CONSTRUCTION OF KINETIC MODELS FOR METABOLIC REACTION NETWORKS: LESSONS LEARNED IN ANALYZING SHORT TERM STIMULUS RESPONSE DATA

I.E. Nikerel, A.B. Canelas, S.J. Jol[†], P.J.T Verheijen, J.J. Heijnen,
Delft University of Technology, The Netherlands, [†]curr. addr.: ETH Zürich, Switzerland
Corresponding author: I.E. Nikerel, Department of Biotechnology, Delft University of Technology
Julianalaan 67, 2628 BC, Delft – The Netherlands, email: i.e.nikerel@tudelft.nl

Abstract. Construction of dynamic models of large scale metabolic networks is one of the central issues of engineering of living cells. However, construction of such models are often hampered by a number of challenges e.g. data availability, compartmentalization and parameter identification coupled to design of *in vivo* perturbations. As a solution to the latter, short-term perturbation experiments are proposed and are proven to be a useful experimental method to obtain insights on the *in vivo* kinetic properties of metabolic pathways.

The aim of this work is to construct a kinetic model using the available experimental data obtained via short term perturbation experiments, where the steady state of a glucose limited anaerobic chemostat culture of *Saccharomyces cerevisiae* was perturbed. In constructing the model, we first determined the steady state flux distribution using the data prior to the glucose pulse and the known stoichiometry. For the rate expressions, we used approximative linlog kinetics, which allows the enzyme-metabolite kinetic interactions to be represented by an elasticity matrix. We performed *a priori* model reduction based on time-scale analysis and parameter identifiability analysis allowing the information content of the experimental data to be assessed. The final values of the elasticities are estimated by fitting the model to the available short term kinetic response data.

The final model consists of 16 metabolites and 14 reactions. With 25 parameters, the model adequately describes the short term response of the cells to the glucose perturbation, pointing to the fact that the assumed kinetic interactions in the model are sufficient to account for the observed response.

1 Introduction

Thanks to the ongoing research efforts on the different "-ome levels", our understanding of cellular systems is increasing every day. Yet, due to the high complexity of the cell and the vast amount of interactions between metabolites, proteins and genes, there are several issues to be further explored. Construction of large scale dynamic models is one of these central issues in further concieving and engineering such systems.

Nevertheless, construction of models predicting cellular behavior is often hampered by a number of challenges ranging from data availability to compartmentalization and from parameter identification to design restrictions of *in vivo* perturbations. To begin with, while the list of quantitatively measured metabolites with high precision techniques is increasing, the list is far from being complete. Furthermore, when we deal with cells with multiple compartments (e.g. yeast), the metabolite measurements are only available as cell-averaged quantities. From the modeling side, while there have been numerous efforts to construct dynamic models for a number of model organisms (one recent example is the kinetic model for *Escherichia coli* presented in [1]), these models usually suffer from parameter unidentifiability for highly nonlinear, enzyme mechanistics based kinetic rate equations with a large number of parameters. Apart from the high number of nonlinear parameters, one other major issue in using mechanistics based kinetics is that the kinetic parameters estimated using data obtained from isolated conditions (*in vitro* conditions) do not represent the cellular *in vivo* behaviour [2].

These fundamental problems emphasize the design of perturbation experiments that would yield data reflecting *in vivo* conditions. Additionally, providing an experimental design that would isolate individual omic levels is important in disentangling the interactome riddle. Focusing on the metabolome level, short term stimulus response experiments are becoming more and more common practice and are proven to be a useful experimental method to obtain insights on the *in vivo* kinetic properties of metabolic pathways. This approach consists of perturbing a well-characterized culture state by external stimuli and subsequently monitoring the response of the intra- and extracellular metabolites over a short period of time ($<$200 s). Recent examples of such data are presented in [3] and [4]. During this time window, it may be assumed that enzyme concentrations do not change, allowing the observed responses to be attributed to kinetic interactions at the metabolome level alone.

Being overwhelmed by available experimental data obtained via short term perturbation experiments, this work aims at presenting a modeling framework that would facilitate the construction of (preferably large scale) kinetic models using such data, while providing solutions to the above mentioned challenges. We aimed to obtain a kinetic description of glycolysis in yeast under anerobic conditions using the data presented in [5].

2 Background

2.1 Construction of mathematical model for metabolic reaction networks

Constructing dynamic models for metabolic reaction networks begins with setting up mass balances for each metabolite, by considering the reaction stoichiometry (summarized in **S**) and assigning a kinetic expression (**v**) for each reaction (eq. 1)

$$\frac{d\mathbf{x}}{dt} = \mathbf{S}\cdot\mathbf{v} - \mu\mathbf{x} \qquad\qquad \mathbf{v} = \mathbf{v}(x,e,p) \tag{1}$$

Traditionally, the kinetics of each reaction is described by Michaelis-Menten type mechanistics kinetic expressions. The shortcomings of such approach being mentionned, approximative kinetics are often used as an alternative to overcome these problems (for a review see [6]). Among alternative formulations, linlog kinetics have been shown to be a useful format with its approximation quality for large changes in metabolite levels. In linlog kinetics, every entitiy is expressed with respect to reference state (eq. 2).

$$\frac{\mathbf{v}}{\mathbf{J^0}} = \frac{\mathbf{e}}{\mathbf{e^0}}\cdot\left(\mathbf{i} + \mathbf{E^x}\cdot\ln\left(\frac{\mathbf{x}}{\mathbf{x^0}}\right) + \mathbf{E^c}\cdot\ln\left(\frac{\mathbf{c}}{\mathbf{c^0}}\right)\right) \qquad\qquad E_{ij}^{x,c} \equiv \varepsilon_{x,c}^{v} \tag{2}$$

where the superscript ^{0} denotes the reference state, and the parameters are the elasticity coefficients for intracellular and extracellular metabolites (ε_x^0 and ε_c^0), defined within the context of metabolic control analysis (MCA) as: $\varepsilon_x^v = \frac{\partial v}{\partial x}\frac{x^0}{J^0}$. Several aspects of this formulation were discussed previously [7, 8, 9, 10, 11]

2.2 Time scale analysis and model reduction

Within a cell, numerous processes (metabolite conversion, enzyme synthesis, gene expression, etc…) occur simultaneously, following a time hierarchy. With respect to given experimental time window, these processes can be classified under different categories ranging from very fast (equilibrium) to very slow (frozen). Time scale analysis aims to classify fast and slow processes/pools in the cell. For such classification, we proposed in [8], to use turn-over time ($\tau = {}^{x_i^0}/_{J_{\text{in},i}^0}$) for each metabolite.

Time scale analysis allows model reduction that is achieved by selecting the fast metabolites in a reaction network (x_{fast}) that have small turnover times relative to the observation window, neglecting their accumulation term (${}^{dx_{\text{fast}}}/_{dt} \approx 0$) and transforming these dynamic metabolite mass balances to algebraic relations which results in a simpler system of differential-algebraic equations (eq. 3.)

$$\begin{aligned}\frac{d\mathbf{x}_{\text{slow}}}{dt} &= \mathbf{S}_{\text{slow}}\mathbf{J^0}\left(\mathbf{i} + \mathbf{E}^{\mathbf{x}}_{\text{slow}}\ln\left(\frac{\mathbf{x}_{\text{slow}}}{\mathbf{x}^{\mathbf{0}}_{\text{slow}}}\right) + \mathbf{E}^{\mathbf{x}}_{\text{fast}}\left(\frac{\mathbf{x}_{\text{fast}}}{\mathbf{x}^{\mathbf{0}}_{\text{fast}}}\right)\right) - \mu\mathbf{x}_{\text{slow}} \\ 0 &= \mathbf{S}_{\text{fast}}\mathbf{J^0}\left(\mathbf{i} + \mathbf{E}^{\mathbf{x}}_{\text{slow}}\ln\left(\frac{\mathbf{x}_{\text{slow}}}{\mathbf{x}^{\mathbf{0}}_{\text{slow}}}\right) + \mathbf{E}^{\mathbf{x}}_{\text{fast}}\left(\frac{\mathbf{x}_{\text{fast}}}{\mathbf{x}^{\mathbf{0}}_{\text{fast}}}\right)\right)\end{aligned} \tag{3}$$

The mass balance for the fast pools can be used to obtain an algebraic relation between fast and slow metabolite pools (eq. 4)

$$\ln\left(\frac{\mathbf{x}_{\text{fast}}}{\mathbf{x}^{\mathbf{0}}_{\text{fast}}}\right) = \mathbf{A}\cdot\ln\left(\frac{\mathbf{x}_{\text{slow}}}{\mathbf{x}^{\mathbf{0}}_{\text{slow}}}\right) \qquad\qquad A_{ij} = \alpha_k \equiv f_k\left(\varepsilon, \mathbf{S}, \mathbf{J^0}\right) \tag{4}$$

Using eqs. 3 and 4, we can lump the rates around the fast pools and obtain lumped kinetic expressions involving only the slow pools.

$$\frac{\mathbf{v}}{\mathbf{J^0}} = \frac{\mathbf{e}}{\mathbf{e^0}}\left(\mathbf{i} + \mathbf{D^x}\ln\left(\frac{\mathbf{x}_{\text{slow}}}{\mathbf{x}^{\mathbf{0}}_{\text{slow}}}\right)\right) \qquad\qquad D_{ij}^{x} \equiv \delta_k \equiv f_k(\varepsilon,\alpha) \tag{5}$$

2.3 *A priori* parameter identifiability analysis

One of the important challenges in dynamic modeling is to perform the parameter identifiability analysis, i.e. assess the information content of the data, given the model. Traditionally, this is performed by first constructing an initial parametrized model, then performing sensitivity analysis afterwards. Recently we have shown in [8] that the use of linlog kinetics allows an alternative method for *a priori* parameter identifiability analysis, allowing subsequently efficient parameter estimation.

The analysis is based on matrix **A** in eq. 4 and $\mathbf{D^x}$ in eq. 5. The entries of these matrices can be estimated directly from the available measurements. Since each entry of this matrices contain the lumped ε-parameters, it is staight forward to determine for which ε-parameter combination, a value can be estimated.

2.4 Estimation of elasticity parameters from dynamic data

While the elaborate details of estimation of elasticity parameters in eq. 2 from transient data has been reported elsewhere [7], we briefly summarize the important points of the procedure here. Considering the mass balance in eq.1, replacing linlog kinetics (eq. 5) for each reaction yields

$$\frac{d\mathbf{x}}{dt} = \mathbf{S}\cdot\mathbf{J^0}\cdot\frac{\mathbf{e}}{\mathbf{e^0}}\cdot\left(\mathbf{i}+\mathbf{D^x}\cdot\ln\left(\frac{\mathbf{x}}{\mathbf{x^0}}\right)\right)-\mu\cdot\mathbf{x}$$

In short term perturbation experiments, the enzyme activities can be assumed to remain constant, leading to $\mathbf{e}=\mathbf{e^0}$, allowing this term to be omitted. Further, at steady state $\mathbf{S}\cdot\mathbf{J^0}\cdot\mathbf{i}=\mathbf{0}$. Implementing these and integrating the mass balance yields:

$$\Delta\mathbf{x}+\mu\int_{t_i}^{t_{i+1}}\mathbf{x}dt=\mathbf{S}\cdot\mathbf{J^0}\cdot\mathbf{D^x}\int_{t_i}^{t_{i+1}}\ln\left(\frac{\mathbf{x}}{\mathbf{x^0}}\right)dt \tag{6}$$

An important feature of eq. 6 is that, while containing several terms, the parameters (D^x) appear in linear fashion, so that it can easily be re-written in a standard linear model form ($\mathbf{a}=\mathbf{Y}\cdot\mathbf{b}$ with $\mathbf{a}=\mathbf{a}(x,\mu)$, $\mathbf{Y}=\mathbf{Y}(S,J^0,x)$, $\mathbf{b}=\mathbf{b}(\varepsilon,\delta)$). This important feature allows the parameters to be estimated via linear regression.

3 Case study

As a case study for constructing the dynamic model, we considered the dynamic metabolome data obtained via short term glucose perturbation experiments to an anaerobic glucose limited chemostat fermentation of *Saccharomyces cerevisiae* [5]. The reference steady state metabolite levels and uptake and secretion rates are given in Table 1, and the response of the system to the glucose perturbation is presented in Figure 1.

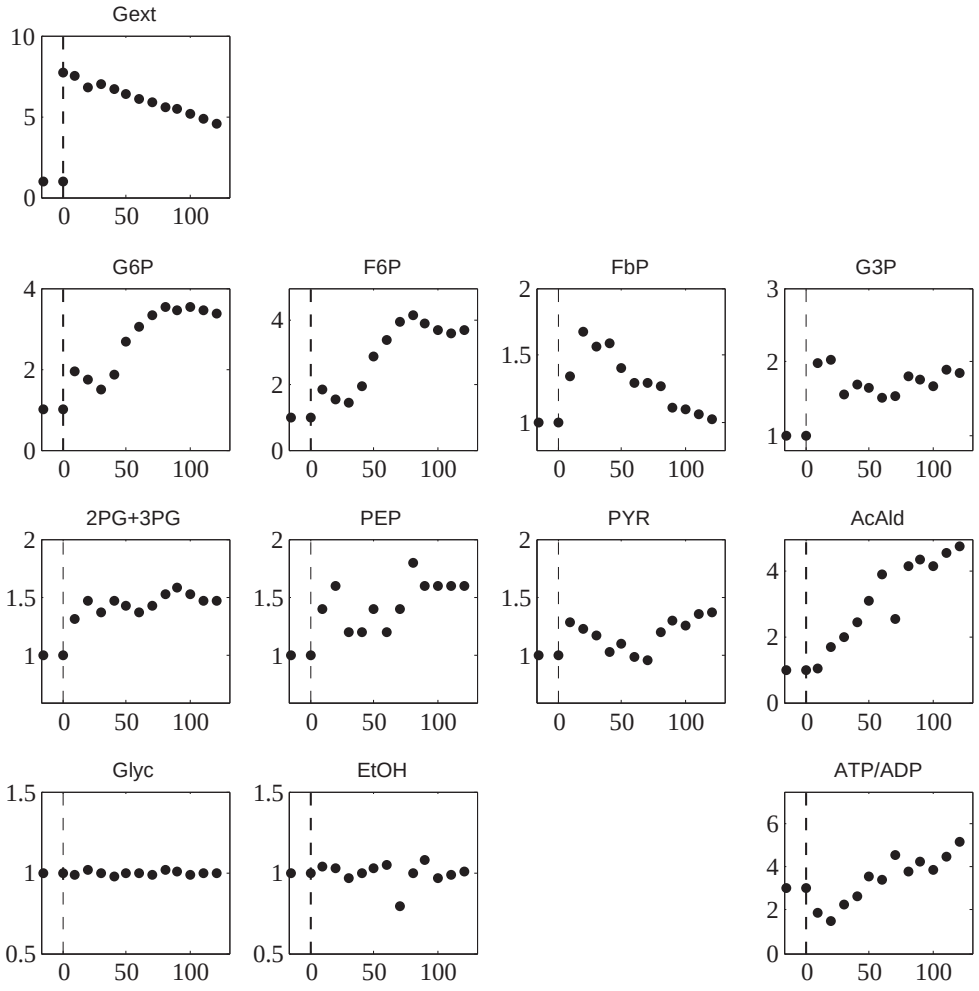


Figure 1: Available experimental data obtained from [5]. Metabolite levels are given in fold changes with respect to the reference state (presented in Table1)

Table 1: Reference conditions for the case study. Further details on experimental conditions can be found in [5]. Biomass (C_X) is in gDW/L, Dilution rate (D) is in h^{-1}, metabolites levels are in $\mu mol/gDW$ (intracellular) and in $mmol/L$ (extracellular). The q-rates and the intracellular fluxes are given in $mmol/gDW/hr$

X^0		J^0			
G6P	1.67	J_{IN}	2.44	J_{PGK}	4.62
F6P	0.25	J_{HK}	2.44	J_{PGM}	4.62
FdP	19.2	J_{PGI}	2.44	J_{GAPDH}	4.62
G3P	0.38	J_{PFK}	2.44	J_{ENO}	4.62
2PG+3PG	0.19	J_{ALD}	2.44	J_{PYK}	4.62
PEP	0.053	J_{G3PDH}	0.25	J_{PDC}	4.62
Pyruvate	0.79	J_{GPP}	0.25	J_{ADH}	4.62

C^0		D = 0.065, C_X= 6.3	
G^{ext}	0.62	$-q_s$	2.8
EtOH	52.68	q_{CO_2}	4.8
Glyc	4.31	q_{EtOH}	4.3
		q_{AcAld}	$0.86\ 10^{-3}$
		$q_{Glycerol}$	0.31

Table 2: The initial elasticity matrix for the skeleton model, considering only substrates and products, allosteric interactions are omitted

	G^{ext}	G^{in}	G6P	F6P	FdP	DHAP	GAP	G3P	Glyc	BPG	3PG	2PG	PEP	PYR	AcAld	EtOH	ATP	NAD
v_{IN}	ε_1	ε_2	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
v_{HK}	·	ε_3	ε_4	·	·	·	·	·	·	·	·	·	·	·	·	·	ε_5	·
v_{PGI}	·	·	ε_6	ε_7	·	·	·	·	·	·	·	·	·	·	·	·	·	·
v_{PFK}	·	·	·	ε_8	ε_9	·	·	·	·	·	·	·	·	·	·	·	ε_{10}	·
v_{ALD}	·	·	·	·	ε_{11}	ε_{12}	ε_{13}	·	·	·	·	·	·	·	·	·	·	·
v_{TPI}	·	·	·	·	·	ε_{14}	ε_{15}	·	·	·	·	·	·	·	·	·	·	·
v_{G3PDH}	·	·	·	·	·	ε_{16}	·	ε_{17}	·	·	·	·	·	·	·	·	·	ε_{18}
v_{GPP}	·	·	·	·	·	·	·	ε_{19}	ε_{20}	·	·	·	·	·	·	·	·	·
v_{GAPDH}	·	·	·	·	·	·	ε_{21}	·	·	ε_{22}	·	·	·	·	·	·	·	ε_{23}
v_{PGK}	·	·	·	·	·	·	·	·	·	ε_{24}	ε_{25}	·	·	·	·	·	ε_{26}	·
v_{PGM}	·	·	·	·	·	·	·	·	·	·	ε_{27}	ε_{28}	·	·	·	·	·	·
v_{ENO}	·	·	·	·	·	·	·	·	·	·	·	ε_{29}	ε_{30}	·	·	·	·	·
v_{PYK}	·	·	·	·	·	·	·	·	·	·	·	·	ε_{31}	ε_{32}	·	·	ε_{33}	·
v_{PDC}	·	·	·	·	·	·	·	·	·	·	·	·	·	ε_{34}	ε_{35}	·	·	·
v_{ADH}	·	·	·	·	·	·	·	·	·	·	·	·	·	·	ε_{36}	ε_{37}	·	ε_{38}

4 Results

4.1 Constructing the Skeleton Model

Considering the experimental context (glucose limited anaerobic fermentation), we use the network shown in Figure 2(a) depicting glycolysis with ethanol and glycerol formation from glucose to establish the mass balances. Branching towards storage carbohydrate synthesis pathway, pentose phosphate pathway (PPP), and TCA have been omitted, since under these conditions, it can be assumed that the production of ethanol through glycolysis is the major source of energy for *S. cerevisiae* and that the contribution of these branches to the overall carbon balance is small. Further, we assumed that the growth rate does not change during the glucose perturbation. Next, the intracellular fluxes at the reference state were calculated using the uptake and secretion rates and flux balance analysis (Table 1)

Setting the nonzero entries of the elasticity matrix Having the stoichiometric matrix obtained from Figure 2(a) and the reference fluxes from Table 1, the next step is to select the non-zero entries of the elasticity matrix based on known enzyme-metabolite kinetic interactions. At first, it was assumed that there are no allosteric interactions and that each reaction is affected by its substrates and products only. As such, the elasticity matrix connecting 15 rates and 18 metabolites, with 38 non-zero entries is presented in Table 2

Lumping unmeasured metabolites As mentioned, one of the common challenges in dynamic modeling of metabolic reaction networks is that, not every metabolite is measured. Comparing the network in Figure 2(a) and the available data in Figure 1, we note that for G^{in}, DHAP, GAP, BPG, NAD and NADH there are no measurements, and that for the 2PG and 3PG pools, only the sum can be measured. For these metabolites, we assume either that the corresponding reaction (e.g. v_{PGM} for 2PG and 3PG) is at pseudo equilibrium or that the respective pool (e.g. G^{in}) is at pseudo steady state so that we lump the rates around these pools (v_{IN} and v_{HK}).

Additionally, while NAD and NADH are not measured they are kept in the network since these can be calculated

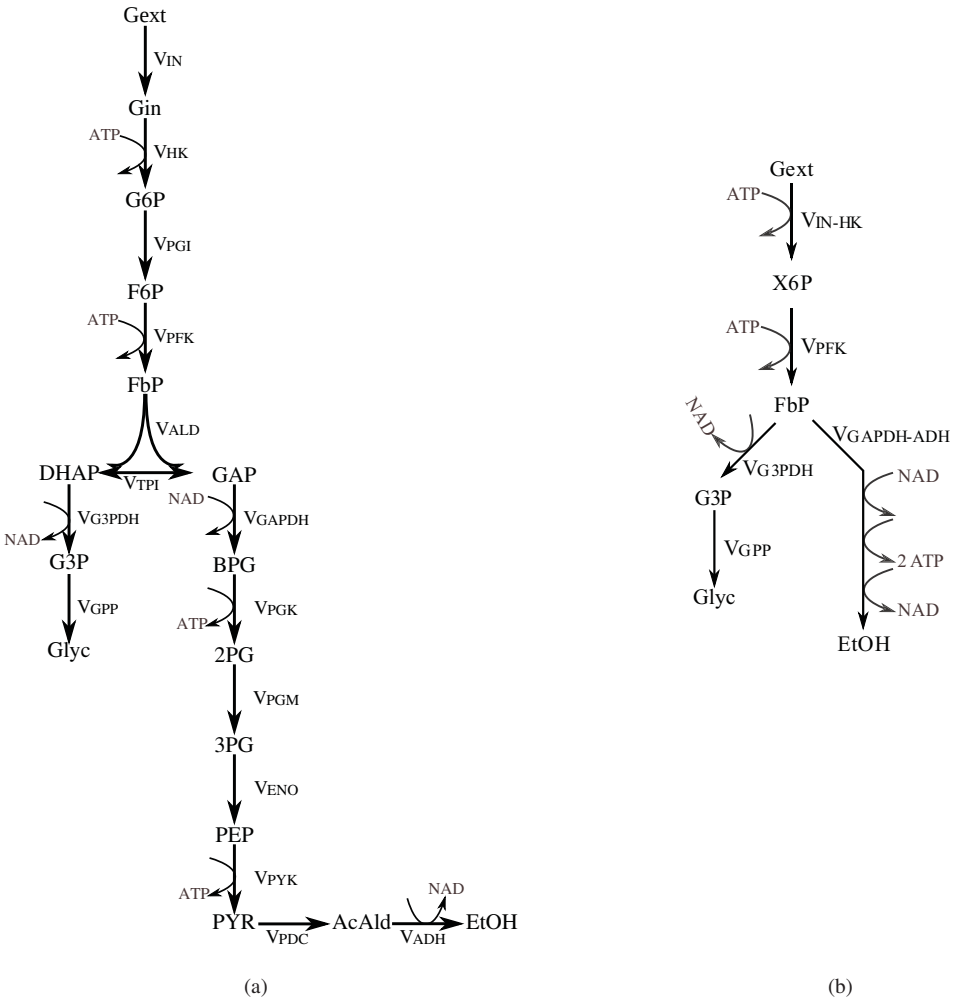


Figure 2: Metabolic network describing glycolysis in *S. cerevisiae* with ethanol and glycerol formation. **a**: The initial full network. **b**: Final reduced network after time scale analysis based model reduction

using equilibrium assumptions. We consider the following net equilibrium from AcAld to EtOH:

$$\mathrm{AcAld} + \mathrm{NADH} \rightleftarrows \mathrm{EtOH} + \mathrm{NAD}$$

Then, the ratio of NAD/NADH follows as

$$\ln\left(\frac{\mathrm{NAD}/\mathrm{NADH}}{\mathrm{NAD}^0/\mathrm{NADH}^0}\right) = \ln\left(\frac{\mathrm{AcAld}}{\mathrm{AcAld}^0}\right) - \ln\left(\frac{\mathrm{EtOH}}{\mathrm{EtOH}^0}\right)$$

Furthermore, we know that during the anaerobic production of ethanol, the energy metabolite ATP is exclusively produced in glycolysis, while it is consumed by many reactions and processes outside our model boundaries. For this, ATP and ADP are only considered as input of the metabolic network, rather than being predicted by the model. Similarly, the calculated redox metabolites NAD and NADH are also considered as input of the metabolic network. In what follows, we consider the ratio $^{\mathrm{ATP}}/_{\mathrm{ADP}}$ and $^{\mathrm{NAD}}/_{\mathrm{NADH}}$ for further analysis, yet note those ratios as ATP and NAD respectively for the sake of briefness.

From the concentration profiles in Figure 1, the ethanol and glycerol profiles show, although noisy, a flat response. This is to be expected as steady state metabolite levels of both ethanol and glycerol are already very high, so that the effect of the additional perturbation can not be observed on this experimental time window therefore, both the elasticity of ethanol on v_{ADH} and the elasticity of glycerol on v_{GPP} cannot be estimated and were removed from the elasticity matrix.

4.2 Time scale analysis and model reduction

Reactions at pseudo equilibrium For these reactions, equations similar to eq. 7 are constructed, the new lumped pools are defined in such a way to eliminate the reaction at equilibrium from the mass balance. For the metabolites G6P and F6P, it is known that the conversion is a pseudo equilibrium rate, hence the dynamics of F6P is

Table 3: The turnover times of the metabolites in the network of figure 2(a), calculated by dividing the metabolite level at steady state by the total incoming flux to the corresponding pool. For FdP′, the turnover time is actually higher since the pool contains GAP and DHAP, but since there is no measurement on those, only the level of FdP is taken into account

Metabolite	τ (s)	Metabolite	τ (s)
X6P	2.83	2PG+3PG	0.14
FdP′	28.3	PEP	0.041
G3P	5.31	Pyruvate	0.61

algebraically linked to G6P:

$$\ln\left(\frac{\text{F6P}}{\text{F6P}^0}\right) = \ln\left(\frac{\text{G6P}}{\text{G6P}^0}\right) \tag{7}$$

Since, in this case, we have the measurements for both metabolites the adequacy of such assumption can be checked by monitoring the mass action ratio during the perturbation experiment (Figure 3(a)). Lastly, the pool X6P is defined in such a way to eliminate v_{PGI} from the mass balance (eq. 8).

$$\text{X6P} = \text{F6P} + \text{G6P} \qquad \begin{array}{rcl} {}^{d\text{G6P}}/_{dt} & = & v_{\text{HK}} - v_{\text{PGI}} \\ {}^{d\text{F6P}}/_{dt} & = & v_{\text{PGI}} - v_{\text{PFK}} \end{array} \qquad \frac{d\text{X6P}}{dt} = v_{\text{HK}} - v_{\text{PFK}} \tag{8}$$

The model reduction is achieved by defining the equilibrium pool (X6P) for each reaction assumed to be at pseudo equilibrium (v_{PGI}) and then removing the pseudo equilibrium reaction and the individual pools (G6P and F6P) from the mass balance. In our case study, in addition to v_{PGI}, v_{ADH} and v_{PGM}, the same procedure is also applied to v_{ALD} and v_{TPI} and the equilibrium pool FdP′ = FdP + GAP + DHAP is defined.

Metabolites at pseudo steady state To inspect which metabolites are fast compared to the rest, time scale analysis was performed by calculating the turn over time for each of the remaining metabolite pools. The resulting turnover times are shown in Table 3.

From this table, it can be concluded that PEP, 2PG+3PG and PYR have the smallest turnover times, hence in addition to G^{in} and BPG, these pools can be assumed to be at pseudo steady state. This implies that the mass balances for these metabolites degenerate into algebraic relations. Considering e.g. G^{in}, the dynamic mass balance becomes:

$$v_{\text{IN}} \approx v_{\text{HK}}$$

substituting the elasticities and rearranging yields

$$\ln\left(\frac{\text{G}^{\text{in}}}{\text{G}^{\text{in}0}}\right) = \alpha_1 \ln\left(\frac{\text{G}^{\text{ext}}}{\text{G}^{\text{ext}0}}\right) + \alpha_2 \ln\left(\frac{\text{G6P}}{\text{G6P}^0}\right) + \alpha_3 \ln\left(\frac{\text{ATP}}{\text{ATP}^0}\right)$$
$$\alpha_1 = -\frac{\varepsilon_1}{\varepsilon_2 - \varepsilon_3} \qquad \alpha_2 = \frac{\varepsilon_4}{\varepsilon_2 - \varepsilon_3} \qquad \alpha_3 = \frac{\varepsilon_5}{\varepsilon_2 - \varepsilon_3}$$

If the measurements would be available for each metabolite in the above equation, we could estimate the α-parameters and check the adequacy of the linear model. The model reduction is achieved by lumping the rates around G^{in}, removing G^{in} from the mass balance and defining the rate expression for the lumped reaction rate by substituting the degenerated algebraic relation into one of the rates for G^{in} (eq. 9)

$$v_{\text{IN-HK}} = J^0_{\text{IN-HK}} \left(1 + \delta_1 \ln\left(\frac{\text{G}^{\text{ext}}}{\text{G}^{\text{ext}0}}\right) + \delta_2 \ln\left(\frac{\text{G6P}}{\text{G6P}^0}\right) + \delta_3 \ln\left(\frac{\text{ATP}}{\text{ATP}^0}\right)\right) \tag{9}$$
$$\delta_1 = \varepsilon_1 + \varepsilon_2\alpha_1 \qquad \delta_2 = \varepsilon_2\alpha_2 \qquad \delta_3 = \varepsilon_3\alpha_3$$

Following the same procedure, for the fast metabolites, BPG, 2PG+3PG, PEP and PYR, the algebraic relations containing α-parameters are constructed.

$$\begin{bmatrix} \ln\left(\frac{\text{BPG}}{\text{BPG}^0}\right) \\ \ln\left(\frac{\text{XPG}}{\text{XPG}^0}\right) \\ \ln\left(\frac{\text{PEP}}{\text{PEP}^0}\right) \\ \ln\left(\frac{\text{PYR}}{\text{PYR}^0}\right) \end{bmatrix} = \begin{bmatrix} \alpha_4 & \alpha_5 & \alpha_6 \\ \alpha_7 & \alpha_8 & \alpha_9 \\ \alpha_{10} & \alpha_{11} & \alpha_{12} \\ \alpha_{13} & \alpha_{14} & \alpha_{15} \end{bmatrix} \cdot \begin{bmatrix} \ln\left(\frac{\text{FdP}}{\text{FdP}^0}\right) \\ \ln\left(\frac{\text{NAD}}{\text{NAD}^0}\right) \\ \ln\left(\frac{\text{ATP}}{\text{ATP}^0}\right) \end{bmatrix}$$

$$\alpha_4 = \frac{2\varepsilon_{21}\left(\varepsilon_{25}\,\varepsilon_{34}\left(-\varepsilon_{30}+\varepsilon_{31}\right)+\varepsilon_{25}\,\varepsilon_{30}\,\varepsilon_{32}-\varepsilon_{29}\,\varepsilon_{31}\,\varepsilon_{34}\right)}{\Xi}$$

$$\alpha_5 = \frac{\varepsilon_{23}\left(\varepsilon_{25}\varepsilon_{34}\left(-\varepsilon_{30}+\varepsilon_{31}\right)+\varepsilon_{25}\,\varepsilon_{30}\,\varepsilon_{32}-\varepsilon_{29}\,\varepsilon_{31}\,\varepsilon_{34}\right)}{\Xi} \qquad \alpha_6 = \frac{\varepsilon_{34}\left(\varepsilon_{25}\,\varepsilon_{30}\,\varepsilon_{33}+\varepsilon_{26}\,\varepsilon_{29}\,\varepsilon_{31}\right)}{\Xi}$$

$$\alpha_7 = \frac{2\,\varepsilon_{21}\,\varepsilon_{24}\left(+\varepsilon_{31}\,\varepsilon_{34}+\varepsilon_{30}\,\varepsilon_{32}-\varepsilon_{30}\,\varepsilon_{34}\right)}{\Xi} \qquad \alpha_8 = \frac{\varepsilon_{23}\,\varepsilon_{24}\left(+\varepsilon_{31}\,\varepsilon_{34}\,+\varepsilon_{30}\,\varepsilon_{32}-\varepsilon_{30}\,\varepsilon_{34}\right)}{\Xi}$$

$$\alpha_9 = \frac{+\varepsilon_{24}\,\varepsilon_{30}\,\varepsilon_{33}\,\varepsilon_{34}-\varepsilon_{22}\,\varepsilon_{30}\,\varepsilon_{33}\,\varepsilon_{34}+\varepsilon_{22}\,\varepsilon_{26}\,\varepsilon_{30}\,\varepsilon_{34}-\varepsilon_{22}\,\varepsilon_{26}\,\varepsilon_{31}\,\varepsilon_{34}-\varepsilon_{22}\,\varepsilon_{26}\,\varepsilon_{30}\,\varepsilon_{32}}{\Xi}$$

$$\alpha_{10} = \frac{2\,\varepsilon_{21}\,\varepsilon_{24}\,\varepsilon_{29}\left(+\varepsilon_{32}-\varepsilon_{34}\right)}{\Xi} \qquad \alpha_{11} = \frac{\varepsilon_{23}\,\varepsilon_{24}\,\varepsilon_{29}\,\left(+\varepsilon_{32}-\varepsilon_{34}\right)}{\Xi}$$

$$\alpha_{12} = \frac{-\varepsilon_{22}\,\varepsilon_{26}\,\varepsilon_{29}\,\varepsilon_{32}+\varepsilon_{22}\,\varepsilon_{26}\,\varepsilon_{29}\,\varepsilon_{34}+\varepsilon_{22}\,\varepsilon_{25}\,\varepsilon_{33}\,\varepsilon_{34}+\varepsilon_{24}\,\varepsilon_{29}\,\varepsilon_{33}\,\varepsilon_{34}-\varepsilon_{22}\,\varepsilon_{33}\,\varepsilon_{29}\,\varepsilon_{34}}{\Xi}$$

$$\alpha_{13} = \frac{2\,\varepsilon_{29}\,\varepsilon_{24}\,\varepsilon_{31}\,\varepsilon_{21}}{\Xi} \qquad \alpha_{14} = \frac{\varepsilon_{29}\,\varepsilon_{24}\,\varepsilon_{31}\,\varepsilon_{23}}{\Xi} \qquad \alpha_{15} = \frac{-\varepsilon_{22}\,\left(\varepsilon_{25}\,\varepsilon_{30}\,\varepsilon_{33}+\varepsilon_{26}\,\varepsilon_{29}\,\varepsilon_{31}\right)}{\Xi}$$

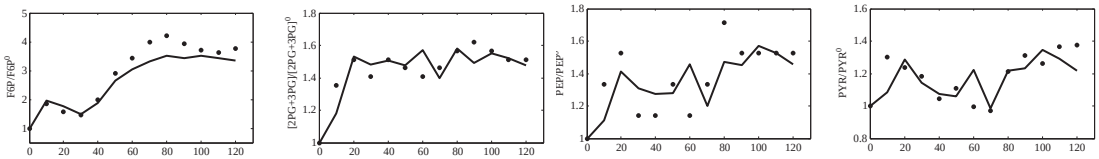
where

$$\Xi = -\varepsilon_{22}\,\varepsilon_{25}\,\varepsilon_{30}\,\varepsilon_{34}+\varepsilon_{22}\,\varepsilon_{25}\,\varepsilon_{31}\,\varepsilon_{34}+\varepsilon_{22}\,\varepsilon_{25}\,\varepsilon_{30}\,\varepsilon_{32}+\varepsilon_{24}\,\varepsilon_{29}\,\varepsilon_{31}\,\varepsilon_{34}-\varepsilon_{22}\,\varepsilon_{29}\,\varepsilon_{31}\,\varepsilon_{34}$$

Having experimental data, a subpart of the entries of matrix **A** can be estimated. This results in

$$\mathbf{A} = \begin{bmatrix} \alpha_4 & \alpha_5 & \alpha_6 \\ 0.1808 & 0.3281 & -0.2398 \\ -0.4187 & 0.4167 & -0.4943 \\ -0.6825 & 0.3448 & -0.6033 \end{bmatrix}$$

The prediction of the fast metabolites from the slow FdP, ATP and NAD are presented in figure 3.



(a) (b) (c) (d)

Figure 3: Considering time-scale analysis for model reduction. **a**: the pools G6P and F6P are in equilibrium, **b–d** the pools 2PG+3PG, PEP and PYR are pseudo steady state pools, calculated from ATP, FbP and NAD. In each case, dots represent the available experimental data, the line represents the model prediction

At last, the model is reduced by lumping the rates around the removed metabolites. In the final model, $v_{\mathrm{GAPDH-ADH}}$ is the lumped reaction of the original v_{GAPDH}, v_{ENO}, v_{PYK} and v_{ADH} reactions.

4.3 *A priori* parameter identifiability analysis

Having reduced the model based on the time scale analysis, *a priori* parameter identifiability analysis can be performed. This is performed by analyzing the the matrices **A** and $\mathbf{D}^x$. Note that the elasticities of the lumped rates are a combination of the original elasticities for the unreduced model (table 2). Taking v_{PYK} rate expression as example and replacing PEP and PYR lead to the lumped expression for $v_{\mathrm{GAPDH-ADH}}$

$$v_{\mathrm{GAPDH-ADH}} = J^0_{\mathrm{GAPDH-ADH}} \cdot \left(1+\delta_4 \ln\left(\frac{\mathrm{FdP}}{\mathrm{FdP}^0}\right)+\delta_5 \ln\left(\frac{\mathrm{NAD}}{\mathrm{NAD}^0}\right)+\delta_6 \ln\left(\frac{\mathrm{ATP}}{\mathrm{ATP}^0}\right)\right)$$

and δ_1, δ_2 and δ_3 are defined as a function of the ε and α parameters

$$\delta_4 = \varepsilon_{31}\,\alpha_{10}+\varepsilon_{32}\,\alpha_{13} \qquad \delta_5 = \varepsilon_{31}\,\alpha_{11}+\varepsilon_{32}\,\alpha_{14} \qquad \delta_6 = \varepsilon_{31}\,\alpha_{12}+\varepsilon_{32}\,\alpha_{15}+\varepsilon_{33}$$

The nonlinear parameter estimation will lead to δ-parameters, and together with the estimated α-parameters, ε_{31}, ε_{32} and ε_{33} can be calculated.

Table 4: Elasticity matrix for the reduced network in figure 2(b)

	G^{ext}	X6P	FdP	G3P	Glyc	EtOH	ATP	NAD
v_{IN-HK}	δ_1	δ_2	·	·	·	·	δ_3	·
v_{PFK}	·	ε_8	ε_9	·	·	·	ε_{10}	·
v_{G3PDH}	·	·	ε_{16}	ε_{17}	·	·	·	ε_{18}
v_{GPP}	·	·	·	ε_{19}	·	·	·	
$v_{GAPDH-ADH}$	·	·	δ_4	·	·	·	δ_5	δ_6

4.4 Estimation of elasticities, calculation of control coefficients

For the metabolic network shown in figure 2(b) we considered the elasticity matrix presented in Table 4, generated an initial estimate by solving eq. 6 and estimated the final values via a nonlinear optimization routine, minimizing the squared error between the simulated and experimental values. Figure 4 shows the simulation with the elasticities obtained from the nonlinear optimization procedure and the original experimental data. The elasticities are shown in table 5. It can be seen that most concentration profiles are estimated reasonably well.

Having estimated the elasticities, we also calculated the concentration, flux and response coefficients for our system using estimated elasticities and the MCA theorems as:

$$\mathbf{C^x} = -\left(\mathbf{S J^0 E^x}\right)^{-1} \mathbf{S J^0} \qquad \mathbf{C^J} = \mathbf{I} + \mathbf{E^x C^x} \qquad \mathbf{R^J} = \mathbf{C^J E^c}$$

The calculated control and response parameters for the ethanol flux are given in Table 5. From this table it can be seen that the ethanol production is mostly controlled by the uptake rate and the available ATP.

Table 5: Final estimates of the elasticities, the calculated control and response coefficient for the ethanol flux

	G^{ext}	X6P	FdP	G3P	EtOH	Glycerol	ATP	NAD
v_{IN-HK}	1.37	0.13	·	·	·	·	0.34	·
v_{PFK}	·	2.055	2.785	·	·	·	1.0049	·
v_{G3PDH}	·	·	8.101	-0.177	·	·	·	1.3323
v_{GPP}	·	·	·	5.32	·	·	·	·
$v_{GAPDH-ADH}$	·	·	10.935	·	·	·	2.612	0.293

	$\mathbf{C^J}_{GAPDH-ADH}$
e_{IN-HK}	1.06
e_{PFK}	-0.07
e_{G3PDH}	-0.05
e_{GPP}	0.01
$e_{GAPDH-ADH}$	0.05

	$\mathbf{R^J}_{GAPDH-ADH}$
G^{ext}	1.46
ATP	0.57
NAD	-0.05

For parameter identifiability, using the estimated values of the α and δ parameters, the individual elasticities for $v_{GAPDH-ADH}$ are found to be: $\varepsilon_{31} = 39.64$, $\varepsilon_{32} = -40.34$, $\varepsilon_{33} = -4.45$. Here, it should be noted that among 18 kinetic parameters ($\varepsilon_{21} \rightarrow \varepsilon_{38}$), only 12 of them (6 α's, 3ε's) can be identified.

5 Discussion and Conclusions

In this work, we presented our model biulding scheme for construction of dynamic models for metabolic reaction networks. As a case study, we developed a kinetic description for glycolysis in yeast cells grown under anaerobic carbon limited conditions. The constructed model was used as data scaffold and served, among many others, to assess the information content of the available highly dynamic metabolome data.

In developing the model, from the beginning, i.e. defining model boundaries, until the end, e.g. parametrization of the model, there are several points considered and adjusted to concur with the case study, e.g. exclusion of PPP, TCA and storage pathways considering cultivation conditions or time scale based model reduction considering the turn-over times. Overall, the constructed model describes adequately the response of the intracellular metabolites to the given glucose perturbation.

While the available dataset is highly dynamic, by implementing time scale analysis based, *a priori* model reduction, we concluded that, the dynamics of a number metabolites, i.e. fast metabolites, are actually not independent but rather their dynamics are dictated by the slow metabolites. The assumption of the insignificance of the differential term for the fast metabolites can be checked by inspecting the adequacy of the algebraic relation (eq. 4). This is presented in figure 3, and we conclude that the model reduction is justified.

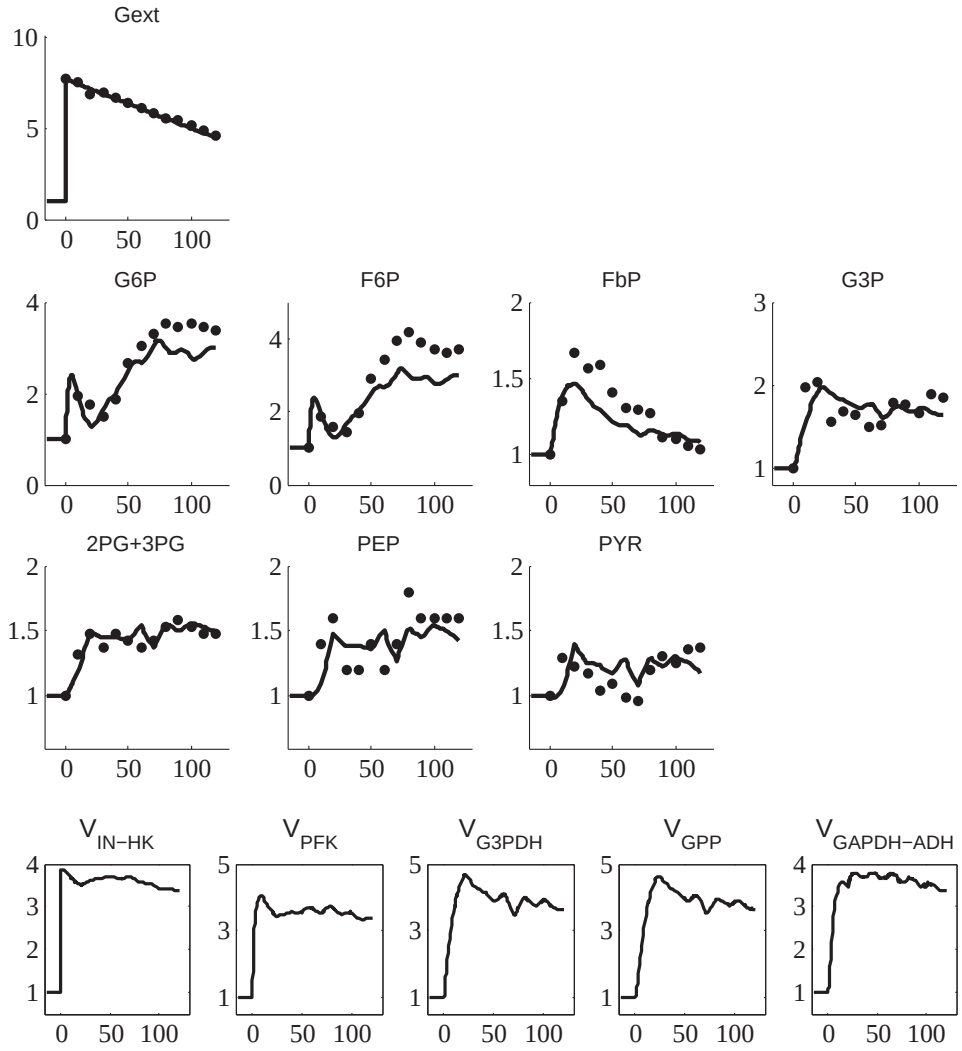


Figure 4: Simulation of the glucose perturbation with the estimated parameters. The metabolites are presented in fold changes.

In the proposed modeling framework, the use of linlog kinetics is central both in *a priori* model reduction and parameterization of the final model. Linlog kinetics is linear in parameters while adequately describing system dynamics. The linearity property is useful especially when carrying parameter identifiability analysis, since the parameter estimation problem (eq. 6) can be translated into a linear model ($\mathbf{a} = \mathbf{Y} \cdot \mathbf{b}$) and the design matrix $\mathbf{Y}$ can be further analyzed using the standard tools of linear algebra as described in [8]. Furthermore, the use of linlog kinetics allows directly interpreting the estimated parameters since these are theoretically defined as the elasticities in MCA. Having the local elasticities, calculation of global properties as control coefficients is straight forward.

The MCA analysis shows that most of the control is on the upper part of glycolysis, in the uptake reaction. This is understandable since the model reduction shows that the lower part of the glycolysis follows the dynamics of the slow metabolites.

Having constructed an initial model, the next step would be to perform model based experimental design to obtain further data on parts of the network where the available information was not enough to estimate a subpart of the parameters. Furthermore, in the subsequent rounds of model building, additional interactions, e.g. allosteric effects, can be incorporated into the model. As an example, a well-known allosteric effect of FbP to v_{PYK} is excluded from the current model in order to prevent over parameterization. This is justified since if $\varepsilon_{FbP}^{v_{PYK}}$ would be introduced into the model, then it would only appear in the definition of δ_4 and the allosteric effect would not be separately observable. Yet this interaction can be incorporated in a later stage of the model provided with additional information.

6 References

[1] Chassagnole, C. and Noisommit-Rizzi, N. and Schmid, J.W. and Mauch, K. and Reuss, M., *Dynamic modeling of the central carbon metabolism of Escherichia coli*, Biotechnology and Bioengineering, 2002, 79(1), 53–73

[2] Teusink, B. and Passarge,Jutta and Reijenga,Corinne A. and Esgalhado,Eugenia and van der Weijden,Coen C. and Schepper,Mike and Walsh,Michael C. and Bakker,Barbara M. and van Dam,Karel and Westerhoff,Hans V. and Snoep,Jacky L., *Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry*, European Journal of Biochemistry, 2000, 267(17), 5313-5329,

[3] Canelas, AB and van Gulik, WM and Heijnen, JJ, *Determination of the cytosolic free NAD/NADH ratio in Saccharomyces cerevisiae under steady-state and highly dynamic conditions.*, Biotechnology and bioengineering, Biotechnology and Bioengineering, 2008, 100(4):734-43.

[4] Kresnowati, M.T.A.P and Suarez-Mendez, CM and van Winden, WA and van Gulik, WM and Heijnen, JJ, *Quantitative physiological study of the fast dynamics in the intracellular pH of Saccharomyces cerevisiae in response to glucose and ethanol pulses*, Metabolic Engineering, 2008, 10(1), 39–54

[5] Üçışık, M. H., *Transient Energetic Metabolic Response of Anaerobic Saccharomyces cerevisiae Culture to Benzoic Acid*, Delft University of Technology, 2006, Master's thesis

[6] Heijnen, J.J., *Approximative kinetic formats used in metabolic network modeling*, Biotechnology and Bioengineering, 2005, 91(5),534-545,

[7] Nikerel, I.E. and van Winden, W.A. and van Gulik, W.M. and Heijnen, J.J., *A method for estimation of elasticities in metabolic networks using steady state and dynamic metabolomics data and linlog kinetics*, BMC Bioinformatics, 2006, 7 : 540

[8] Nikerel, I.E and van Winden, W.A and Verheijen, P.J.T and Heijnen,J.J., *Model reduction and* a priori *kinetic parameter identifiability analysis using metabolome time series for metabolic reaction networks with linlog kinetics*, Metabolic Engineering, 2009, 11, 20-30

[9] Visser, D. and Schmid, J.W. and Mauch, K. and Reuss, M. and Heijnen, J.J., *Optimal re-design of primary metabolism in Escherichia coli using linlog kinetics*, Metabolic Engineering, 2004, 6(4), 378–390,

[10] Visser, D. and Heijnen, J.J., *Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics*, Metabolic Engineering, 2003, 5(3), 164–176,

[11] Wu,L. and Wang,W. and van Winden,W.A. and van Gulik,W.M. and Heijnen,J.J., *A new framework for the estimation of control parameters in metabolic pathways using lin-log kinetics*, European Journal of Biochemistry, 2004, 271(16), 3348-3359,