CONTROLO 2008
8th Portuguese Conference on Automatic Control
University of Trás-os-Montes and Alto Douro, Vila Real, Portugal
July 21-23, 2008

STRUCTURAL ANALYSIS OF METABOLIC NETWORKS: A CASE STUDY ON *LACTOCOCCUS LACTIS*

Susana Vinga [*,**,1,2] **Karl Thomaseth** [***] **João M. Lemos** [*,†]
Ana Rute Neves [****] **Helena Santos** [****] **Ana T. Freitas** [*,†]

[*] *INESC-ID - R Alves Redol 9, 1000-029 Lisboa, Portugal*
[**] *FCM-UNL - C Mártires Pátria 130, 1169-056 Lisboa, Portugal*
[***] *ISIB-CNR - Corso Stati Uniti 4, 35127 Padova, Italy*
[****] *ITQB-UNL - R Qta Grande 6, 2780-156 Oeiras, Portugal*
[†] *IST-UTL - Av Rovisco Pais, 1049-001 Lisboa, Portugal*

Abstract: The dynamic modelling of metabolic networks constitutes a major challenge in systems biology. The time evolution of metabolite concentration in living cells is usually modelled by complex systems of non-linear differential equations with a large number of parameters. The identification of the model structure and the estimation of its parameters from experimental data is a difficult task, for which there is currently no automatic and straightforward solution. This paper shows that a prior structural analysis of the model that addresses parameter sensitivity and identifiability issues can significantly improve the subsequent reverse engineering step. The application of established model building and analysis procedures can thus have a positive impact in the development of complex biological systems models.

Keywords: nonlinear systems; parameter estimation; identifiability; sensitivity analysis.

1. INTRODUCTION

1.1 *Motivation*

The dynamical description and modelling of metabolic networks constitutes a major challenge in systems biology. The development of experimental techniques for data acquisition is producing at growing rate high quality biochemistry information. From the modeller point of view, new tools are continuously being proposed to estimate more efficiently and accurately dynamic models that can be useful to design new informative experiments and to propose and validate new hypotheses. Given the abundance of information, a major modelling difficulty has become the formulation of complex, structured, biologically meaningful models for predicting multidimensional observations of biological system variables. Although the increase in model complexity requires new tools, longstanding model building methodologies for structural formulation, parameter estimation and validation (Carson *et al.*, 1983), are still applicable and can effectively support the identification of complex biological systems.

1.2 *The problem*

The formulation of structured models of biological systems is greatly influenced by the wealth of detailed information available on single biological processes and on their interactions. A feasible and often preferred model building strategy is therefore the top-down approach that starts with gathering prior information on the biological system structure, which is then translated into mathematical model equations to yield a system of non-linear differential equations. These latter become useful only if adequate values can be assigned to a large number of model param-

[1] E-mail: svinga@kdbio.inesc-id.pt
[2] Part of this work has been done under project DynaMo (PTDC/EEA-ACR/69530/2006) from FCT, Portugal.

eters that are usually either unknown or have been determined in controlled isolated conditions. To infer biologically meaningful parameters the model must be fitted to multidimensional experimental data profiles, which is typically complicated by practical unidentifiability due to overparameterization. Due to model complexity it is usually difficult or even unfeasible to carry out a formal *a priori* identifiability analysis (Audoly *et al.*, 2001; Bellu *et al.*, 2007). The problem becomes then to reduce the model order, either by assigning fixed values to a subset of the parameters or, preferably, by simplifying the subjacent differential equations. To retain the biological meaning of the model the simplification must be carried out heuristically, but can be guided using formal analysis tools as shown in this work.

1.3 *State-of-the-art review*

1.3.1. *Biochemistry* One interesting system that will be used as a case study is the glycolytic pathway. Glycolysis consists in the degradation of one molecule of glucose onto two molecules of pyruvate with the concomitant production of adenosine triphosphate (ATP). The topology of this metabolic network has been known for a long time, including the activation and inhibitory signals present, which can be represented as extra edges in the graph. The dynamics of this system is however still not completely understood. With Nuclear Magnetic Resonance (NMR) it is nowadays possible to obtain *in vivo* multivariate time series of metabolite concentrations for several organisms and different pathways and perturbations, allowing thus the real-time observation of these processes. This type of information, in particular the glycolytic pathway in *Lactococcus lactis*, will be the basis for this modelling study (Neves *et al.*, 2002).

1.3.2. *Models for metabolic networks* The dynamic evolution of metabolite concentrations is usually modelled with systems of non-linear differential equations. In particular, a methodology for establishing the format of these equations was proposed, in (Savageau, 1969), under Biochemical Systems Theory (BST), and proved to be very useful. In this approach each flux is approximated by a power law, which corresponds to a Taylor series expansion in logarithmic space. Therefore, the flux rate $\dot{X}_i$ of each metabolite concentration X_i is given as a function of the concentrations of all metabolites $X_j, j = 1, ..., n$. Under BST, the fluxes can be expressed in two different ways, either using S-systems or the Generalized Mass Action (GMA) formalism, written respectively as:

$$\begin{aligned} \dot{X}_i &= \alpha_i \Pi_{j=1}^{n} X_j^{g_{ij}} - \beta_i \Pi_{j=1}^{n} X_j^{h_{ij}} \\ \dot{X}_i &= \Sigma_{k=1}^{m} \pm \gamma_{ik} \Pi_{j=1}^{n} X_j^{f_{ijk}} \end{aligned} \quad (1)$$

In S-systems the production and degradation fluxes are grouped into two factors with parameters (α_i, g_{ij}) and (β_i, h_{ij}) respectively. In GMA each term is kept separately, with parameters γ_{ij} and f_{ijk}. All these parameters have biochemical interpretation: the proportional parameters are the rate constants and the exponents are the kinetic orders of the corresponding chemical reactions. Further details of the advantages of BST are fully described and explored elsewhere (Voit, 2000).

1.3.3. *Parameter estimation* A major bottleneck of modeling metabolic networks is the estimation of parameters from measured variables. This step consists on finding the parameters that minimize an objective function, usually the weighted sum of squares of the residuals between the simulated and experimental data points. This procedure leads to a non-linear least squares problem (Seber and Wild, 1989). Additional constraints on the parameters can be introduced that reflect biochemical restrictions on the rate and kinetic values. Since the systems are usually non-linear, there are no current automatic methods that guarantee convergence to a global optimum. However, some techniques specifically developed for metabolic networks can find local solutions by performing the simulation of differential equations combined with convenient non-linear numerical optimization methods (Mendes and Kell, 1998).

1.3.4. *Structural Analysis* Global identifiability is a structural property of models and is related with the possibility of separately and uniquely recover the unknown parameters from complete error-free input-output data (Audoly *et al.*, 2001; Bellu *et al.*, 2007). This is still a difficult problem for non-linear systems, where no general algorithms exist that can determine *a priori* if the parameter optimization procedure has an unique solution.

Local identifiability regards systems that have a finite number of distinguishable parameter values, *i.e.*, for almost every solution there is a neighborhood where no other solutions exist. This type of identifiability is related to sensitivity analysis, in particular with the properties of the sensitivity matrix of the linearized system (Jacquez and Perry, 1990). In fact, the parameters or their linear combinations that do not influence the measured state variables are unidentifiable.

1.4 *Paper contribution*

This paper presents a case study of application of the traditional stability analysis based on the Jacobian matrix of the model equation, combined with the approach of singular value decomposition of model output sensitivities to show how a preliminary structural model can be reformulated in simplified form to substantiality improve the parameter estimation task. In fact, by performing the preliminary analysis of the structural model the paper shows that it is possible

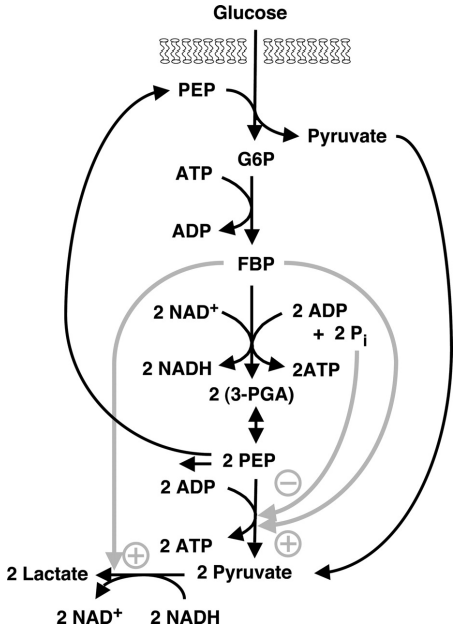

Fig. 1. Metabolic pathway of glycolysis in *L. lactis*. This figure represents a simplified model of glycolysis, from glucose uptake to the production of lactate. Also shown (in grey) activation and inhibitory signals. In (Voit *et al.*, 2006*b*).

to identify a cluster of state variables with fast equilibration dynamics that can be lumped into a single state variable. This procedure eliminates practically unidentifiable fast modes and allows to estimate a reduced parameter set that accurately reproduces, upon simulation, the original experimental time series.

1.5 *Paper structure*

Section 2 presents the dynamical model of glycolysis that served as a starting point to this work. In Section 3 a structural stability and sensitivity analysis of this model is performed and a new, reduced model is proposed. In Section 4 the parameter estimation is carried out and the new model is validated against the original experimental data. Finally Section 5 discusses the improvements obtained with this procedure and some possible future adjustments.

2. DYNAMIC MODELING OF THE METABOLIC NETWORK

The starting point of the present work was a previous glycolysis model proposed recently (Voit *et al.*, 2006*b*). A simplified network of the biochemical reactions was used, reproduced in Figure 1. The corresponding differential equations for this pathway were defined using the BST approach and read as follows:

$$
\begin{aligned}
\dot{X}_1 &= -\beta_1 X_1^{h_{11}} X_2^{h_{12}} X_5^{h_{25}} \\
\dot{X}_2 &= \beta_1 X_1^{h_{11}} X_2^{h_{12}} X_5^{h_{25}} - \beta_2 X_2^{h_{22}} \mathrm{ATP}^{h_{2\mathrm{ATP}}} \\
\dot{X}_3 &= \beta_2 X_2^{h_{22}} \mathrm{ATP}^{h_{2\mathrm{ATP}}} - \beta_3 X_3^{h_{33}} P_i^{h_{3P_i}} \mathrm{NAD}^{h_{3\mathrm{NAD}}} \\
\dot{X}_4 &= 2\beta_3 X_3^{h_{33}} P_i^{h_{3P_i}} \mathrm{NAD}^{h_{3\mathrm{NAD}}} + \alpha_4 X_5^{g_{45}} - \beta_4 X_4^{h_{44}} \\
\dot{X}_5 &= \beta_4 X_4^{h_{44}} - \beta_1 X_1^{h_{11}} X_2^{h_{12}} X_5^{h_{25}} - \alpha_4 X_5^{g_{45}} \\
&\quad -\beta_{51} X_3^{h_{513}} X_5^{h_{515}} P_i^{h_{51P_i}} - \beta_{52} X_5^{h_{525}} \\
\dot{X}_6 &= \beta_1 X_1^{h_{11}} X_2^{h_{12}} X_5^{h_{25}} + \beta_{51} X_3^{h_{513}} X_5^{h_{515}} P_i^{h_{51P_i}} \\
&\quad -\beta_{61} X_6^{h_{616}} X_3^{h_{613}} \mathrm{NAD}^{h_{61\mathrm{NAD}}} - \beta_{62} X_6^{h_{626}} \\
\dot{X}_7 &= \beta_{61} X_6^{h_{616}} X_3^{h_{613}} \mathrm{NAD}^{h_{61\mathrm{NAD}}}
\end{aligned}
\tag{2}
$$

Due to several problems found during the estimation procedure, several metabolites were considered as input signals. This was the case for glucose, where it was very difficult to reproduce the sigmoid type decay with this type of formalism. Other metabolites, such as Pi, NAD+/NADH and ATP, also constitute a major problem since they participate in many other reactions within the cells. Therefore it is extremely hard to isolate the fluxes only due to glycolysis from other production and consumption rates resulting from other pathways. Nevertheless the fitting of this model, although far from being fully accomplished, provides important insights regarding the design of the pathway (Voit *et al.*, 2006*a*).

The data used was obtained through NMR, as referred in the Introduction. In particular, multivariate time series of metabolite concentrations of glucose (X_1), G6P (X_2), FBP (X_3), 3-PGA (X_4), PEP (X_5), pyruvate (X_6), lactate (X_7), Pi, NAD+, NADH and ATP were obtained when the system was perturbed with glucose pulses of different concentrations and under several experimental conditions. In the present work, a bolus of 40 mM of glucose is applied in aerobic and anaerobic conditions (Neves *et al.*, 2002). It is noteworthy that it is not always possible to measure all metabolite concentrations during the procedure due to intrinsic detection limits of the experimental method.

3. STRUCTURAL ANALYSIS

Structural analysis is an essential preliminary step for parameter identification in metabolic networks. It is useful to recognize the presence of redundant structural information, due to practically unobservable modes, or of non-influential parameters, due to linear dependency, that will both lead to difficulties during parameter estimation, so-called practical unidentifiability. For this purpose some prior information on model parameters is required around which the model equations are linearized. This analysis suffers from self-reference, which limits the validity of the results. This apparently circular paradox must be solved heuristically by assigning reasoned numerical parameter values and by adjusting them as good as possible to adapt model predictions to actual measurements. The more general global *a priori* identifiability analysis would not require this prior information, but would be hardly applicable because the models considered do not fall into yet solvable model categories (Bellu *et al.*, 2007).

3.1 *Local stability and identifiability of dynamics*

To analyze local stability of dynamic modes consider a generic system of non-linear differential equations: $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p})$, $\mathbf{x}(0) = \mathbf{x}_0$, where $\mathbf{x}(t)$ is the nominal trajectory associated with nominal model inputs, $\mathbf{u}(t)$, and parameter values, $\mathbf{p}$. A small perturbation in the initial condition, $\delta\mathbf{x}_0$, determines a perturbation in the state trajectory that obeys: $\delta\dot{\mathbf{x}}(t) = \mathbf{A}\,\delta\mathbf{x}(t)$; $\delta\mathbf{x}(0) = \delta\mathbf{x}_0$, with Jacobian matrix $\mathbf{A} = \nabla_{\mathbf{x}} f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p})$. For simplicity it is assumed that $\mathbf{A}$ is time-invariant, such that $\delta\mathbf{x}(t)$ can be expressed analytically as a function of the eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ and eigenvectors $\mathbf{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathbf{A}$, that is $\mathbf{A}\,\mathbf{v}_i = \lambda_i\,\mathbf{v}_i$, or in matrix notation $\mathbf{A}\,\mathbf{V} = \mathbf{V}\,\Lambda$, with $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$. In particular,

$$\begin{aligned}\delta\mathbf{x}(t) &= e^{\mathbf{A}\,t}\delta\mathbf{x}_0 = \mathbf{V}\,e^{\Lambda\,t}\mathbf{V}^{-1}\delta\mathbf{x}_0 \\ &= \sum_i c_i\,\mathbf{v}_i\,e^{\lambda_i\,t} \end{aligned} \quad (3)$$

It follows that $\delta\mathbf{x}(t)$ vanishes only if all the eigenvalues have negative real part, *i.e.* if the system is asymptotically stable.

From a model identification perspective, fast modes associated to a real eigenvalue λ become unobservable if they vanish with much shorter time constant $\tau = 1/\lambda$ than the temporal resolution of measurements determined by the sampling schedule $\{t_k, k = 1, \ldots, N\}$. The model output equations are described, in general, by non-linear functions, such as: $\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t), \mathbf{p}, t)$. In the present study, $\mathbf{g}$ is the identity because all metabolites (state variables) are measurable with NMR. Therefore, the whole state trajectories spanned by the eigenvectors of the fast modes can be considered unmeasurable and the model parameters associated with these trajectories become in practice very difficult to estimate. In this analysis the Jacobian matrix is actually time-varying and must be calculated for some nominal parameter values. These latter were taken from the preliminary solution of (Voit *et al.*, 2006*b*) and results in Table 1 have been calculated in quasi steady state conditions by running the simulation sufficiently long to exclude the initial perturbation dynamics.

Table 1. Eigenvalues (1st row) and eigenvectors of the linearized system.

	-1506	-0.877	-0.394	-0.305	-0.0383	$-8.84{\cdot}10^{-6}$	0
1	0	0	0.362	0	0	0	0
2	0	0	-0.401	0	0.843	0	0
3	0	0.245	-0.032	0	0.038	0	0
4	-0.707	0.117	0.269	0	0.42	0.982	0
5	0.707	0.022	0.052	0	0.082	0.191	0
6	0	-0.962	-0.795	1	0.325	0	0
7	0	-0.002	0.001	-0.001	-0.009	-0.001	1

The eigenvalue with largest absolute value (-1506) differs three orders of magnitude from the one ranked next. The corresponding eigenvector: $\mathbf{v}_1$ = [0, 0, 0, -0.707, 0.707, 0, 0] indicates that the fast mode consists in a fast convergence to equilibrium of X_4 (3-PGA) and X_5 (PEP). Since the transformation $X_4 + X_5$ belongs to the null space of $\mathbf{v}_1$, the model can be simplified by lumping X_4 and X_5 into a single state variable. By canceling out the fast mode, some model parameters are also eliminated, *i.e.* α_4, β_4, g_{45} and h_{44} are replaced by k_{45} (see Section 3.3), although the eliminated parameters do not seem the most critical as regards practical identifiability as discussed next.

3.2 *Local parameter identifiability analysis*

The classic local identifiability analysis is based on the evaluation of linear dependency among parameter sensitivities of model outputs (Jacquez and Perry, 1990). In short, a small parameter perturbation, $\delta\mathbf{p}$, around a nominal value $\mathbf{p}$ produces a perturbation in model outputs: $\delta\mathbf{y}(t) \approx \nabla_{\mathbf{p}}\mathbf{y}(t, \mathbf{p})\,\delta\mathbf{p}$. By organizing the matrix of multi-output observations with discrete-time sampling, $\{\mathbf{y}_i(t_j)\}_{i,j}$, into a single column vector $\mathbf{Y}$, one can express measured output variations due to parameter perturbations as: $\delta\mathbf{Y} = \mathbf{S}\,\delta\mathbf{p}$, where $\mathbf{S}$ is the sensitivity matrix $\{\partial\mathbf{Y}_i/\partial\mathbf{p}_j\}_{i,j}$. To eliminate the effect of parameter scaling, relative variations can be considered computing: $\delta\mathbf{p}/\mathbf{p} \triangleq \mathrm{diag}(\mathbf{p})^{-1}\,\delta\mathbf{p}$ implies scaling of sensitivities such that $\delta\mathbf{Y} = \bar{\mathbf{S}}\,(\delta\mathbf{p}/\mathbf{p})$, with $\bar{\mathbf{S}} = \mathbf{S}\,\mathrm{diag}(\mathbf{p})$. Absolute (relative) parameter variations from nominal values are locally identifiable if $\mathbf{S}$ ($\bar{\mathbf{S}}$) has full rank. By using singular value decomposition (SVD) one obtains the factorization $\mathbf{S} = \bar{\mathbf{U}}\Sigma\bar{\mathbf{V}}^T$, with $\bar{\mathbf{U}}$ and $\bar{\mathbf{V}}$ eigenvector matrices of $\mathbf{S}\mathbf{S}^T$ and $\mathbf{S}^T\mathbf{S}$, respectively, and Σ diagonal with non-negative decreasing elements (singular values, SV) equal to the square root of the eigenvalues of $\mathbf{S}^T\mathbf{S}$. The rank, the number of non-zero SV, is in practice determined by normalizing all SV with respect to the largest one and by setting a cutoff, *e.g.* the square root of machine precision. The practically unidentifiable parameter subspace is spanned by the last columns of $\bar{\mathbf{V}}$with (nearly) zero SV.

Figure 2 shows the SV for the scaled sensitivities calculated for the glycolysis model. SVD was applied to model sensitivities simulated up to 13 min, where the initial transient dynamics vanishes, using the solution proposed previously. The last SV appears to be in practice negligible when compared to the first one. According to the last column of $\bar{\mathbf{V}}$, it corresponds to parameter h_{525}, which is the exponent of the loss rate of PEP ($\dot{X}_5$) in equation 2. Also the two preceding SV (23 and 24) indicate a possible lack of identifiability for a subspace of parameters. From the corresponding columns of $\bar{\mathbf{V}}$ this involves parameters β_{61}, β_{62}, h_{51Pi}, h_{616}, h_{626} that intervene in the transformation from pyruvate to lactate, suggesting that the model may be overparametrized in the description of this final transformation.

3.3 *Model reduction*

A first modification to the initial model was to tentatively describe extracellular glucose decay. This was

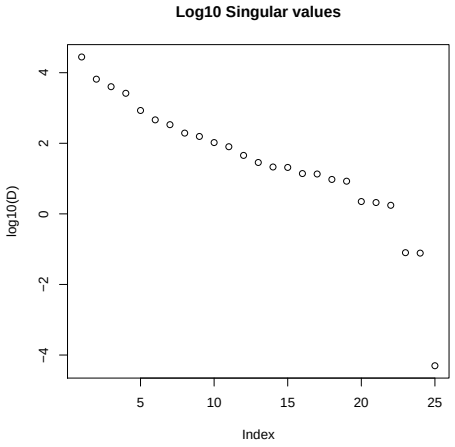

Fig. 2. Singular values of the scaled sensitivity matrix of the glycolysis model.

achieved by changing the equation for that flux ($\dot{X}_1$) to include a time-dependency to correctly describe a sigmoid type consumption.

According to the previously described structural analysis, showing that the reversible reaction between 3-PGA (X_4) and PEP (X_5) is extremely fast, both pools were merged into a unique state variable, $X_{45} = X_4 + X_5$. This is also justified by the data that support the hypothesis of an invariant ratio between the concentrations of these two metabolites. By defining this ratio as k_{45}, *i.e.* $X_4 = k_{45}\, X_5$, the equations to compute the individual concentrations from X_{45} become: $X_4 = k_{45}\, X_{45}/(1 + k_{45})$ and $X_5 = X_{45}/(1 + k_{45})$, respectively.

The results from the practical identifiability analysis, suggesting an over-parameterization of some parts of the model, were not yet taken into account because the fitting to experimental data of the above modified model evidenced some limitations that will require a more extensive revision.

For the Pi, ATP and NAD+/NADH the previous decision of keeping these time series as input signals to the model was maintained. In fact, given the extreme difficulty of expressing the variation of these ubiquitous metabolites, it is better to use them as external signals of the system. Since they present some irregularities and need to be extrapolated beyond the defined interval, cubic splines and interpolation were used to infer the parameters and correctly simulate the obtained solution. The equations of this new proposed model are:

$$
\begin{aligned}
\dot{X}_1 &= -k(1 + \alpha t^{\beta})X_1 \\
\dot{X}_2 &= \beta_1 X_1^{h_{11}} X_2^{h_{12}} X_5^{h_{25}} - \beta_2 X_2^{h_{22}} \mathrm{ATP}^{h_{2\mathrm{ATP}}} \\
\dot{X}_3 &= \beta_2 X_2^{h_{22}} \mathrm{ATP}^{h_{2\mathrm{ATP}}} - \beta_3 X_3^{h_{33}} P_i^{h_{3Pi}} \mathrm{NAD}^{h_{3\mathrm{NAD}}} \\
\dot{X}_{45} &= 2\beta_3 X_3^{h_{33}} P_i^{h_{3Pi}} \mathrm{NAD}^{h_{3\mathrm{NAD}}} - \beta_1 X_1^{h_{11}} X_2^{h_{12}} X_5^{h_{25}} \\
&\quad -\beta_{51} X_3^{h_{513}} X_5^{h_{515}} P_i^{h_{51Pi}} - \beta_{52} X_5^{h_{525}} \\
\dot{X}_6 &= \beta_1 X_1^{h_{11}} X_2^{h_{12}} X_5^{h_{25}} + \beta_{51} X_3^{h_{513}} X_5^{h_{515}} P_i^{h_{51Pi}} \\
&\quad -\beta_{61} X_6^{h_{616}} X_3^{h_{613}} \mathrm{NAD}^{h_{61\mathrm{NAD}}} - \beta_{62} X_6^{h_{626}} \\
\dot{X}_7 &= \beta_{61} X_6^{h_{616}} X_3^{h_{613}} \mathrm{NAD}^{h_{61\mathrm{NAD}}}
\end{aligned}
\tag{4}
$$

4. PARAMETER ESTIMATION AND VALIDATION

4.1 *Estimation procedure*

The model equations were implemented using the modelling software tool Pansym (Thomaseth, 2003), which takes as input a structural description of a system and generates symbolically the model's differential and output equations by means of algorithmic procedures pertaining to Bond Graph theory. The equations are further manipulated symbolically to determine, analytically, derivatives necessary to assemble the model Jacobian matrix as well as the sensitivity differential equations of model dynamics necessary to calculate the output sensitivities with respect to estimated parameters. The software produces Fortran source code for the numerical simulation by means of standard ordinary differential equation solvers for non-stiff (4th/5th order Runge Kutta) or stiff systems (LSODA). The simulation routines are interfaced with the non-linear parameter estimation routines made available by the open-source statistical programming environment R (http://www.r-project.org/).

The solutions obtained with this procedure for both experiments (anaerobic and aerobic conditions) are represented in Table 2.

Table 2. Solution obtained for the parameters. Glucose pulse of 40 mM under aerobic and anaerobic conditions (see Eq. 4).

Parameter	Aerobic	Anaerobic
k	0.0530251	0.124738
α	0.0419958	0.134194
β	2.68092	2.6674
β_1	7.20321	5.86735
h_{11}	0.997546	1.25193
h_{12}	-1.48643	-1.06961
h_{25}	0.38576	0.288687
β_2	0.345889	0.379794
h_{22}	1.54399	2.83465
$h_{2\mathrm{ATP}}$	1.51599	0.26203
β_3	0.338423	0.181082
h_{33}	1.09298	1.02783
h_{3Pi}	0.258372	-0.137778
$h_{3\mathrm{NAD}}$	-0.0966562	0.174484
β_{52}	0.134164	0.000447956
h_{525}	0.0940446	0.103001
β_{51}	0.862421	0.683548
h_{513}	0.7663	0.854603
h_{515}	0.0382342	0.0921911
h_{51Pi}	0.211149	-0.279396
β_{61}	0.0324743	0.0306174
h_{616}	0.675486	0.84484
h_{613}	1.03221	0.958361
$h_{61\mathrm{NAD}}$	-0.0519436	0.38804
β_{62}	1.74742	1.915812
h_{626}	1.40312	1.08493
k_{45}	2.04035	1.54695

4.2 *Validation against experimental data*

After performing the optimization procedure described above it is possible to assess the accuracy of the model

obtained by comparing the simulated with the original data. Figure 3 represents the original experimental points and the corresponding simulations generated with the estimated parameters (Table 2).

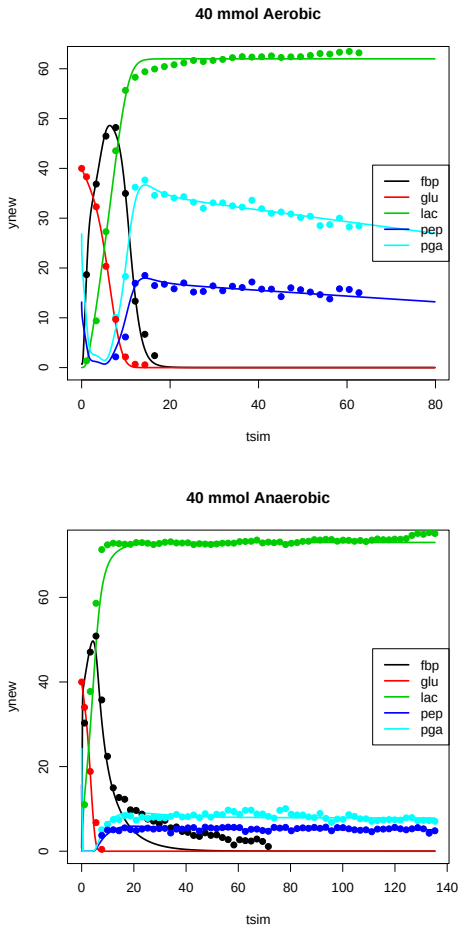

Fig. 3. Simulation of obtained solutions. a) Aerobic conditions. b) Anaerobic conditions.

The analysis of the curves shows a good fitting of the original experimental data in both aerobic and anaerobic conditions.

5. DISCUSSION AND POSSIBLE FURTHER ADJUSTMENTS

The analysis of the results shows that previous problems found in parameter estimation might have been partially due to overmodelling of the reversible reaction 3-PGA/PEP. In fact, this reaction occurs almost instantaneously, and therefore these two pools can be considered at equilibrium, with constant K=1.5-3, as estimated from data obtained by *in vivo* NMR.

Another interesting point is the longer tail of FBP. This might be the result of an incomplete model with some missing equation and not a parameter estimation problem. The analysis of these simulation results provides important clues about missing factors and induce the modeller to rethink the original equations to cope with incomplete information.

6. CONCLUSION

The structural analysis of metabolic networks significantly improves model identification and parameter inference. This case study shows that the elimination of a fast reaction solves the problem of optimizing the remaining parameters. In fact, the pooling of the two state variables allowed the convergence of a standard non-linear least squares algorithm used for fitting the modified model to experimental data. This provided a more accurate picture on the model's capability to fit experimental data and on its limitations. The analysis of the metabolic model structure and the subsequent careful choice of metabolites to be included can thus constitute a key step to overcome some of the problems previously encountered during the estimation procedure.

7. REFERENCES

Audoly, S., G. Bellu, L. D'Angio, M. P. Saccomani and C. Cobelli (2001). Global identifiability of nonlinear models of biological systems. *IEEE Trans Biomed Eng* **48**(1), 55–65.

Bellu, G., M. P. Saccomani, S. Audoly and L. D'Angio (2007). DAISY: a new software tool to test global identifiability of biological and physiological systems. *Comput Methods Programs Biomed* **88**(1), 52–61.

Carson, E.R., C. Cobelli and L. Finkelstein (1983). *The mathematical modeling of metabolic and endocrine systems*. John Wiley & Sons. New York.

Jacquez, J. A. and T. Perry (1990). Parameter estimation: local identifiability of parameters. *Am J Physiol* **258**(4 Pt 1), E727–36.

Mendes, P. and D. Kell (1998). Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* **14**(10), 869–83.

Neves, A. R., R. Ventura, N. Mansour, C. Shearman, M. J. Gasson, C. Maycock, A. Ramos and H. Santos (2002). Is the glycolytic flux in *Lactococcus lactis* primarily controlled by the redox charge? kinetics of NAD(+) and NADH pools determined in vivo by 13C NMR. *J Biol Chem* **277**(31), 28088–98.

Savageau, M. A. (1969). Biochemical systems analysis. I. some mathematical properties of the rate law for the component enzymatic reactions. *J Theor Biol* **25**(3), 365–9.

Seber, G.A.F. and C.J. Wild (1989). *Nonlinear Regression*. Wiley. New York.

Thomaseth, K. (2003). Multidisciplinary modelling of biomedical systems. *Comput Methods Programs Biomed* **71**(3), 189–201.

Voit, E., A. R. Neves and H. Santos (2006*a*). The intricate side of systems biology. *Proc Natl Acad Sci U S A* **103**(25), 9452–7.

Voit, E. O., J. Almeida, S. Marino, R. Lall, G. Goel, A. R. Neves and H. Santos (2006*b*). Regulation of glycolysis in *Lactococcus lactis*: an unfinished systems biological case study. *Syst Biol (Stevenage)* **153**(4), 286–98.

Voit, Eberhard O. (2000). *Computational analysis of biochemical systems : a practical guide for biochemists and molecular biologists*. Cambridge University Press. New York.