IN SILICO MODELLING AND METABOLIC ENGINEERING OF ESCHERICHIA COLI TO SUCCINIC ACID PRODUCTION

Zsolt BODOR[1], Andrea (IUHASZ) FAZAKAS[2], Szabolcs LÁNYI[3], Beáta ÁBRAHÁM[4]

Succinic acid is identified to have a great economical potential in a biobased economy. To improve the production in Escherichia coli metabolic engineering should be carried out. For new mutant strain design is important to in silico simulate and analyze the metabolic changes, network of a cell under different environmental and/or genetic perturbations. To investigate the genetic and environmental perturbations, the relationship between biomass and succinate yield, the NADH oxidation connection to growth rate and succinate production, in silico metabolic analysis was carried out using constraint-based metabolic flux simulations in minimal medium with glucose and glycerol as carbon sources.

Keywords: succinic acid, metabolic engineering-modelling, systems biology, FBA

Abbreviations

COBRA: constraint-based reconstruction and analysis
FBA: flux balance analysis
NADH: nicotinamide adenine dinucleotide
SBML: systems biology markup language
mmol $gDW^{-1}h^{-1}$: millimoles per gram dry cell weight per hour
PEP: phosphoenolpyruvate
Acetyl-CoA: Acetyl coenzyme A

1. Introduction

A variety of useful metabolites can be produced by cell factories including pharmaceutical chemicals, fine and commodity chemicals [1].

[1] Faculty of Applied Chemistry and Material Science, University POLITEHNICA of Bucharest, Romania, e-mail: bodorzsolt@sapientia.siculorum.ro
[2] Faculty of Applied Chemistry and Material Science, University POLITEHNICA of Bucharest, Romania
[3] Faculty of Science, Department of Bioengineering, Sapientia Hungarian University of Transylvania, Cluj, Romania
[4] Faculty of Science, Department of Bioengineering, Sapientia Hungarian University of Transylvania, Cluj, Romania

The succinic acid (butanedioic acid), has a high economical impact (expected to reach $496 million in 3 years) and it is used in a number of industries including: polymers, food, surfactants and detergents, is a supplement for pharmaceuticals [2-7], it can be converted to biodegradable plastics [8,9]. The U.S. Department of Energy announced succinic acid as one of the 12 top chemical building blocks produced by microorganisms [4,8,10].

There are many bacteria being capable to produce succinate: *Anaerobiospirillum succiniciproducens* [7], *Mannheimia succiniciproducens* [8], *Actinobacillus succinogenes* [11] or even metabolically engineered *E. coli* strains in complex medium [8,9], but the complexity of nutrients or the necessary environmental conditions make these methods very expensive [6].

The level of succinate produced by native strains of *E. coli* in minimal medium is very low. Many scientists have described genetic engineering approaches to improve succinate production in *E. coli* by different methods: overexpressing different genes [11,12], with combination of different gene deletions [6], or using complex medium [8,9].

One way to improve succinate production is to redirect the carbon flow towards succinate by blocking pathways of competing metabolic products such formate, lactate, ethanol and leaving only the succinate pathway to achieve redox balance during substrate utilization [5].

Systems biology can now play a role in the metabolic engineering process by guiding interventions to divert metabolite flux within a microbial cell.

COBRA Toolbox [13,14] is an open-source and modular platform, incorporating strain optimization tasks, algorithms such as Flux Balance Analysis (FBA) [15,16]. Using FBA and a complex genome-scale metabolic model we can predict the production of a desired product and optimize the production rate under different environmental conditions. In addition, effects of environmental and genetic perturbations on the metabolic network can also be simulated, using complex calculation methods such as, robustness analysis, phase plane analysis, dynamic modelling, etc [13,14,17].

The main aim of the study is *in silico* screening of genes to be removed from *E. coli* for the overproduction of succinic acid from glucose and glycerol under anaerobic conditions using minimal medium, to test the relationship between growth and succinate production and finally to analyze the effect of NADH oxidizing pathways on cell viability and succinic acid production.

We firstly predicted *in silico* the flux distributions inside and outside or the cell under specific conditions and substrates followed by the knock-out of genes to test the effect on cell metabolism. Robustness analysis, dynamic growth simulations (batch growth) and phenotypic phase plane analysis [13] of *E. coli* K12 MG1655 on glucose and glycerol minimal media under aerobic, microaerobic and anaerobic conditions were carried out.

With *in silico* studies of genetic engineering [18-20] we can reduce the time and cost of wet experiments and can design industrially important strains [1].

2. Experimental

For simulations the latest version of the *Escherichia coli* metabolic network model formulated by Orth *et al.* [21] (2011) was downloaded from BioModels online database (http://www.ebi.ac.uk/biomodels-main/). Calculations were made in MATLAB R2012a (Mathworks Inc,; Natick, MA) utilizing the SBML Toolbox (version 4.1.0, http://sbml.org/software/sbmltoolbox/) and the COBRA Toolbox (version 2.0.5, http://opencobra.sourceforge.net/openCOBRA/Welcome.html) [13,14]. Optimization was undertaken using the Gurobi (version 5.1.0).

2.1. Flux balance analysis (FBA)

It is a constraint-based modelling to determine flux distribution in the model using linear optimization of an objective function- typically biomass production.

Stoichiometric, steady-state balances on all metabolites are imposed as linear constraints on the basic equation:

$$S*v=0 \quad (1)$$

- where, S is an m*n matrix and m is the number of metabolites and n is the number of reactions in the model. The linear optimization problem can be formulated as follow:

$$\text{Maximize: } z=c^T v \quad (2)$$

- where c denotes the vector defining the weights for each of the fluxes in v.

The lower and upper bounds: $a_i \leq v_i \leq b_i$ – representing the constraints for reaction irreversibility and substrate uptake from the environment. All flux units are in mmol $gDW^{-1}h^{-1}$, except the biomass flux, which has units of h^{-1}. Reversible reactions have lower bounds of -1000 while irreversible reactions have a lower bound of 0, in both cases the upper bounds are 1000. Three different environmental conditions were tested, namely aerobic, microaerobic and anaerobic under minimal medium using glucose or glycerol as sole carbon sources including the wild-type and the genetically modified strains too. In each simulation the maximization of the biomass was set as the objective function.

2.2. Minimal media determination and substrate utilization prediction

Minimal media was used for our simulations and the carbon source was glucose or glycerol. Glycerol was chosen as an alternative carbon-source because

it is an abundant carbon feedstock resulted from biodiesel production. Consumption rate (substrate uptake rate) for the main carbon substrate was set to 10 mmol $gDW^{-1}h^{-1}$. Under microaerobic conditions an oxygen uptake rate of 5 mmol $gDW^{-1}h^{-1}$ was used while 0 mmol $gDW^{-1}h^{-1}$ for anaerobic conditions.

Growth rates and substrate utilization rates were compared with those obtained experimentally elsewhere (strong correlation between results).

2.3. Analysis of maximum biomass production rate and the secretion of succinic acid

Cellular growth simulation, gene deletions were carried out with FBA using Cobra Toolbox.

The production rate of succinic acid was determined in each simulation. To improve the succinate production our assumption was to eliminate the competing pathways. Three major pathways were spotted and eliminated step by step: the first one was the formate pathway- the *pfl* (pyruvate formate lyase) gene was knocked-out by setting the lower and upper bounds to 0. We decided to start with *pfl* because under anaerobic and microaerobic conditions the production rate of formate was the highest one followed by lactate and ethanol. After elimination of the lactate dehydrogenase-*ldh* and alcohol dehydrogenase-*adh* genes the phosphotransferase system was our next target pathway to increase the yield of succinate. The created mutant strain was analyzed, including FBA analysis, dynamic growth simulations (batch growth), phenotypic phase plane analysis (Phpp) and Flux Variability Analysis (FVA) (data not shown here), genes and reactions essentiality analysis (data not shown here) was carried out using different environmental and genetic conditions.

Dynamic growth simulations

FBA analyzes can be used to examine dynamic processes including the microbial growth in batch cultures by combining FBA with an iterative approach based on a quasi-steady-state assumption [14]. With this simulation the growth rate and the metabolites production and consumption can be estimated. The initial substrate concentration was set to 10 mmol L^{-1} while the initial biomass concentration to 0.035 g L^{-1}. Time step is 25 min and the maximum number of steps is 150 in order to allow observing the full diauxic shift.

Phenotypic phase plane analysis (Phpp)

Phpp was used to vary to parameters simultaneously, to see the interactions between two reactions under different conditions (mentioned before)

and the same time to determine the impact on growth rate and plot the results as a phenotypic phase plane. Simulations were carried for wild-type and mutant strains while varying substrates and oxygen/carbon dioxide uptake rates.

3. Results and discussion

Succinate can be produced by *E. coli* under anaerobic and microaerobic conditions but the quantity of the excreted succinic acid is very low. Our simulation to increase production was performed with substrate flux consideration of 10 mmol $gDW^{-1}h^{-1}$.

Relationship between growth rate and succinate production was computed followed by the effect of NADH oxidation to growth rate and succinic acid production.

Flux Balance Analysis

FBA was used to calculate the growth rate of *E. coli* and the flux distribution inside the cell for wild-type and the newly designed strains. In the Table 1 the specific maximum growth rates are presented under different substrate, environmental and genetic conditions.

Table 1

Maximum growth rates comparison of *E. coli* on different substrates using aerobic, microaerobic and anaerobic conditions

Substrate	Strain	Growth rate (h^{-1})		
		Aerobic	Microaerobic	Anaerobic
Glucose	Wild-type/ Mutant	0.98/ 0.97	0.49/ 0.38	0.24/ 0.09
Glycerol	Wild-type/ Mutant	0.56/ 0.56	0.33/ 0.29	0.08/ 0.001
Succinate (mmol $gDW^{-1}h^{-1}$)	Glucose	-/-	-/6.69	0.08/12.5
	Glycerol	-/-	-/0.32	0.03/6.00

Simulations were carried out in minimal medium with either glucose or glycerol (uptake rate 10 mmol $gDW^{-1}h^{-1}$)

As we can observe the biomass production rate on glucose is higher than on glycerol in both case (wild-type and mutant). The reason is the difference between substrates molecular weights, but after calculations it was clear that mass yield of cell-mass on glucose was 0.43 gDW g $glucose^{-1}$ while in case of glycerol 0.53 under aerobic conditions. It is also apparent that under microaerobic conditions the cell-mass yield on glycerol is still higher; meanwhile under anaerobic conditions glucose gives a higher yield (0.13 instead of 0.11). A significant increase of succinate yield was obtained only with *pfl, ldh, adh, GLCpts* quadratic deletion, a molar yield of 1.25 mol mol $glucose^{-1}$. The fourth

identified target pathway was the phosphoenolpyruvate-dependent phosphotransferase (*GLCpts*) system which phosphorylates uptaken sugars by producing pyruvate from phosphoenolpyruvate.

Inactivation of the *pts* system results in an increase in the PEP pool, allowing the generation of more succinate.

Engineering the redox metabolism was found to be effective to improve the succinate production from glucose as the sole carbon source. As we expected the carbon flux to formate was the dominant ~ was the sum of the fluxes to acetate and ethanol.

The optimal succinate production of the mutant was 12.5 mmol $gDW^{-1}h^{-1}$ with a growth rate of 0.09 under anaerobic conditions and 6.69 under microaerobic conditions, growth rate of 0.38.

The NADH generated during glycolysis is reoxidized in the process when the organic intermediates are reduced, the reducing equivalents are fully consumed. The cell tries to redress the reducing equivalents and in every genetic mutation different metabolites are produced. To maintain the redox balance Acetyl-CoA is converted to lactate and ethanol and ATP is produced from the acetate pathway. The residual acetate produced in the *ackr* mutant is resulted from the pyruvate oxidase (*poxB*) activity.

Glycerol is an abundant and inexpensive carbon source, generated by biodiesel technology.

During anaerobic conditions the glycerol fermentation resulted formate, ethanol and acetate and a small amount of succinate (0.0026 mol mol $glycerol^{-1}$), without lactate (Table 1).

Effects of eliminating by-product pathways using glycerol as sole carbon source

As we can observe in Table 1 almost in each simulation the mutant strain growth rate was slower. The deletion of the *pfl* reaction blocked the cell growth under anaerobic conditions. The explication could be that Acetyl-CoA is an essential metabolite for biosynthesis that is produced primarily by *pfl* during fermentative growth [3]. We decided to test different co-substrates (data not shown) to enhance the cell viability. Best solution was found in case of glucose as co-substrate. The production of different metabolites such as formate and ethanol was necessary for the biomass synthesis during the glycerol metabolism [5]. The ATP is consumed under biomass synthesis process and reducing equivalents (NADH) are produced. The regeneration of both is resolved by the cell producing other by-products. Conversion of glycerol to formate and ethanol fulfils energy requirements by generating ATP via substrate-level phosphorylation.

The role of glucose as co-substrate in glycerol metabolism was tested. The succinate yield was linearly dependent on glucose uptake rate, however we decided to use an uptake rate for glucose only 1 mmol $gDW^{-1}h^{-1}$ just to improve the cell viability.

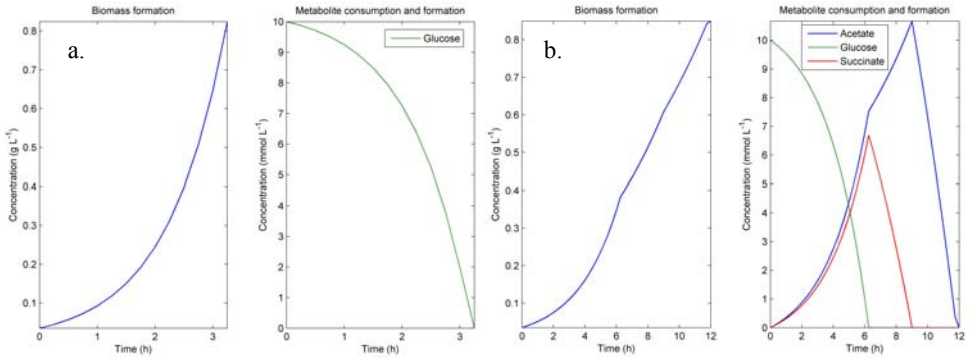
The triple mutant (*pfl*, *ldh*, *adh*) was able to produce succinate with a production rate of 6 mmol $gDW^{-1}h^{-1}$ under anaerobic conditions (growth rate 0.001 h^{-1}) and 0.32 using microaerobic conditions (growth rate 0.29 h^{-1}). In each simulation the biomass production is negatively affected by the succinic acid production.

With these genetic modifications the synthesis of succinate remained as the primary route of NAD^+ regeneration [9]. Eliminating the pathways of the metabolites like formate, lactate and ethanol is considered as an ideal strategy to improve succinate yield from glucose and glycerol. The inactivation of *pfl*, *ldh* and *adh* increased the yield of succinic acid under anaerobic conditions on minimal medium, but cell growth was decreased (NADH regeneration inability).

Dynamic growth simulations

Using dynamic FBA we can examine the dynamics of growth rate and the production and consumption rate of different metabolites under different conditions. In this study the wild-type and mutant strains behavior was examined, in the Fig.1 are presented the results for mutant strains.

I.



c.

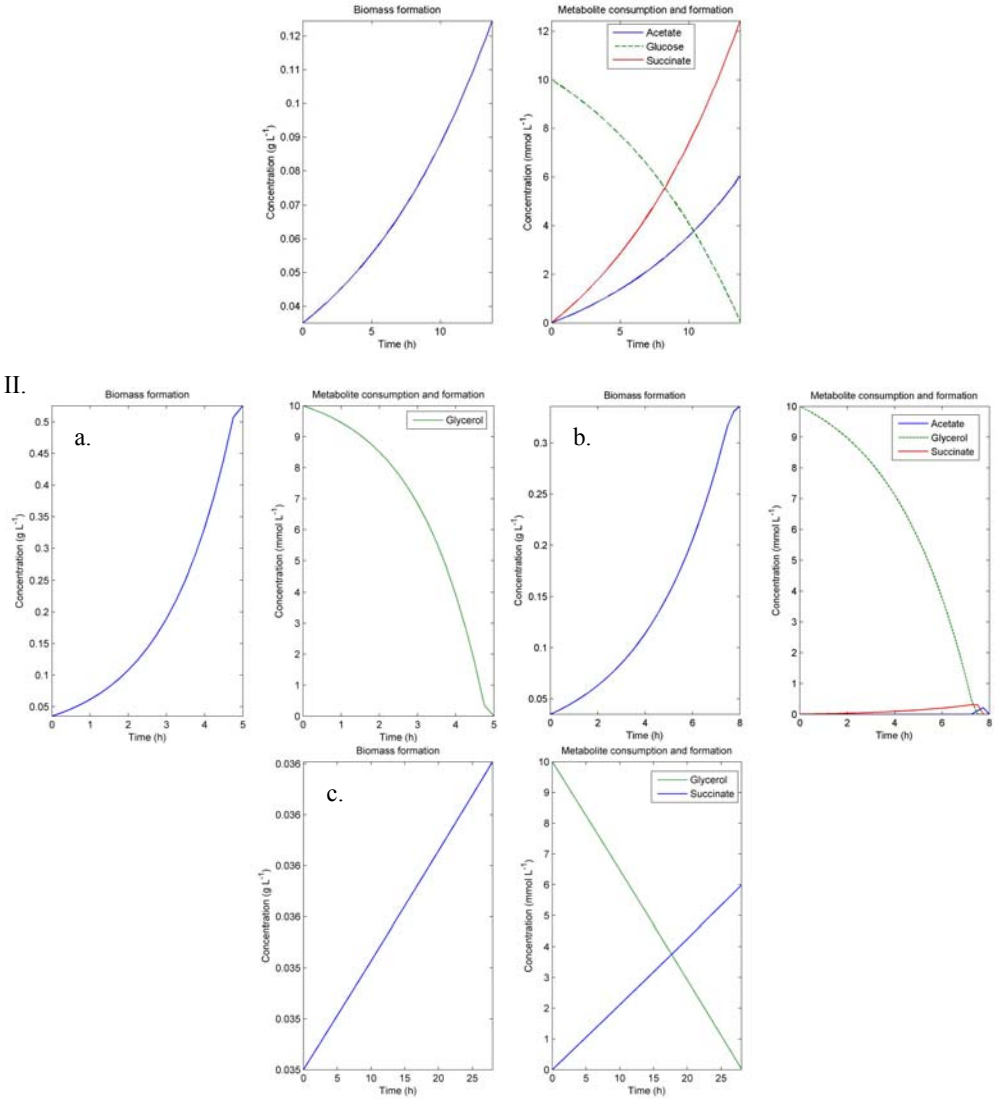


Fig.1. Dynamic FBA of mutant strains – the predicted biomass concentration, substrate and metabolite concentrations are shown as a function of time, I. Glucose substrate-a) aerobic, b) microaerobic, c) anaerobic conditions; II. Glycerol substrate

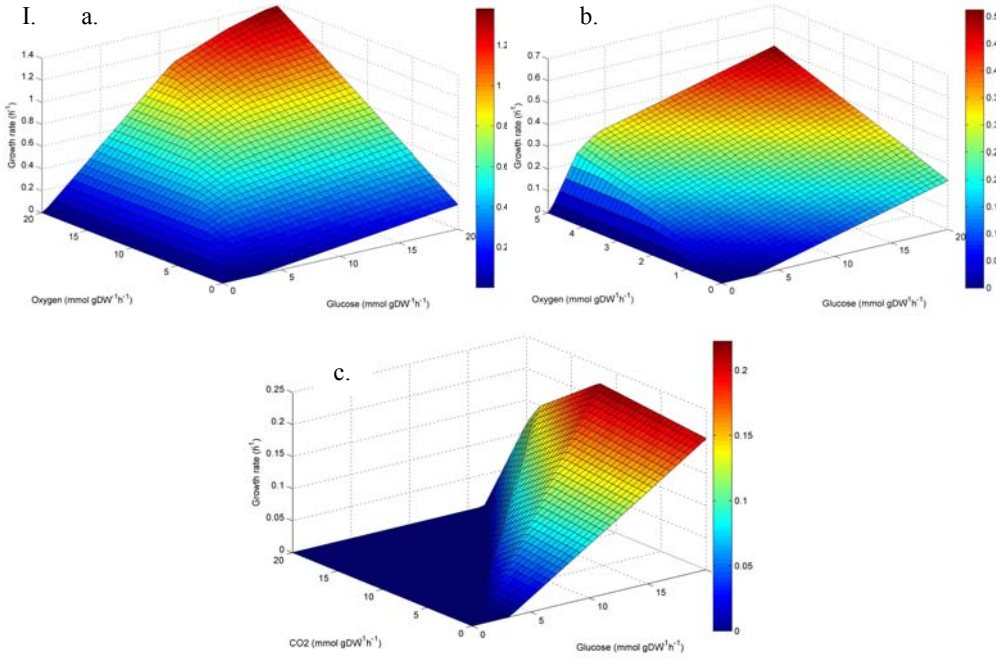
Fig. 1 shows the dynamic flux balance batch culture predictions for the mutant *E. coli* strains (succinic acid producers) under aerobic, microaerobic and anaerobic conditions on glucose (I) and glycerol (II) with minimal media. As we can observe growth rate was higher under aerobic and microaerobic conditions (results are in concordance with experimentally obtained), but the production rate of succinic acid was the highest under anaerobic conditions. Diauxic growth was

observed in case *b* where the metabolites produced before were metabolized. The batch time at which the predicted glucose was completely exhausted was over 3 h in *a*, 6 h at *b* and 14 h at *c*. The resulting batch time was longer under anaerobic conditions, approximately 15 h.

Using glycerol the model predicted similar diauxic growth pattern but a much longer batch time due to its slower growth on this substrate.

Phenotypic phase plane analysis

To analyze the relationship between genotype-phenotype phenotypic phase plane analyses was carried out for the wild-type and mutant strains and the resulting optimal metabolic states were studied. Results for mutant strains are presented in Fig.2.



II.

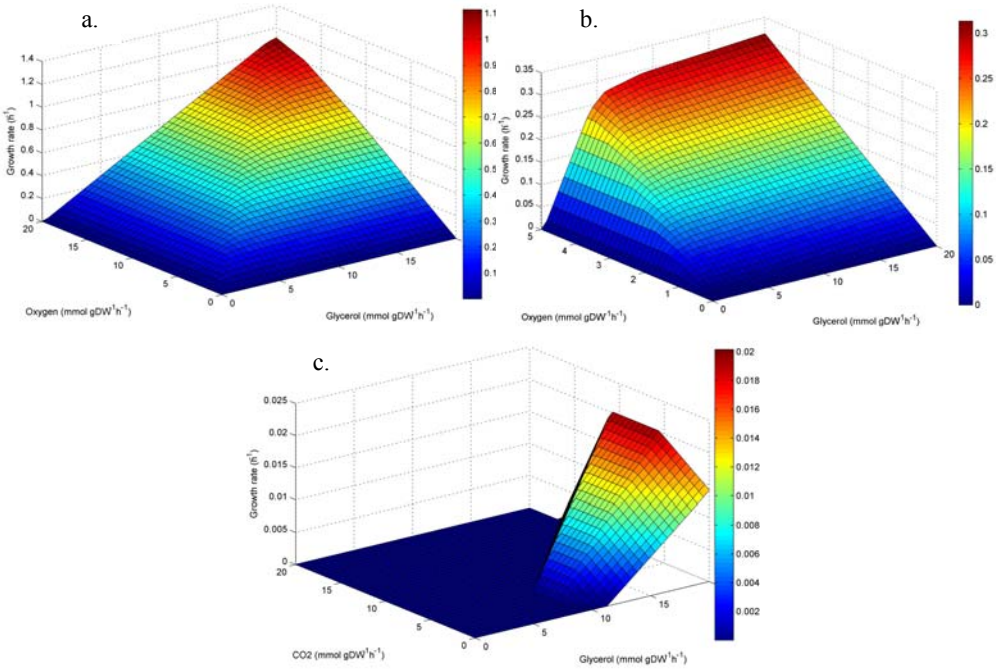


Fig.2. Phenotypic phase planes for growth of mutant strains with varying glucose (I) and glycerol (II) uptake rates, including the oxygen and carbon dioxide too. Three different environmental conditions were analyzed: a)aerobic, b)microaerobic and c)anaerobic

Phpp was performed to investigate the relationship between glucose/glycerol and oxygen uptake rate under aerobic and microaerobic conditions and on the other hand between glucose/glycerol and carbon dioxide.

It is clear from these plots that each surface has distinct regions, meaning qualitatively distinct phenotypes. There is a strong correlation between oxygen glucose/glycerol uptake rates, because oxygen is necessary to fully oxidize the substrate. Under anaerobic conditions in both case the CO_2 concentration negatively affected the mutant growth rate despite of the high glucose/glycerol uptake rate.

4. Conclusions

COBRA Toolbox was used to investigate the *E. coli* global metabolic capability to produce succinic acid from glucose and glycerol under anaerobic conditions, including the aerobic and microaerobic conditions. The flux distribution of the entire metabolism was estimated.

Analyses of the model gave insights into the metabolic phenotypes and possibility to test the genetic engineering effect on cell behavior under different

environmental conditions. Phpp analysis was used to predict the cell's metabolic states at various levels of glucose/glycerol, oxygen and carbon dioxide availability.

Modifying the complex redox balancing system suggested us the importance of this system to increase the rate of succinate production. As it was observed, by modifying the pyruvate metabolism a significant increase in succinate production could be obtained in both case of carbon sources.

It should be mentioned as well that eliminating the initial steps in glucose metabolism (PTS system) has a positive effect on succinate production. It is clear that complex modifications are needed to obtain a significant increased succinic acid yield.

Acknowledgements

This work was supported by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/107/1.5/S/76903 and by Collegium Talentum.

R E F E R E N C E S

[1] *S. Ohno, C. Furusawa, H. Shimizu*, "*In silico* screening of triple reaction knockout *Escherichia coli* strains for overproduction of useful metabolites", Journal of Bioscience and Bioengineerin, **vol. 2**, 2013, pp. 221-228.

[2] *S.J. Lee, D.Y. Lee, T.Y. Kim, B.H. Kim, J. Lee, S.Y. Lee*, "Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and *in silico* gene knockout simulation", Applied and Environmental Microbiology, **vol. 71**, 2005, 7880-7887.

[3] *H. Wu, Z. Li, L. Zhou, Q. Ye* "Improved succinic acid production in the anaerobic culture of an *Escherichia coli pflB ldhA* double mutant as a result of enhanced anaplerotic activities in the preceding aerobic culture", Applied and Environmental Microbiology, **vol. 73**, 2007, 7837-7843.

[4] *J. Wang, J. Zhu, G.N. Bennett, K.Y. San*, "Succinate production from different carbon sources under anaerobic conditions by metabolic engineered *Escherichia coli* strains", Metabolic Engineering, **vol. 13**, 2011, 328-335.

[5] *M.D. Blankschien, J.M. Clomburg, R. Gonzalez*, "Metabolic engineering of *Escherichia coli* for the production of succinate from glycerol", Metabolic Engineering, **vol. 12**, 2010, 409-419.

[6] *X. Zhang, K.T. Shanmugam, L.O. Ingram*, "Fermentation of Glycerol to Succinate by Metabolically Engineered Strains of *Escherichia coli"*, Applied and Environmental Microbiology, **vol. 76**, 2010, 2397-2401.

[7] *S. Hyohak, Y.L. Sang*, "Production of succinic acid by bacterial fermentation", Enzyme and Microbial Technology, **vol. 39**, 2006, 352-361.

[8] *X. Zhang, K. Jantama, K.T. Shanmugam, L.O. Ingram*, "Reengineering *Escherichia coli* succinate production in mineral salts medium", Applied and Environmental Microbiology, **vol. 75**, 2009, 7807-7813.

[9] *B.J. McKinlay, C. Vieille, G.J. Zeikus*, "Prospects for a bio-based succinate industry". Applied and Microbiol Biotechnology, **vol. 76**, 2007, 727-740.

[10] *T. Werpy, G. Peterson*, "Top value added chemicals from biomass", U.S. Department of Energy, Washington DC, 2004.

[11] *P. Kim, M. Laivenieks, C. Vieille, J.G. Zeikus*, "Effect of overexpression of *Actinobacillus succinogenes* phosphoenolpyruvate carboxykinase on succinate production in *Escherichia coli"*, Applied and Environmental Microbiology, **vol. 70**, 2004, 1238-1241.

[12] *L. Liang, R. Liu, J. Ma, K. Chen, M. Jiang, P. Wei*, "Increased production of succinic acid in *Escherichia coli* by overexpression of malate dehydrogenase", Biotechnology Letters, **vol. 33**, 2011, 2439-2444.

[13] *S.A. Becker, A.M. Feist, M.L. Mo, G. Hannum, B.Ø. Palsson, M.J. Herrgard, "Quantitative prediction of cellular metabolism with constraint-based* models: the COBRA Toolbox", Nature Protocols, **vol. 2**, 2007, 727-738.

[14] *J. Schellenberger, R. Que, R.M. Fleming, I. Thiele, J.D. Orth, A.M. Feist, D.C. Zielinski, A. Bordbar, N.E. Lewis, S. Rahmanian, J. Kang, D.R. Hyduke, B.O. Palsson*, "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0., Nature Protocols, **vol. 4**, 2011, 1290-307.

[15] *J.M. Lee, E.P. Gianchandani, J.A. Papin*, "Flux balance analysis in the era of metabolomics", Briefings in Bioinformatics, **vol. 7**, 2006, 140-150.

[16] *C. Xu, L. Liu, Z. Zhang, D. Jin, J. Qiu, M. Chen*, "Genome-scale metabolic model in guiding metabolic engineering of microbial improvement", Applied Microbiology and Biotechnology, **vol. 97**, 2013, 519-539.

[17] *D. McCloskey1, B.Ø. Palsson, A.M. Feist*, "Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli", Molecular Systems Biology, vol. 9, 2013, 661-676.

[18] *B. Papp, B. Szappanos, R.A. Notebaart*, "Use of Genome-scale metabolic models in evolutionary systems biology", Methods in Molecular Biology, **vol. 759**, 2011, 483-497.

[19] *A.L. Barabási, Z.N. Oltvai*, "Network biology: understanding the cell's functional organization", Nature Review Genetics, **vol. 5**, 2004, 101-113.

[20] *B. Szappanos, K. Kovács, B. Szamecz, F. Honti, M. Constanzo, A. Baryshnikova, G.G Dietrich, M.J. Lercher, M. Jelasity, C.L. Myers, B.J. Andrews, C. Boone, S.G. Oliver, C. Pál, B. Papp*, "An integrated approach to characterize genetic interaction networks in yeast metabolism", Nature Genetics, **vol. 43**, 2011, 656-662.

[21] *J.D. Orth, T.M. Conrad, J. Na, J.A. Lerman, H. Nam, A.M. Feist, B.O. Palsson*, "A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011", Molecular Systems Biology, **vol. 7**, 2011, 535-544.