Modelling and Optimisation of Central Metabolism Response to Glucose Pulse

Sonja Čerić, Želimir Kurtanjek
University of Zagreb, Faculty of Food Technology and Biotechnology,
Pierottijeva 6, 10000 Zagreb, CROATIA
zkurt@pbf.hr

Abstract. *The application of global optimisation algorithms for modelling of enzyme kinetics and parameters of the central metabolism is analyzed. The model is a set of 10 highly complex stiff ordinary differential equations with 132 adaptive parameters. The efficacies of Nedler-Mead, simulated annealing, and differential evolution for minimization of variance between simulations of a model of the central metabolism and experimental data of the corresponding intracellular key metabolites are compared. Experimental data obtained during response of population of E. coli cultivated in a batch bioreactor are applied. Data are obtained during transients of 15 seconds upon a glucose impulse.*

Results are discussed in view of metabolic flux analysis and genetic engineering.

Keywords. metabolic networks, Nedler-Mead, simulated annealing, differential evolution

1. Introduction

Advances of bioinformatics based on automatic determination of genome leads to a possibility to development of computer models of a living cell. Living cells are highly complex regulated and adaptive dynamic systems, and computer models need to account for the main regulation mechanisms by which a cell grows and adapts to a changing environment. One of the modelling approaches, which has a potential to provide insight in the key features, is based on dynamic material and energy balances of metabolite [1]. This approach is seriously limited by the difficulties in estimation of enzyme kinetic functions and numerous parameters. Choice of the kinetic functions is usually limited to the fundamentally founded and generally accepted Michaelis-Menten functions with several activation and inhibition effects. Unfortunately, kinetic parameters, at present knowledge, can not be determined from fundamental principles, and have to be based on empirical methods.

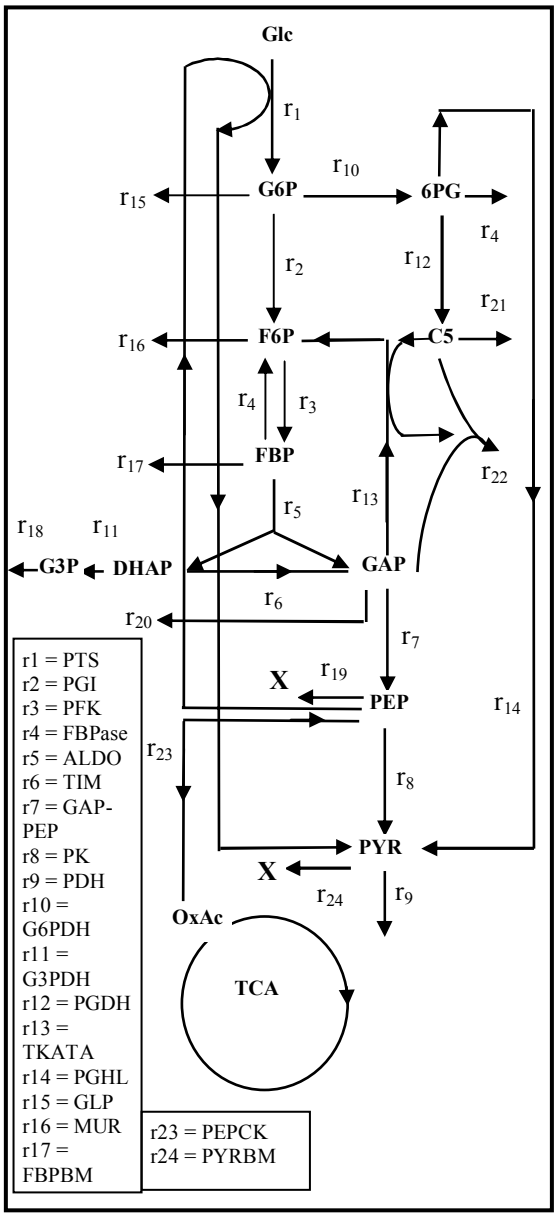

Figure 1. Model of flux network of *E. coli* central metabolism.

Usually, the parameters are estimated by the statistical least squares method to minimize variance between a model predictions and experimental observations. This approach leads to numerically very difficult optimisation. Due to significant measurement errors, modelling inaccuracies, and overlapping effects of kinetic parameters, a response surface has multiple local minima, and numerous "flat" patches. A good initial point in parametric space is usually unknown, and an unwary algorithm that always accepts a step which minimizes a function freezes in a local minimum. In such situations gradient based methods have poor efficacy, and require multiple restarts with new guess of initial points. A potential alternative is application of non-gradient methods with a chance to perform as a global optima search algorithm.

Here are compared three attractive non-gradient methods, of which two mimic how nature, biology (differential evolution, a variant of genetic algorithm) and thermodynamics (simulated annealing), adapts and self-organise to optimise its structure. These methods are multidimensional search algorithms, which are generally considered to have a chance to perform global optimisation. In this work they are applied for minimisation of the sum of squares of errors between the experimental data and a model prediction from the experiment with central metabolism response to glucose impulse.

2. Model

Model is based on experimental data published by D. Degenring [2,3] from rapid sampling experiment with *E. coli* exposed to glucose pulse. The results from simulations of the original model showed significant difference from experimental data for some metabolites and the original model was altered and improved by S. Čerić [3]. The changes introduced include insertion of two novel reactions, (PEP carboxykinase and conversion of pyruvate into biomass), alteration of three kinetic rate equations (PTS, phosphoglucose-isomerase and aldolase). The altered model predicts increase through Entner-Dourdof pathway, and in this work the model is further extended by closing this flux into pyruvate pool. The alterations of the original model are depicted in Fig. 1 as the rates r_{14} and r_{23}.

The dynamic balances (1) are given by the product of the stoichiometric matrix $\boldsymbol{\alpha}$ and the vector of reactions **r**. The stoichiometric matrix has a full rank which proofs independence of the mass balances. In Fig. 1. are listed components of the reaction vector, i.e. fluxes. The total number of fluxes is 24, of which 11 fluxes are intracellular, 13 are extracellular (1 is inbound measured assimilation of glucose, and 12 are unmeasured outbound fluxes from the central metabolism into anabolite reactions). Based on the flux count and the number of intracellular balances (1) determined is the model degree of freedom of 13. High degree of freedom of the model profoundly effects efficacy of the model adaptation and validity.

$$\alpha = \begin{pmatrix}
+1 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 \\
-1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & +1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & +1 & +1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & +1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & +1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 \\
0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 \\
0 & +2 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & -3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & +1 & -1 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & -6 & 0 & 0 & -4 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0
\end{pmatrix};$$

$$\frac{d}{dt}\vec{c} = \boldsymbol{\alpha}^T \cdot \vec{r}\left(\vec{c}, \vec{c}_u, \vec{\beta}\right) \qquad (1)$$

The mass balances of the intracellular species $\boldsymbol{c}$ are dependent on unbalanced cofactors $\boldsymbol{c}_u$. The cofactors are measured under experimental transient conditions and are included into the model as continuous interpolated functions. The rates include forward and backward reactions and reaction kinetic functions are of Michaelis-Menten type. For example, the first flux is given by:

$$r1 = v_{PTS} = \frac{V_{f,pts} \cdot Glc \cdot PEP}{\left(K_{m,Glc,pts} + Glc\right) \cdot \left(K_{m,PEP,pts} + PEP\right)} \cdot \frac{1}{\left(K_{i,G6P,pts} + G6P\right)^{npts}} \qquad (2)$$

which is kinetics of assimilation of extracellular glucose by phospho-transferase systems (PTS). The kinetic expression of all of the reactions is given by S. Čerić [4]. The kinetic model includes

132 parameters which need to be estimated through minimisation of the sum of squares (3) of differences between predicted by model intracellular concentrations $\boldsymbol{c}$ and the measured concentrations $\boldsymbol{c}_{\mathrm{m}}$.

$$S^2\left(\vec{\beta}\right) = \left(\vec{c} - \vec{c}_m\right) \cdot \left(\vec{c} - \vec{c}_m\right)^T \qquad (3)$$

The experimental methods are highly complex and automated by a robotic system [2-3] enabling fast sampling with frequency of 4 Hz. Data are collected during the first 17 s of the transients induced by the glucose impulse. Collected are data for 21 intracellular species, each 4 times sampled in a second, which yields in total a data set of 1428 concentrations. Random dispersion of the experimental data is very pronounced, with approximate variability coefficient of 0,25. An example of the response of intracellular glucose-6-phosphat (G6P) with dispersion of data points is depicted in Fig. 2.

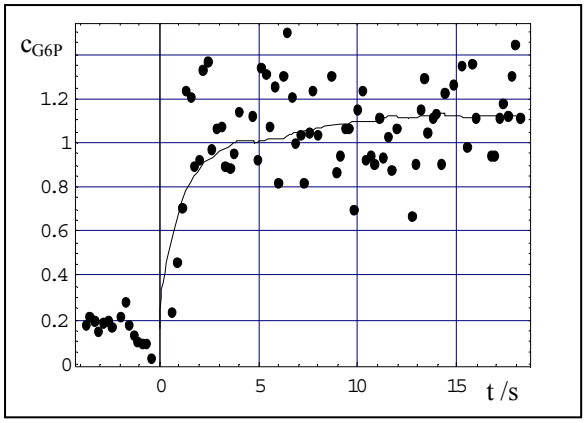

Figure 2. Experimental data (□) and the model simulation (┘) of G6P (mmol/L) intracellular concentration upon glucose impulse.

3. Model optimisation

In this work are applied and compared three potentially effective algorithms for global optimisation over a continuous multidimensional space of kinetic parameters in the model of central metabolism. The model simulation and parameter optimisation is realised in Wolfram Research "*Mathematica*" [5].

3.1. Nedler-Mead optimisation

A simple, but very "code cost effective" method of Nedler-Mead is applied. The method can be described as a "down hill " simplex descent. The method has a "reputation" of being the best first choice, which will work provided it has a suitable initiation. It starts with an initial (usually randomly picked) simplex of N+1 vertices in the N dimensional space. The simplex edges are taken as unit vectors, and at the vertices objective function is evaluated, and values sorted in decreasing order. In the next step the vertex with the highest value is improved. Centroid of N vertices is calculated, excluding the highest (labelled with i=1), by:

$$\vec{x}_{mean} = \frac{1}{N} \cdot \sum_{i=2}^{N+1} \vec{x}_i \qquad (4)$$

From this a new search direction is determined by reflection from the centroid

$$\vec{x}_1^{new} = \vec{x}_{mean} + \left(\vec{x}_{mean} - \vec{x}_1\right) \qquad (5)$$

If $f\left(\vec{x}_1^{new}\right) < f\left(\vec{x}_{N+1}\right)$ then the new vertex is in a downhill direction. A better point is attempted by doubling the move. If after the reflection the new point is still the highest, a reflection and shrinking is attempted:

$$\vec{x}_1^{new} = \vec{x}_{mean} + \frac{1}{2} \cdot \left(\vec{x}_{mean} - \vec{x}_1\right) \qquad (6)$$

If this does not improve, then just shrinking is attempted:

$$\vec{x}_1^{new} = \vec{x}_{mean} - \frac{1}{2} \cdot \left(\vec{x}_{mean} - \vec{x}_1\right) \qquad (7)$$

Finally, if this also fails, then all the vertices are shrunk toward the best one

$$\vec{x}_i^{new} = \vec{x}_i - \frac{1}{2} \cdot \left(\vec{x}_i - \vec{x}_{N+11}\right) \quad i = 1,2..N \qquad (8)$$

Through successive iterations, a simplex tumbles downhill with change of scale continuously adapting in size and location to configuration of the response surface. However, for efficiency and success to locate a global minimum, a good starting simplex is required, and restart with different initial simplexes is needed.

3.2. Simulated annealing algorithm

Simulated annealing is inspired by thermodynamic consideration of energy distribution of multi-component systems. By the law of statisti-

cal thermodynamics, a probability p to find a system in a state with energy E at temperature T is given by Boltzman law:

$$p(E,T) \propto e^{-E/kT} \tag{9}$$

where k is Boltzman constant. Value of the objective function f here represents energy E of a system. Energy of a system state is calculated from the initial, usually randomly picked, set of initial points and initial temperature. A new state reachable from the current state is randomly selected and its energy is evaluated. If energy is lower, the new state is always adopted. But if it higher it is accepted based on the probability of observing a fluctuation of size $exp(-\Delta E/kT)$, decided by comparison with a random number from interval [0,1] with uniform probability distribution. Successively, "temperature" of system is lowered and smaller energy fluctuations become more statistically significant. Theoretically, as $T \to 0$ system approaches the global minimum. However, in practice, search is stopped when a maximum number of iterations exceed a predefined limit. Restarting the procedure with a new set of initial points increases a chance for obtaining global minimum.

3.3. Differential evolution algorithm

The differential evolution (DE) method is a variant of the genetic algorithm (GE). It starts with a population of n (predetermined size of population) random vectors $\vec{x}_1$, $\vec{x}_2$, . $\vec{x}_n$ of real numbers from N dimension, called "genes". In every iteration, for each $\vec{x}_i$ integers a, b, and c are randomly chosen yielding construction of a corresponding mate

$$\vec{y}_i = \vec{x}_a + \gamma \cdot (\vec{x}_b - \vec{x}_c) \tag{10}$$

where γ is fixed and predetermined scaling factor. Then $\vec{x}_i$ is mated with $\vec{y}_i$ according to the given crossover probability. Gene exchange is performed by exchange of vector components. In addition, o point mutation randomly occurs at randomly selected component. Product of mutation is a child vector $\vec{z}_i$ which competes by fitness evaluation with its parent $\vec{x}_i$ for a place in a new population.

The method is very robust, but to increase a chance for convergence to a global minimum a restart of iterations from new initial populations is required.

4. Results and discussion

The model consists of 10 stiff ordinary differential equations with 132 adaptive parameters. The initial conditions are known from the experimental data [4]. A good initial estimate of the parameters is also known from the published data [2-4]. However, in order to test efficacy of the optimisation algorithms, initial set of parameters are randomly selected. An example of a typical iteration sequences is depicted in Fig. 3.

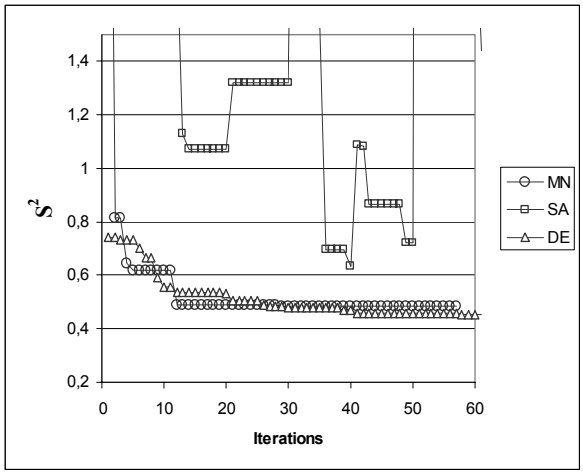

Figure 3. Minimization of the sum of squares S^2 by the algorithms: Nelder-Mead (MN), Simulated Annealing (SA), and Differential Evaluation.

Nelder-Mead and differential evolution converge relatively fast, and in approximately 60 iterations are sufficient for acceptable minimum. Acceptable level of the sum of squares depends on the level of random components in input (measured data). When NM is compared to DE, an indication of a better performance of NM is noticeable. Simulated annealing method does not reach the minimum at the same number of iterations. Similar results are obtained for repeated experiments with different sets of initial points. Tracking of the parameters during iterations is depicted in Fig. 4. For illustration purpose only the dominant parameter V_{fpts} in the model of initial intake of glucose by PTS system. The result shows that pa-

rameters do not converge to the same values.

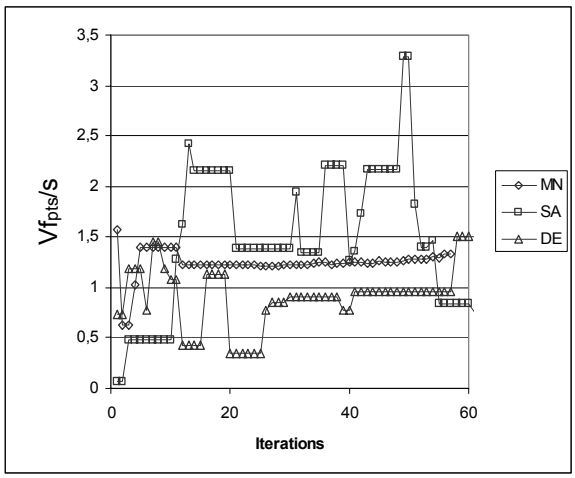

Figure 4. Estimation of the maximum PTS rate parameter by the algorithms: Nelder-Mead (MN), Simulated Annealing (SA), and Differential Evaluation (DE).

This clearly shows that the response surface of the model has several, possibly many equally acceptable minima. From statistical point of view, it can be inferred that the model parameters have large joint confidence intervals.

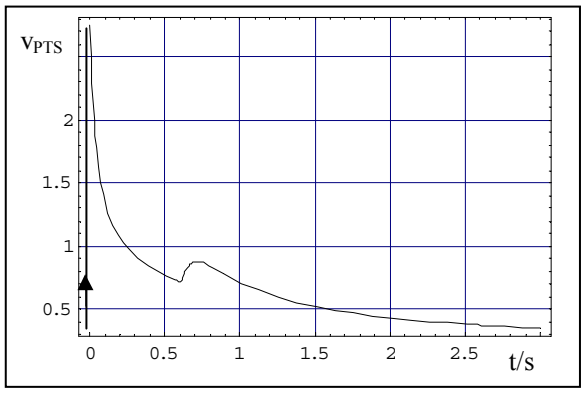

Figure 5. Simulated v_{PTS} (mmol/Ls) of PTS transferase rate upon glucose impulse.

Results of the model simulation are presented in Fig. 2. and Fig. 5. Response of the concentration of intracellular glucose-6-phosphat shows statistically appropriate unbiased interpolation with randomly distributed error. Most of the error dispersion can be interpreted as a result due to analytical, i.e pure error.

The main objective of the modelling is to reveal regulation of the intracellular process. The strong regulation effect of the phosphotransferase system can be observed in Fig. 5. Cells kept in environment with almost completely exhausted from source of carbon (glucose) immediately commence with high rate of assimilation at the very start of the impulse. By the model predictions, Fig. 5., maximum assimilation rate (2-5 mmol/Ls) occurs during 0,1 s, after which a strong feedback regulation mechanism drastically decreases the rate in an exponentially decaying manner, and in 15 s approaches to about 10 % of the initial maximum.

The model predicts further distribution of the metabolites due to the glucose impulse. The distribution of the metabolite flux is a highly regulated process, and the model enables flux analysis, gives mechanism of regulation, and possible determination of the "bottle necks" in the network.

Implications of the modelling and analysis are strain improvement by genetic engineering methods and improved process control of industrial biotechnology processes.

5. Conclusions

Analysis of dynamics of metabolic networks is a difficult task and challenge for computer modelling, simulation and optimisation. Here are investigated three numerical methods for kinetic parameter optimisation, namely Nedler-Mead, stochastic annealing, and differential evolution.

The methods are compared for their potential ability to provide global optima by adaptation and random search. The method of Nedler-Mead and differential evolution gave better efficacy compared to simulated annealing. However, convergence properties are case and initial point dependent, and need to be tested or "tuned" for each particular project. The attractive property of the methods is their independence on gradient evaluation, and their better performance for large systems compared to classical methods for parameter estimation, such as Marquard-Levenberg.

Analysis of the optimisation on the parameter space (variable) clearly shows the problem of overlapping effects of many parameters. Simulation results confirm the well known effect of interdependent relation between maximum rate and saturation (and inhibition) constants in Michaels-Menten type kinetic models. Inability to resolve individual kinetic effects can be elevated by use of models with simpler kinetic structure, such as "log type" kinetics.

However, here the main purpose of the modelling is not focused on precision of estimation of

kinetic parameters, but to use the model for evaluation of the flux rates through the network. Determination of distribution of fluxes provides clues to metabolic limitations in specific pathways and indicates possible flux control enzymes and cofactor dependencies. Computer modelling of metabolic networks and the flux analysis provides rational computer aided approach to genetic engineering method for improvement of strains and industrial biotechnological processes.

6. Acknowledgements

Communication and assistance of D. Degenring is gratefully acknowledged.

7. References

[1] Bailey, J. E. Toward a Science of Metabolic Engineering. *Science.* 252, (1991) 1668-1675.

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

[3] Degenring, D., Froemel, C., Dikta, G., Takors, R. Sensitivity analysis for the reduction of complex metabolism models. *J. Process Control* . 14, (2004) 729-745.

[4] Čerić, S. " Dynamic model of *Escherichia coli* central metabolism response upon glucose impulse", Diploma work, University of Zagreb, Faculty of Food Technology and Biotechnology, Zagreb (2005).

[5] Wolfram Research "*Mathematica*", v. 5.1 (2005).