-Lopez, F. G., Armaou, A., Nikolaev, E. V, and Maranas, C. D. (2006). A computational procedure for optimal engineering interventions using kinetic models of metabolism. *Biotechnology Progress*, *22*(6), 1507–17. doi:10.1021/bp060156o

Xu, G., and Wang, L. (2014). An Improved Geometric Programming Approach for Optimization of Biochemical Systems. *Journal of Applied Mathematics*, *2014*, 1–10. doi:10.1155/2014/719496

Received August 29, 2014, accepted September 15, 2014