9th INTERNATIONAL INDUSTRIAL SIMULATION CONFERENCE 2011

ISC'2011

EDITED BY

Simonetta Balsamo

and

Andrea Marin

JUNE 6-8, 2011

VENICE, ITALY

A Publication of EUROSIS-ETI

OPTIMIZATION PROTOCOL OF BIOCHEMICAL NETWORKS FOR EFFECTIVE COLLABORATION BETWEEN INDUSTRIALISTS, BIOLOGISTS AND MODELLERS

Ivars Mozga
Egils Stalidzans
Biosystems Group, Faculty of Information Technology,
Latvia University of Agriculture
Liela iela 2, LV3001, Jelgava, Latvia
ivars.m@gmail.com

KEYWORDS
Biology, Numerical methods, Optimization, Model analysis, Dynamic.

ABSTRACT

Due to the growing number and size of biochemical network computer models an efficient co-operation among experts industrialists, biologists and modellers in industrial biotechnology is becoming a topical issue. In order to speed up the introduction of systems biology achievements into production an organism independent universal collaboration protocol is necessary incorporating a sequence of actions starting with a choice of criterion and model up to industrial tests.
Protocol of cooperation of industrialists, biologists and modellers engineering the performance of a microorganism using systems biology approach to improve the efficiency of biotechnological process is developed. Method is described in form of an algorithm that consists of four consecutive stages starting with determination of the product of interest and optimization criteria and ending by an estimation of industrial process feasibility analysis taking into account the features of the available industrial equipment.
Application of protocol in yeast glycolysis optimization is given as example.
The protocol facilitates efficient resource usage and time savings in biochemical network optimization process due to rational sequence of operations and reduction of optimization duration.

INTRODUCTION

Interdisciplinary scientific fields generally and systems biology (Bruggeman and Westerhoff, 2007; Kitano, 2002) in particular suffers from efficient methods of interdisciplinary collaboration between biologists, modellers, mathematicians, chemists, bioengineers and others solving joint tasks (Aebersold, Hood, and Watts, 2000; Ideker, 2004). Additional problem is the industrial implementation of scientific achievements because of different attitude and expectations of scientists and industrialists about industrially attractive biotechnological process (Otero and Nielsen, 2010). This kind of problem can be addressed by a development of a collaborative optimization protocol.
This paper describes a protocol of cooperation of industrialists, biologists and modellers engineering the performance of a microorganism using systems biology approach to improve the efficiency of biotechnological process. Method is described in form of an algorithm that consists of four consecutive stages starting with determination of the product of interest and optimization criteria and ending by an estimation of industrial process feasibility analysis taking into account the available equipment.

STATE OF THE ART

Currently engineering of intracellular biochemical processes is a growing field in biotechnology aiming to reduce impact of several efficiency related topics: increase of product/substrate ratio, reduction of energy costs, reduction of side products and others. The development of new bioprocesses has also relatively new topics to cope with. Growing number of mathematical models of cellular bioprocesses stimulate increasing use of simulations and optimization of mathematically described process (Li et.al. 2010). Usually dynamic models of intracellular biochemical networks are described in a form of a set of non-linear differential equations that can be optimized using time consuming numerical methods (Hirmajer, Balsa-Canto, and Banga, 2009).
Usually optimization is made by biologists creating a hypothesis of a process improvement involving a small number of system elements because of difficulty to deal with high number of cross-talks and interactions between elements of systems – reactants and enzymes of biomolecular reactions inside the cell. Hypotheses are usually generated by biologists that are specialized in a particular organism. Then hypothesis is tested in a dynamic model. In case of success a biological experiment is following to test the feasibility of a process in a living organism. That is essential because engineering is performed using a small scale models (up to 30 reactions) of the process of interest because full scale dynamic models (thousands of reactions) are not developed yet.
Problem of the above mentioned approach is the high probability that not a complete space of optimal solutions is searched systematically (Schulz, Bakker, and Klipp,

2009) even in 30 reactions models. Thus complex interactions and non-intuitive industrially interesting solutions may be not hypothesized and tested. Additional problem is the limitation on the number of organisms that are known to particular teams of biologists. That increase risk of finding suboptimal solution of a problem because of lack of experience.

Another topic that reduce the intensity of implementation of scientific achievements is that industrialists are usually involved in late stages of optimization process and early rejection of biologically attractive but industrially uninteresting solutions are rejected in a stage when significant amount of resources are already spent.

There is a need for industrially oriented protocol of collaboration of industrialists, biologists and modellers in development of biotechnologically effective engineered organisms exploring all the space of possible solutions with early rejections of industrially uninteresting solutions. This article describes protocol of optimization of concentrations of reaction speed regulating enzymes. Concentrations of enzymes can be altered by influencing the transcription and translation (Klipp, 2005) intensity of enzymes . Enzyme concentration can be both increased and reduced thus respectively increasing and decreasing reaction flow (speed of the reaction). The protocol can be adapted for a wider range of tasks during optimization of biochemical networks.

OPTIMIZATION PROTOCOL

The protocol consists of four interlinked stages as shown in Figure 1. Each stage is meant to narrow the field of search of the best solutions as early as possible.

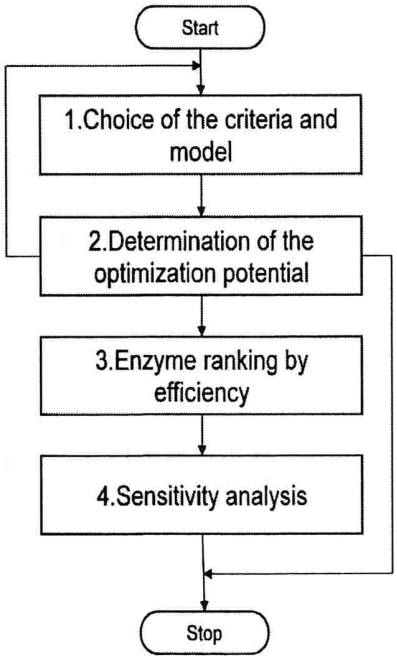

Figure 1 Four stages of optimization protocol

Stage 1: choice of the criteria and model

The first stage – choice of the criteria and model (Figure 2) is collaboration between industrialists, biologists and modellers to set the scope of the optimization work. It has a great influence on costs and duration of the optimization process. Main contributors at this stage are the industrialists. They should clearly express their interests both in mathematical terms and in form of other limitations including legislation, environmental issues and others which may be unknown by biologists and modellers. Still collaboration with biologists and modellers is needed to assure that the criteria's can be calculated from the chosen model. Extension of a model may be needed to link it with the criteria.

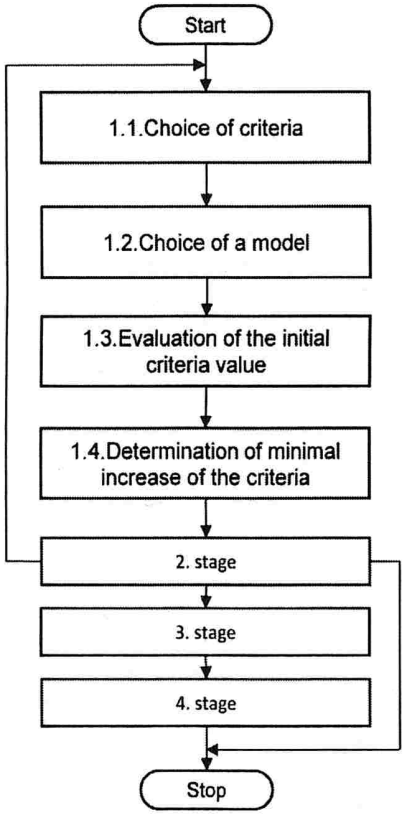

Figure 2 Algorithm for the stage "choice of the criteria and model"

Choice of criteria (1.1.) is a complex task. Too narrow criteria can limit the search in a very narrow area of possible solution space and end without industrially interesting solution at early stage. Too general criteria might cause the opposite drawback – search can be too wide and limitations that were not indicated at early stage can come up at the fourth stage and solutions might be rejected already after spending significant resources to analyze it. A model that includes criteria related processes should be chosen (step **Choice of a model (1.2.)**) from respective databases (Chen et.al. 2010; Oliver and Snoep, 2004) by modellers in collaboration with biologists. The model can be slightly adapted to have a steady state (Klipp, 2005) that is a prerequisite for industrial biotechnological process. The next step is the **evaluation of the initial criteria value (1.3.)** (by modeller) of the original model before optimization. In the next step the **determination of minimal increase of the criteria (1.4.)** should be done by industrialist operating mainly with economical factors. Thus it is defined that criteria values

below the minimal increase are not interesting for industrialist and further research is meaningless if minimal increase is not reached.

Stage 2: determination of the optimization potential

The task of the second stage (Figure 3) is the determination of the optimization potential. It is necessary to find out if the chosen criteria and organism can give increase of optimization criteria above the value defined in the first stage. In case of failure the first stage has to be repeated.

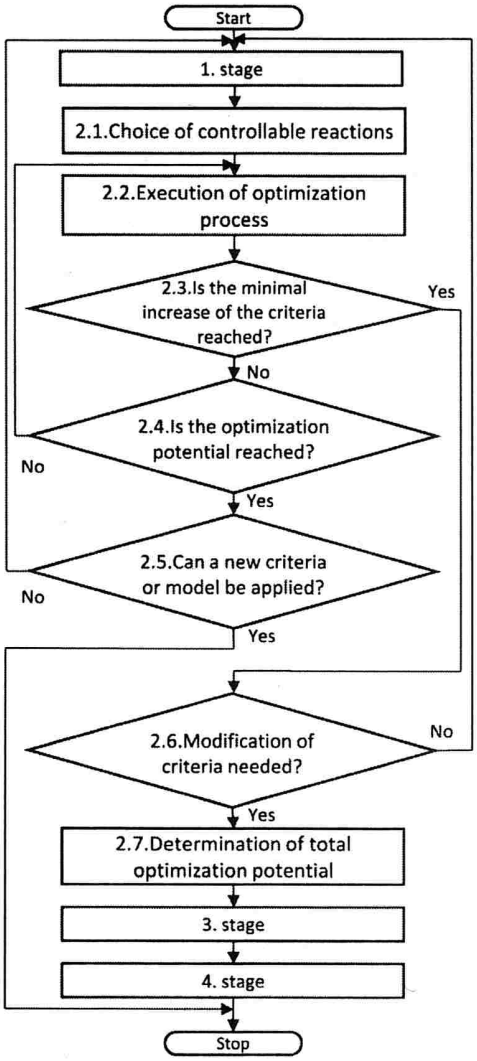

Figure 3 Algorithm for the stage "determination of the optimization potential"

The first step of this stage is **the choice of controllable reactions (2.1.)** in the model that happen by the help of an enzyme and therefore their flow can be adjusted by change of enzyme concentration. Reactions can be selected by biologists and modellers using bases SABIO-RK (Wittig et.al. 2006), BRENDA (Scheer et.al. 2006) or KEGG (Kanehisa et.al. 2010). The next step is the **execution of optimization process (2.2.)** performed by modeller making optimization run where all the reaction related parameters are allowed to change in a wide range, for instance, from -99% of the initial value of the original model up to + 1000%. Modellers should consult with biologists because 1) some reactions may need some biomass related flow and 2) to interpret biologically correctly the reaction parameters of optimal solution. This step can be performed using general numerical optimization software or a systems biology related tools and model formats like COPASI (Hoops et.al. 2006), Potters Wheel (Maiwald and Timer, 2008), SB Toolbox2 (Schmidt and Jirstrand, 2006) and others. In case of a big model with high number of parameters the optimization run can take long time (several days) until optimization potential is clarified. Therefore optimization process should be observed by a modeller asking **"is the minimal increase of the criteria reached?" (2.3.).** In case of a negative answer the question **"is the optimization potential reached? (2.4.)** should be estimated. In other words it means: is the optimization reached maximal value? Positive answer indicates that the optimization has ended and it's potential is not high enough and industrialist should be asked: **"can a new criteria or model be applied?" (2.5.)** to see if adaptation of criteria's can help. Negative answer means that the optimization protocol has not found industrially feasible solution even in modelling level. In case of positive answer to the question 2.3. the criteria has to checked in the step **"Modification of criteria needed?" (2.6.).** The question should be asked to the industrialist as the optimization process and growth of criteria could lead to unrealistic values of parameters. Negative answer to the question 2.6. should be followed by **determination of total optimization potential (2.7.)** done by modeller running the optimization until progress of optimization criteria reaches it's best value for given model. An indication for that is a long stagnation of optimization process.

Stage 3: enzyme ranking by efficiency

The third stage enzyme ranking by their efficiency in increase of the criteria (Figure 4) is performed by the modeller except for the step 3.5. which is done by industrialist. Some of enzymes (via reactions that they control) are strongly contributing in increase of criteria while some of them show very small or no influence on the criteria. Change of concentration of particular enzyme is costly. So it would be of advantage to find out 1) how the values of criteria are growing depending on the number of modified enzymes and 2) which would be the best enzymes to influence in case of modification of set of one, two, three and so on enzymes. For instance in case of one enzyme the most efficient is the enzyme A reaching 40% (increase of 40%) of total optimization potential, in case of two enzymes those are enzymes B and D reaching 65% (increase =65-40=25%) and in case of three enzymes those are A, D and E covering 73% (increase =73-65=8%) of total optimization potential.

The increase of criteria becomes smaller and smaller while the number of enzymes in combinations are increasing. That leads to a situation when adding of more enzymes to be modified becomes too expensive for the small increase of expected efficiency.

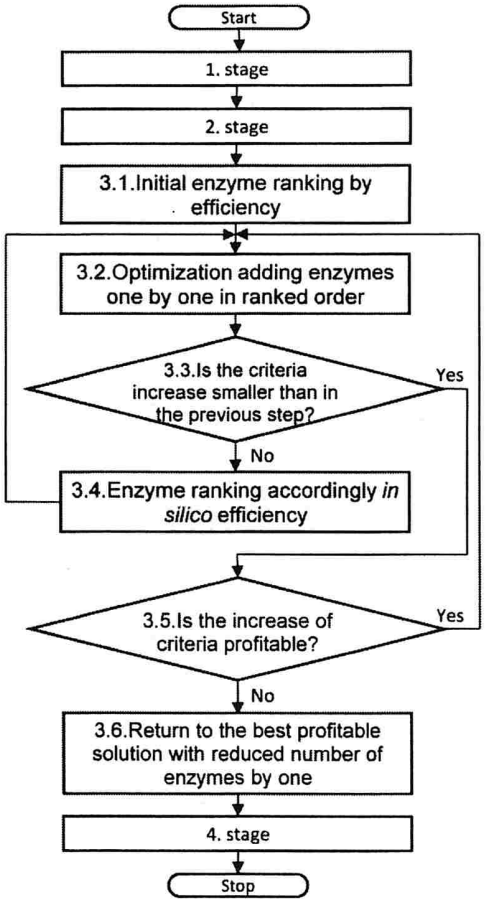

Figure 4 Algorithm for the stage "enzyme ranking by efficiency"

Scanning of all the possible combinations (no order important and no repetition allowed) of enzymes to be modified the number of optimizations K can be calculated by formula:

$$K = \sum_{r=1..j} \frac{n!}{r!(n-r)!} \qquad (1)$$

where r – number of available enzymes to choose from, j - up to which number of enzymes combinations are counted. Thus a model with 15 enzymes that can be influenced in case of up to five enzyme combinations would have 4943 possible combinations to optimize. In case of up to 10 there would be 30826 combinations and in case of 15 there would be 32767 combinations. Target of this stage is to find the best combinations of enzymes for any number of enzymes to be modified in shortest time possible.

The first step is the **initial enzyme ranking by efficiency (3.1.).**. Depending on the model several methods can be used by the modeler to determine the influence of concentration change of a particular enzyme on the other reactions and the value of the criteria. Some of them are Metabolic Control Analysis (Crabtee et.al.1985), (Fell, 2005), (Fell and Sauro, 1985), (Reder,1988), ranking accordingly to the efficiency increase in case of single optimized enzyme or others. It is assumed that the efficiency rank will not be predicted accurately because of nonlinearity of nonlinear differential equations. **Optimization adding enzymes one by one in ranked order (3.2.)** has to be performed by modeller in a cycle with the next step **"Is the criteria increase smaller than in the previous step?" (3.3.)** as long as the answer is negative. Step 4.3. is followed in cycle by **enzyme ranking accordingly in silico (computer simulation) efficiency (3.4.)** changing the efficiency rank of enzymes accordingly the optimization outcome. If the efficiency increase in the step is smaller than for the previous number of enzymes (3.3.) the question **"Is the increase of criteria profitable?" (3.5.)** comes up. Negative answer leads to the end of the cycle and **return to the best profitable solution with reduced number of enzymes by one (3.6.)** that will be analysed in the fourth stage. During the step 3.2. each case of criteria increase and corresponding set of optimization parameter values has to be saved as some of solutions may have weak stability and suboptimal solutions may become the best feasible ones in the fourth stage.

Stage 4: stability analysis of best solutions

During the fourth stage "Sensitivity analysis of best solutions" (Figure 5) the best technically feasible stationary state should be found taking into account the dynamic parameters of both the cellular dynamics and industrial control system. Modellers and industrialists are the main contributors at this stage.

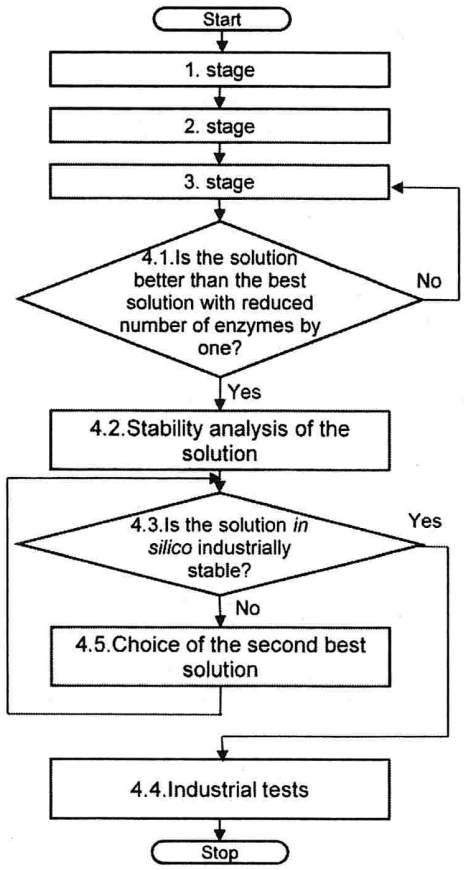

Figure 5 Algorithm for the stage "sensitivity analysis of best solutions"

The stage starts with a question **"is the solution better than the best solution with reduced number of enzymes by one?" (4.1.)**. In case of negative answer the best

solution of combination with less number of enzymes has to be preferred because due to reduced complexity and realization costs. Thus return to the step 3.6. is needed.

Stability analysis of the solution (4.2.) is the next step to be performed. Robustness of bioprocess has to be estimated under known dynamic parameters of cellular biochemical process of interest and the accuracy and dynamic parameters of the available control system of of the bioreactor or other industrial system. Different approaches can be used depending on the available information about the control system. One of methods is the implementation of dynamic model of the control system into the model of the bioprocess and run the stability analysis of simultaneous interaction of both systems.

If the answer to the question **"Is the solution *in silico* industrially stable?" (4.3.)** is positive, the **industrial tests (4.4.)** can be started on level of biological experiments and in case of success may be implemented in the industrial process. In case of negative answer to the question 4.3. the **choice of the second best solution (4.5.)** has to be performed by a modeller and return to the step 4.1 thus forming a cycle.

The optimization progress recorded in the step 3.2. is used in the step 4.5. If the solution that corresponds to the best value of criteria is not stable, the solution with next best criteria value has to be analyzed for stability.

APPLICATION CASE

The protocol has been applied to increase the profitability of yeast glycolysis in production of ethanol from glucose using model published by Hynne and colleagues (Hynne et.al.2001). The application case is related to the production of biofuels. The model contains 24 reactions and 25 reactants (Figure 6). 15 reactions are performed by the help of enzymes with variable concentrations that influence flow of the corresponding reaction. Using the protocol and software tool COPASI combinations of modified enzyme concentrations was found. The optimization criteria was:

$$C = \frac{Ethanol flow}{Glucose\ flow} + 5 * Ethanol\ flow \quad (2)$$

where
Ethanol flow – product produced,
Glucose flow – feedstock consumed.

As a result increase of the criteria from 4,99 (Ethanol flow = 0,804 mmol/min, Glucose flow = 0,832 mmol/min) was increased to 13,3 (Ethanol flow = 2,27 mmol/min, Glucose flow = 1,15 mmol/min).

The fourth stage of the protocol was not performed as it depends on equipment parameters and may be specific from case to case. The increase of ethanol production was reached without ensuring the production of biomass and therefore might perform in biological experiments with lower efficiency than in computer simulations.

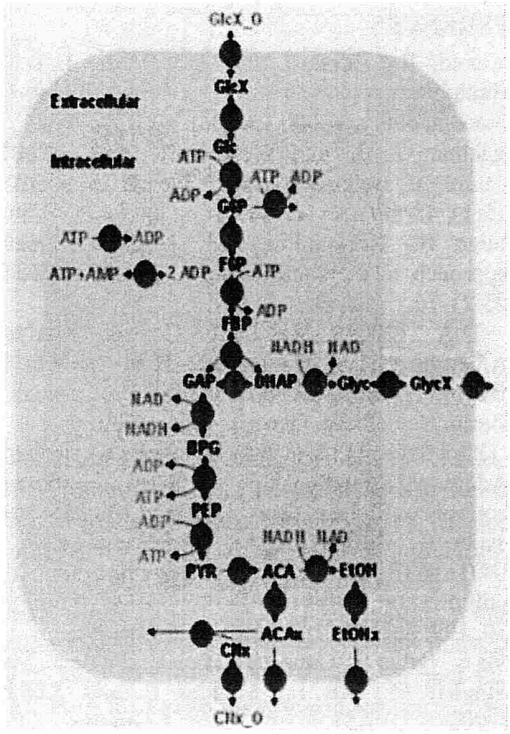

Figure 6 Yeast glycolysis model (Hynne, 2001) Screenshot from JWS online website (http://jjj.biochem.sun.ac.za/database/hynne/index.html)

CONCLUSION

The developed protocol describes cooperation of industrialists, biologists and modellers optimizing the industrial performance efficiency of a microorganism. The protocol consists of four consecutive stages: 1) choice of the criteria and model, 2) determination of the optimization potential enzyme 3) ranking by efficiency and 4) sensitivity analysis of best solutions.

Early rejecting of industrially and/or biologically unfeasible solutions increase the efficiency of optimization.

The protocol propose systematic scan the solution space for the best solutions thus checking also contra intuitive solutions and eliminating subjective approach of particular scientist.

Significant in silico increase of yield and and ethanol production is achieved in case of yeast glycolysis models (Hynne et.al. 2001).

ACKNOWLEDGEMENTS

This work is funded by a project of European Structural Fund Nr. 2009/0207/1DP/1.1.1.2.0/09/APIA/VIAA/128 'Latvian Interdisciplinary Interuniversity Scientific Group of Systems Biology' www.sysbio.lv

REFERENCES

Aebersold, R., Hood, L. E., and Watts, J. D. (2000). Equipping scientists for the new biology. Nature biotechnology, 18(4), 359. doi: 10.1038/74325.

Bruggeman, F. J., and Westerhoff, H. V. (2007). The nature of systems biology. Trends in microbiology, 15(1), 45-50. doi: 10.1016/j.tim.2006.11.003.

Crabtree B., Newsholme E.A. (1985) A quantitative approach to metabolic control. Curr Top Cell Regul 25:21-76.

Fell D.A. (2005) Metabolic Control Analysis. In: Alberghina L., Westerhoff H.V. (eds.) Systems Biology, Definitions and Perspectives, Springer-Verlag Berlin Heidelberg, Germany, pp. 69-80.

Fell D.A., Sauro H.M. (1985) Metabolic Control Analysis: Additional relationships between elasticities and control coefficients. Eur J Biochem 148:555-561.

Hirmajer, T., Balsa-Canto, E., and Banga, J. R. (2009). DOTcvpSB, a software toolbox for dynamic optimization in systems biology. BMC bioinformatics, 10, 199. doi: 10.1186/1471-2105-10-199.

Hoops S., Sahle S., Gauges R., Lee C., Pahle J., Simus N., Singhal M., Xu L., Mendes P. and Kummer U. (2006). COPASI — a COmplex PAthway SImulator.Bioinformatics 22, 3067-74.

Hynne F., Danø S., Sørensen P.G. (2001) Full-scale model of glycolysis in Saccharomyces cerevisiae. Biophys Chem, 94(1-2):1, pp.21-63.

Ideker, T. (2004). Systems biology 101--what you need to know. Nature biotechnology, 22(4), 473-5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15085805.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M.; KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. 38, D355-D360 (2010).

Kitano, H. (2002). Computational systems biology. Nature, 420(6912), 206-10. doi: 10.1038/nature01254.

Klipp E., Herwig R., Kowald A., Wierling C. and Lehrach H. (2005) Systems Biology in Practice, Concepts, Implementation and Application, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 467 p.

Li C., Donizelli M., Rodriguez N., Dharuri H., Endler L., Chelliah V., Li L., He E., Henry A., Stefan M.I., Snoep J.L., Hucka M., Novère N.L. and Laibe C. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. BMC Systems Biology 2010, 4:92

Maiwald T. and Timmer J. *Dynamical Modeling and Multi-Experiment Fitting with PottersWheel.* Bioinformatics 2008, 24(18):2037-2043

Olivier, B.G. and Snoep, J.L. (2004) Web-based modelling using JWS Online, Bioinformatics, 20:2143-2144

Otero, J. M., and Nielsen, J. (2010). Industrial systems biology. Biotechnology and bioengineering, 105(3), 439-60. doi: 10.1002/bit.22592.

Reder C. (1988) Metabolic control theory: a structural approach. J Theor Biol 135:175-201.

Scheer M., Grote A., Chang A., Schomburg I., Munaretto C., Rother M., Söhngen C., Stelzer M., Thiele J., Schomburg D. BRENDA, the enzyme information system in 2011. Nucleic Acids Res., 39:670-676 (2011)

Schmidt, H. and Jirstrand, M. (2006). Systems Biology Toolbox for MATLAB: a computational platform for research in Systems Biology. Bioinformatics, 22(4), 514–515.

Schulz, M., Bakker, B. M., and Klipp, E. (2009). TIde: a software for the systematic scanning of drug targets in kinetic network models. BMC bioinformatics, 10, 344. doi: 10.1186/1471-2105-10-344.

Wittig U., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., Anstein, S., Saric, J. and Rojas, I. SABIO-RK: Integration and Curation of Reaction Kinetics Data. In proceedings of the 3rd International workshop on Data Integration in the Life Sciences 2006 (DILS'06). Hinxton, UK. Lecture Notes in Bioinformatics, 4075: 94-103(2006).

AUTHOR BIOGRAPHY

IVARS MOZGA studied IT in Latvia University of Agriculture. Since 2006 he is a PhD student in Latvia University of Agriculture specializing in computer control of biological systems. Doctoral studies are connected to modeling of biomolecular processes within cellular structures. Researcher in a project of European Structural Fund Nr. 2009/0207/1DP/1.1.1.2.0/09/APIA/VIAA/128 'Latvian Interdisciplinary Interuniversity Scientific Group of Systems Biology' www.sysbio.lv

EGILS STALIDZANS studied in Riga Technical University (Latvia), Faculty of Computer Science and Information Technology where defended his PhD in 2005. Research is mainly related to development of mathematical models under information insufficiency. In 2010 he became associated professor and senior researcher at Latvia University of Agriculture, Faculty of Information technology. Since 2006 he leads Biosystems Group. He is interested in application of Systems Biology approach in control of industrial bioprocesses. Project coordinator in a project of European Structural Fund Nr. 2009/0207/1DP/1.1.1.2.0/09/APIA/VIAA/128 'Latvian Interdisciplinary Interuniversity Scientific Group of Systems Biology' www.sysbio.lv