I. Troch, F. Breitenecker, eds. ISBN 978-3-901608-35-3

MODEL AND GLOBAL SENSITIVITY ANALYSIS OF *E. COLI* CENTRAL METABOLISM

A. Tušek, Ž. Kurtanjek
University of Zagreb, Zagreb, Croatia

Corresponding Author: A. Tušek, University of Zagreb, Faculty of Food Technology and Biotechnology
Pierottijeva 6, 1000 Zagreb, Croatia; atusek@pbf.hr

Abstract. This work provides analysis of kinetic behavior of the central metabolism of *E. coli* upon glucose impulse during the initial transients of 15 seconds. The analysis is based on the model derived from dynamic measurements of the key intracellular metabolites. Response of the central carbon metabolism (glycolysis and pentose phosphate pathway) is decoupled from the anabolic and TCA systems by the transient measurements of the cofactors and oxaloacetate concentrations. The kinetic parameters are initially estimated by the nonlinear Least Squares Method (with Marqardt minimization) and improved by several global optimization algorithms (simplex Nelder-Mead, Simulated Annealing and Differential Evolution). However, due to severe ill-conditioned problem, large errors in the estimates are inherently present. The focus of this research is to reveal which are the most important enzyme effectors, reflected by the corresponding kinetic parameters, responsible for modeling of the input-output fluxes of the central metabolism. Applied is the Fourier Amplitude Sensitivity Test (FAST) for global sensitivity analysis. Identified are the key kinetic parameters responsible for the following fluxes: phosphotransferase system (PTS), nucleotide biosynthesis and pyruvate to biomass. The results could be potentially applicable for understanding of the metabolism regulation and for rational application of genetic engineering.

1 Introduction

The mathematical model applied in this work is based on the experimental data by Degenring *et al*. [1-2]. The data are collected by the automated sampling method and biochemical tests. Measured are responses of the intracellular glycolysis metabolites of *Eschericha coli* cultivated in a batch reactor under balanced growth conditions. The glucose impulse was introduced after a period of glucose deprivation. Each sample is immediately frozen upon automatic withdrawal from the bioreactor in order to ensure deactivation of the consequent biochemical reactions which would change the intracellular composition of the metabolites. Intracellular composition is sampled with 4 Hz frequency. The model consists of mass balance equations derived from the glycolysis network with included phosphate penthose pathway and Entner-Dourdoroff shunt. The kinetic rate functions and initial parameter values are selected from the biochemical data basis BRENDA and ExPASy [1-2]. Mathematical model that gives the best fit to the experimental data is composed of 22 biochemical reactions and 122 kinetic parameters. In the process of the model adaptation, the parameters are fitted to the experimental data by the method of least squares and Marquardt optimization algorithm. The obtained model describes the network of biochemical reactions included in central metabolism of *E. coli* leading from glucose to energy and cell building components (Figure 1). Balances for the cofactors (ATP, ADP, NADH, NAD), acetylcoenzyme A, citrate and oxaloacetate are not explicitly accounted in the model, i.e. are not included into the reaction kinetics, as they were experimentally determinated in form of time varying interpolation functions [1-2]. Experimental account of the cofactors enables dynamic decoupling of the central metabolism from TCA cycle and the anabolic system. Čerić and Kurtanjek [3] improved the original model by Degenring *et al*. [1-2] by: closure of Entner-Doudoroff pathway with pyruvate balance, introduction of phosphoenolpyruavate carboxylase and carboxykinase reactions in the balance of phosphoenolypyravate, account for loss of pyruvate in biomass synthesis, change in kinetic rate expressions for several enzymes, and re-estimation of the kinetic parameters by application of the global optimisation algorithms. The initial parameter estimates are used as initial values for application of the several global optimization algorithms: simplex Nelder-Mead, Simulated Annealing (SA) and Differential Evolution (DE) method (a variant of the genetic algorithm GA). The DE method gave the most of the improvement in minimization of the variance between the model predictions and experimental data. The modified model correctly predicts observed oscillatory response to glucose impulse in concentrations of pyruvate and D-ribose-5-phosphate. The analyzed reaction map is depicted in Figure 1 and the kinetic rate expressions and parameters are available from URL address given in [4].

Although, the parameter fitting process resulted in a relative small error, very high dimension of the parametric space resulted in a severely ill-conditioned problem and consequently the parameter confidence intervals are very large leading to significant uncertainties in the "true" parameter values. To resolve the problem of the parameter selection, a parameter sensitivity analysis is conducted. Due to strong regulation of the metabolic fluxes, standard local (infinitesimal) one-to-one parameter sensitivity analysis does not provide insight into the highly regulated and globally concerted behaviour. Hence, this work is aimed for a global sensitivity analysis to

provide key parameters by simultaneous change of the kinetic parameters in a finite range of 132 dimensional space.

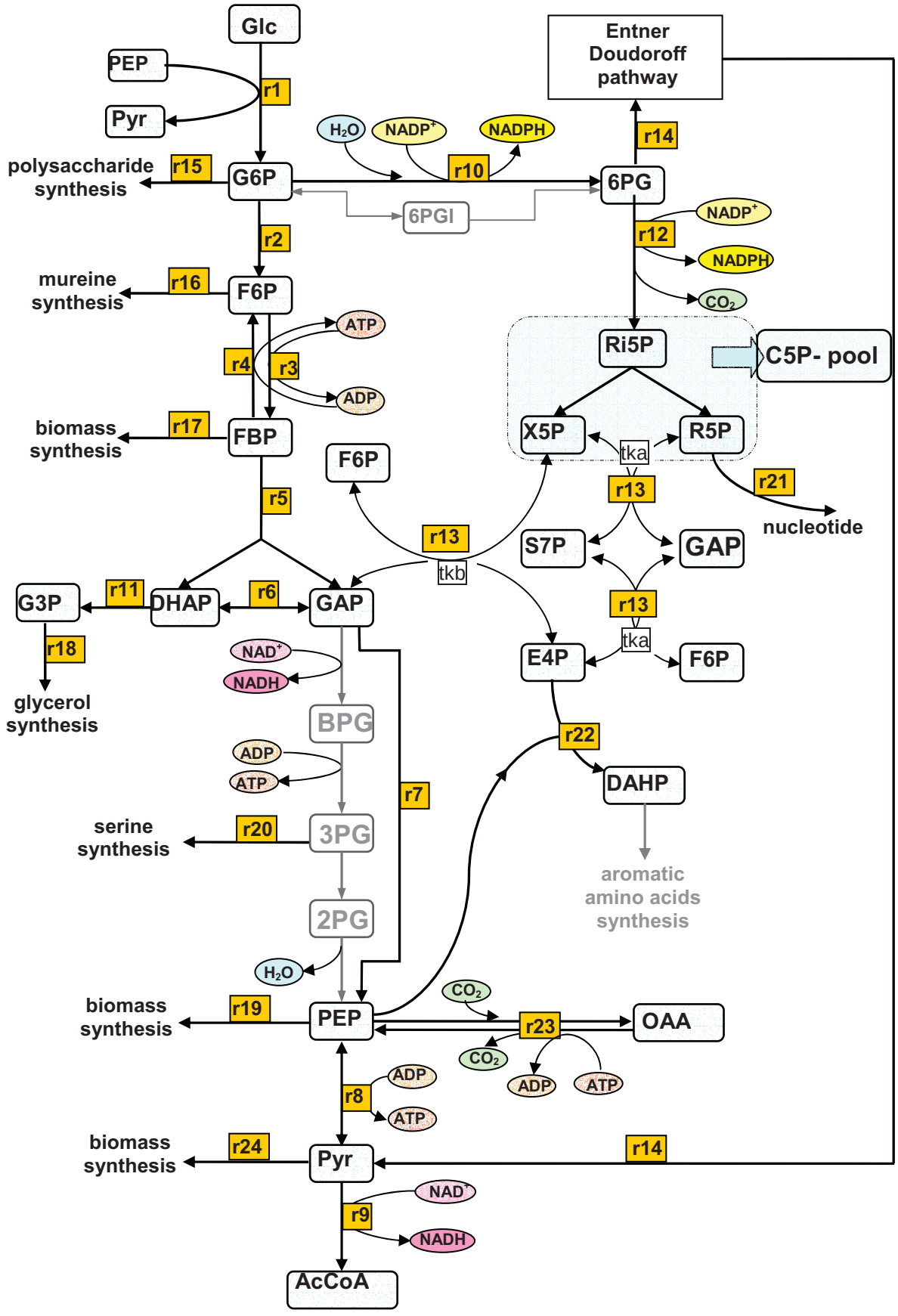

Figure 1. Metabolic network of *E. coli* central metabolism.

For numerical evaluation of the global sensitivity analysis FAST algorithm over Monte Carlo simulation is chosen due to its computing efficiency. By the FAST method a number of needed simulations is considerably reduced [5-6] and here is effectively executed on a standard PC computer with the advantage of numerical efficiency provided with Wolfram Research *Mathematica* NDSolve integration for ODE stiff systems [7]. For comparison of numerical efficacy, number of needed simulations by Monte Carlo simulation could be approximately 15 000 (assuming minimal 300 random simulations for each parameter) which is considerably higher than 2 000 simulations needed by the applied FAST algorithm. The needed number of simulations by FAST is determined by trail and error method until the sensitivities become unaffected by the number of samples.

2 Global sensitivity analysis

Global sensitivity analysis is a statistical method of analysis of effects of relative change in model responses due to variation of model input parameters over a reasonable finite range [5-6]. Effects of uncertainties of model parameters are analyzed numerically by Fourier Amplitude Sensitivity Test (FAST) method. Application of FAST algorithm provides sensitivities to large and simultaneous change of the complete set of the model parameters. This method is based on transformation of multidimensional space of model parameters to one dimension space of single parameter s. Normalized parameters x_i are sampled from assumed range of minimal and maximal values and uniform probability density function. The values are sampled along piece wise linear functions defined by:

$$x_i = \frac{1}{2} + \frac{1}{\pi} \cdot \arcsin\left(\sin\left(\pi \cdot \omega_i \cdot \mathrm{s} + \varphi_i\right)\right) \tag{1}$$

The model parameters are the specific rate kinetic constants which are linearly depended on the scale transformations of the standard (normalized) parameters. The parameter s covers the rage $s \in [-1,1]$. Each parameter transformation x_i is associated with two randomly selected parameters, frequency ω_i and phase angle φ_i. The parameters are randomly associated with the odd frequencies in the range from 29 to 301 and the phase angles in the range from -2π to 2π. The unbiased choice of frequencies and phase angles is checked by calculation of the covariance matrix of the complete set of the parameters. In this work obtained is the maximum covariance of order 10^{-5} which ensures effectively random sampling process needed for the sensitivity analysis. Importance of individual parameters is analyzed by general parameter sensitivity theory based on their corresponding variance contributions [5-6]. Results of the output function are decomposed into Fourier series with coefficients:

$$A_\omega = \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} f(s) \cdot \cos(\omega \cdot s) \cdot ds \tag{2}$$

$$B_\omega = \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} f(s) \cdot \sin(\omega \cdot s) \cdot ds \tag{3}$$

The total dispersion (variance) D_T is determined from the Fourier coefficients:

$$D_T = 2 \cdot \sum_{\omega=1}^{\infty} \left(A_\omega^2 + B_\omega^2\right) \tag{4}$$

And individual parameter contribution in the total dispersion is calculated by the corresponding parameter harmonics:

$$D_i = 2 \cdot \sum_{\omega=k \cdot \omega_i}^{\infty} \left(A_\omega^2 + B_\omega^2\right) \tag{5}$$

The first and second order sensitivity coefficients are calculated by:

$$S_i = \frac{D_i}{D_T} \tag{6}$$

$$S_i^{tot} = 1 - \frac{D_{-i}}{D_T} \tag{7}$$

3 Results and discussion

Model of *E. coli* central metabolism includes 24 biochemical reactions which are strongly depended. The main effect is glucose intake by PTS system which provides a cell with free energy and production of cell building blocks. It also functions in a regulatory capacity, controlling the rates of cellular carbon and energy metabolism [8]. To analyze the glucose consumption throughout the metabolic network 1 C-mol balance for the input and output fluxes are evaluated, and the results are depicted in Figure 2. There are two positive in-fluxes, PTS and the reaction of PEP production from OAA. These two fluxes are considered as the inputs to the central metabolism. They are closely related since for glucose transport into the *E. coli* cell by PTS mechanism PEP is required. Since the glucose impulse is introduced after carbon source starvation, glucose transport (Figure 2, curve A) into the cell is very fast and high amount of PEP is needed, which results in very high activity of the reaction leading from OAA into PEP (Figure 2, curve B). In the opposite direction, reaction (from PEP to OAA) serves as an anaplerotic reaction of TCA cycle; it is required to replenish the pool of tricarboxylic acid cycle intermediates. Reaction of gluconeogenic conversion of oxaloacetate to phosphoenolpyruvate is considered to be inactive in *E. coli* grown in glucose containing media [9]. However, under these experimental conditions it is a dominant reaction tided to PTS. Here, under the given experimental conditions, prior to the glucose impulse, intracellular stored energy is used for the maintenance of the vital cellular functions, such as functionality of a cell membrane. Three the most dominate output fluxes are represented in Figure 2 with negative values. They are denoted by the corresponding letters: from pyruvate to AcCoA (curve C); polysaccharide synthesis (curve D); and from pyruvate to biomass (curve E). The input-output 1 C-mol flux balances for the central metabolism reveals that 50 % of glucose intake is directed toward polysaccharide synthesis, 30 % is the flux of oxaloacetate to phosphoenolpyruvate (PEP) production needed for PTS activity. The flux toward biomass synthesis from pyruvate accounts for 11.7 % of the glucose carbon intake. Analysis of the 1 C-mol fluxes also reveals that PTS is the most active during first few fractions of a second, followed by the flux form OAA to PEP.

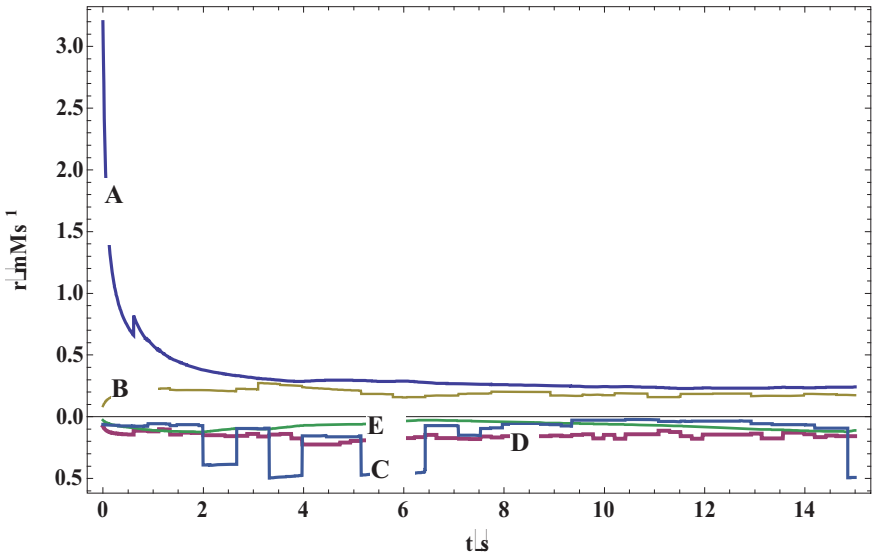

Figure 2. 1-C mol input PTS and the output fluxes during 15 s from the central metabolism. The fluxes are marked as: A) PTS, B) OAA to PEP, C) pyruvate to AcCoA, D) polysaccharide synthesis, E) from pyruvate to biomass.

The complete set of parameters in the expanded model has132 kinetic parameters which are transformed along Lissajous curves to one dimensional space of a single parameter *s*. Each of the parameters is randomly and mutually uncorrelated in the range of ± 50 % of its nominal value with the uniform probability distribution function.

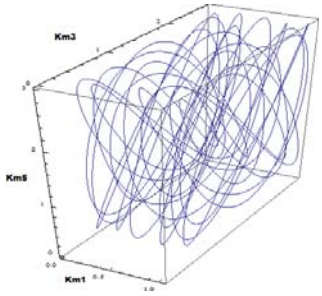

Figure 3. Subset of the parametric space, spanned by the saturation constants for PTS (Km1), phosphoglucosisomerase (Km3), and polysaccharide synthesis (Km5), filled with Lissajous curve for uncorrelated parameter sampling.

As an illustration of the filling process of the parametric space, a case of a 3D parameter subspace (saturation constants for PTS (Km1), phosphoglucosisomerase (Km3), and polysaccharide synthesis (Km5)) is presented in Figure 3. From the transformed set of parameters sampled are 2000 cases in steps of $\Delta s = 0.001$ from the range $s \in [-1,1]$ and each of them is used for simulation glucose impulse distribution throughout the central metabolism given by the 10 differential equations. The number of needed samples (simulations) is determined by a trial and error method until the power spectrum obtained by Fourier decompositions becomes essentially invariant to the sample size. Results of simulation, i.e. each input or output metabolic flux of interest, are expanded into a finite Fourier series. The cut-off frequency in the expansion is approximately three times higher of the highest fundamental harmonic, which has been experimentally shown to be sufficient to reach the region of negligible power. An example of the power spectrum for input PTS flux is shown in Figure 4. The result shows the fast decay of the power spectrum, i.e. that most of the contributions in the spectrum are covered in the range of the first three highest harmonics.

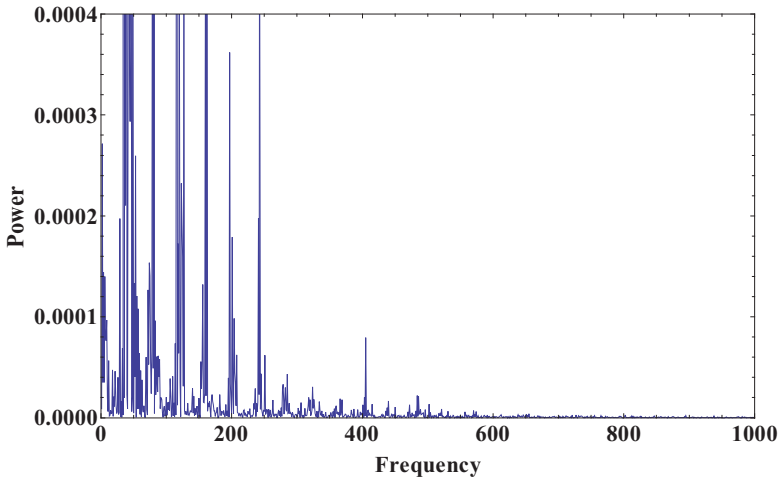

Figure 4. Power spectrum of the Fourier transformation of PTS flux during the first 5 seconds after the glucose impulse.

Results of the output function are decomposed into the Fourier series, and sensitivity coefficients are calculated. Sensitivities on the complete kinetic parameters of the three fluxes (PTS, nucleotide synthesis, and flux from pyruvate to biomass) during first 5 s after the impulse fluxes are analyzed and presented in Figure 5. The percentages of the flux sensitivities for individual enzymes S_{ENZYME} are calculated as the sum of the partial sensitivities S_i of each kinetic parameter included into the corresponding model of the enzyme kinetics:

$$S_{ENZYME} = \sum_{i=k}^{i=l} S_i \quad \text{for } i = k, k+1, k+2\Lambda\ , l-1, l \in \text{Model}(\text{Enzyme}) \tag{8}$$

where the kinetic parameters included into the activity of a particular enzyme are indexed from k to l. The same formula (8) is applied in the cases when several reaction steps are lumped into a single pseudo reaction, i.e, when a pseudo enzyme is present in the model. The results presented in Figure 5 show the well known fact that PFK is the key enzyme for regulation of glycolysis. In the phosphotransferase system (Figure 5A), the most sensitive are the parameters of phosphofructokinase (60 %), PTS is important for regulation of phosphofructokinase activity, and the parameters included in PTS mechanism (20 %), while the rest 20 % is the contribution of the rest of enzymes (parameters). This observation holds for the PTS transients after the first few fractions of the second upon the glucose impulse. The controlling effect of PFK is initially negligible, but becomes important in the region as a steady PTS flux is established (Figure 2). Sensitivity analysis of nucleotide synthesis flux (Figure 5B) reveales that it mostly depends on the parameters of the nucleotide biosynthesis mechanism (39 %) and on the mureine synthesis parameters (30 %). Pyruvate to biomass flux (Figure 5C) shows the highest sensitivity on the parameters responsible for biomass forming mechanism (23%), PTS (15 %), PFK (14 %) and aldolase (12 %). On average, the contribution of the rest of the metabolic rates is between 15-30 %.

4 Conclusions

The presented results of the non-stationary 1-C mol balances and the global parameter sensitivity analysis enable inference on metabolic flux regulation and detection of the most sensitive parameter which is related to interaction of a given enzyme and substrate, cofactors and metabolites. Its main implications are derived from the systemic or global sensitivity results which are properties of “the whole” rather than a single one-to-one local analysis. The classical local analysis is constrained to the selection of a steady state (cell homeostasis), while in the impulse experiment the assumption of a steady state is inappropriate. Selected is a relatively wide range of

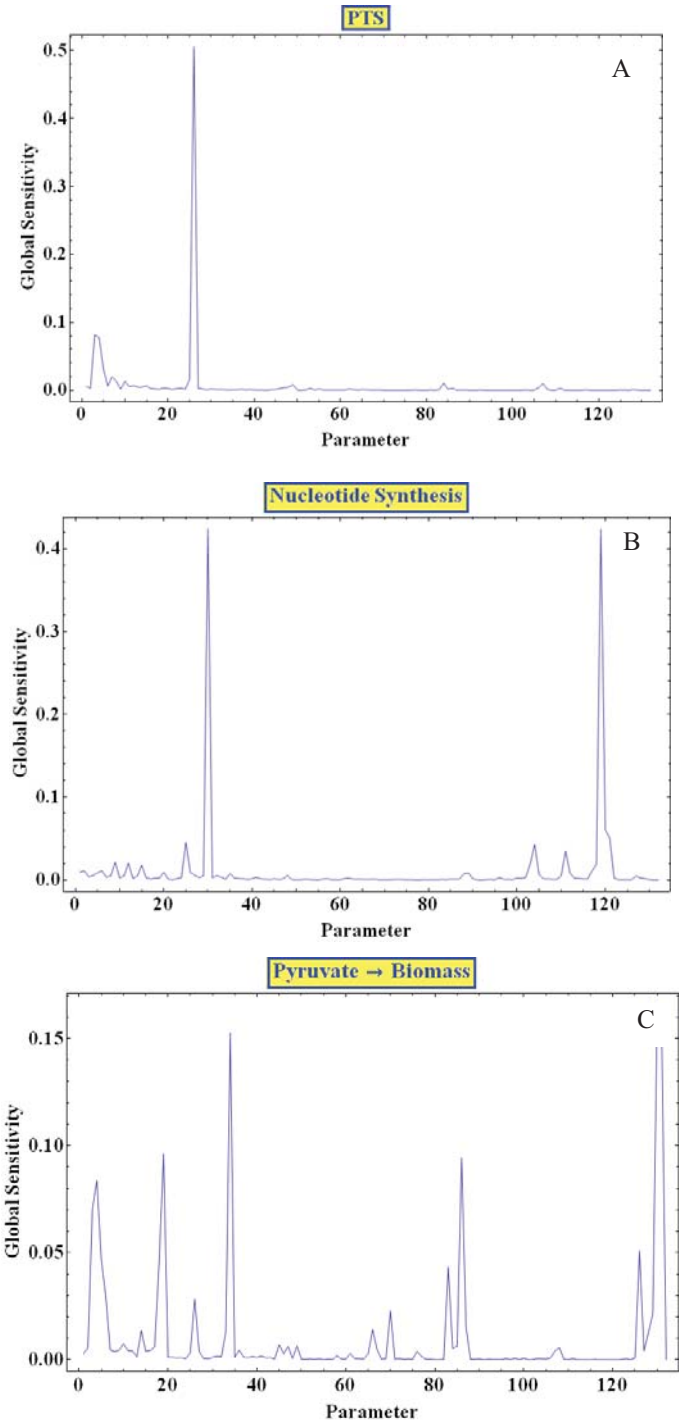

Figure 5. Global sensitivities S_i of the input-output fluxes on the kinetic parameters evaluated during 5 s after the glucose impulse. The maximum sensitivities are:

A) (1) enzyme phosphofructokinase S_{26}(npfk3) = 0.506;
(2) PTS mechanism $S_3(K_{m2})$ = 0.08, $S_4(K_{i1})$ =0.077

B) (1) mureine synthetase $S_{30}(K_{m9})$ = 0.432; (2) nucleotide synthetase S_{119}(nrppk1) = 0.423;
(3) phosphofructokinase $S_{25}(K_{i1})$ =0.045;
(4) phospho-gluconat-dehidrogenase S_{104}(npgdh2) = 0.043;
(5) transketolase-transaldolazeS_{111}(vftkata) = 0.034

C) (1) pyruvate to biomass S_{130}(vfpyrbm) = 0.165; S_{139}(Km44) = 0.152;
(2) aldolase S_{34}(vmaxaldo) =0.1527;
(3) phosphofructokinase S_{19}(L) = 0.00961;
(4) PTS $S_4(K_{i1})$ = 0.083;
(5) pyruvate-dehydrogenase $S_{86}(K_{m28})$ = 0.094.

the parameter variation, -50 % to + 50 % of the nominal values, which reflects how each of the model parameters is responsible for the model performance when all the other parameters are simultaneously varied. The analysis is aimed for identification of metabolic regulation. It potentially elucidates the main allosteric effects in enzyme regulation. On the other hand, results of the global sensitivity are also crucial for model improvement and parameter optimization. It enables a modeler with selection of the most suitable model for a given set of experimental data, leading to model reduction and increase of its predictive potential.

From systems view, this work constitutes an attempt towards dynamic mathematical modelling of regulation *E. coli* central metabolism upon glucose impulse perturbation, which is needed for the rational optimization of biotechnological processes. Due to the modelling of the very early response of the perturbation, limited to the first 15 seconds, effects of the genetic level regulation may be considered ineffective. In addition, by experimental account of the intracellular cofactors, achieved is separation from the anabolic effects on the regulation on the level of protein phosphorylation. Hence, the main result of this work is focused on the modelling of metabolic regulation (effectors) and discern of the key enzyme-metabolite interactions (reflected by the corresponding kinetic parameters) on the main metabolic fluxes by the application of the global sensitivity analysis.

However, for a rational development of optimized genetically modified biotechnological processes a cell systemic approach is needed which simultaneously accounts for metabolite level regulation and the global genetic regulatory network and molecular signalling network of the regulation of the central carbon metabolism [10].

5 Nomenclature

Metabolites

AcCoA	acetyl-coenzyme A
ADP	adenosindiphosphate
ATP	adenosintriphosphate
BPG	glycerate-1,3-bisphosphate
CIT	citrate
C5P	lumped pentose phosphate pool
DAHAP	7-phospho-2-dehydro-3-deoxy-D-arabinoh-eptonate
DHAP	glycerine phosphate
ED	Entner-Doudoroff pathway
E4P	D-erythrose-4-phosphate
FBP	β-D-fructose-1,6-bisphosphate
F6P	β-D-fructose-6-phosphate
GAP	glyceraldehydes-3-phosphate
GTP	guanosine triphosphate
G3P	glycerol-3-phosphate
G6P	α-D-glucose-6-phosphate
NAD	diphosphopyridindinucleotide (oxidized)
NADH	diphosphopyridindinucleotide-phosphate (oxidized)
MUR	mureine
OxAc	oxaloacetat
PEP	phosphoenolpyruvate
PP	phospho-pentose pathway
2PG	glycerate-2-phosphate
3PG	glycerate-3-phosphate
6PG	6-phospho- D-gluconate
PYR	pyruvate
R5P	D-ribose-5-phosphate
Ri5P	D-ribulose-5-phosphate
S7P	D-sedoheptulose-7-phosphate
X5P	D-xylulose-5-phosphate

Enzymes

pts	phosphotransferase system
tka	transketolase A
tkb	transketolase B

Abrevations

ATP	adenosine triphosphate

FAST Fouirer Amplitude Sensitivity Test
NADH nicotinamide adenine dinucleotide
NADPH nicotinamide adenine dinucleotide phosphate
r_i i-th flux (metabolic reaction rate)

6 References

[1] Degenring, D., Fromel, C., Dikata, G. and Takors, R.: *Sensitivity analysis for the reduction of the complex metabolism models.* Journal of Process Control, 2004, 14:729-745.

[2] Degenring, D.: *Erstellung und Validierung mechanistischer Modelle für mikrobiellen Stoffwechsel zur Auswertung von Substrat-Puls-Experimenten*. Dissertation. University of Rostock, 2004.

[3] Čerić, S. and Kurtanjek, Ž.: Model *identification, parameter estimation and dynamic flux analysis of E. coli central metabolism*. Chemical and Biochemical Engineering Quarterly, 2006, 20(3):243-253.

[4] http://www.pbf.hr/hr/zavodi/zavod_za_procesno_inzenjerstvo/laboratorij_za_mra/modeliranje_biotehnoloskih_procesa

[5] Saltelli, A., Ratto, M., Tarantola, S. and Campolongo, F.: *Sensitivity analysis of chemical models.* Chemical Reviews, 2005, 105(7):2811-2827.

[6] Cukier, R.I., Fortuin, C.M., Shuler, K.E., Petschek, A.G. and Schaibly, J.H.. *Study of the sensitivity of the coupled reaction systems to uncertainties in rate coefficients. I theory.* The Journal of Chemical Physics, 1970, 3873-3878.

[7] Wolfram Research "*Mathematica*", v. 6.0 (2008).

[8] Saier, M.H. JR.: *Bacterial Phosphoenolpyruvate: Sugar Phosphotransferase Systems: Structural, Functional, and Evolutionary Interrelationships*. Bacteriological Reviews, 1977, 41(4):856-871.

[9] Sauer, U., Lasko, D.R., Fiaux, J., Hochuli, M., Glaser, R., Szyperski, T., Wüthrich, K. and Bailey, J.E.: *Metabolic flux ratio analysis of genetic and environmental modulations of Escherichia coli central carbon metabolism.* Journal of Bacteriology, 1999, 181(21):6679-6688.

[10] Hardiman, T., Lemuth, K., Keller, M., Reuss, M. and Siemann-Herzberg, M.: *Topology of the global regulatory network of carbon limitation in Escheichia coli*, Journal of Biotechnology, 2007, 132(2): 359-374.