

EXPLORING THE EFFECT OF VARIABLE ENZYME CONCENTRATIONS IN A KINETIC MODEL OF YEAST GLYCOLYSIS

JÓZSEF BRUCK^{1,2} WOLFRAM LIEBERMEISTER¹
jbruck@staff.hu-berlin.de lieberme@molgen.mpg.de

EDDA KLIPP¹
klipp@molgen.mpg.de

¹ *Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany*

² *Humboldt University Berlin, Department of Biology, Chair of Theoretical Biophysics, Invalidenstr. 42, 10115 Berlin, Germany*

Metabolism is one of the best studied fields of biochemistry, but its regulation involves processes on many different levels, some of which are still not understood well enough to allow for quantitative modeling and prediction. Glycolysis in yeast is a good example: although high-quality quantitative data are available, well-established mathematical models typically only cover direct regulation of the involved enzymes by metabolite binding. The effect of various metabolites on the enzyme kinetics is summarized in carefully developed mathematical formulae. However, this approach implicitly assumes that the enzyme concentrations themselves are constant, thus neglecting other regulatory levels – e.g. transcriptional and translational regulation – involved in the regulation of enzyme activities. It is believed, however, that different experimental conditions result in different enzyme activities regulated by the above mechanisms. Detailed modeling of all regulatory levels is still out of reach since some of the necessary data – e.g. quantitative large scale enzyme concentration data sets – are lacking or rare. Nevertheless, a viable approach is to include the regulation of enzyme concentrations into an established model and to investigate whether this improves the predictive capabilities. Proteome data are usually hard to obtain, but levels of mRNA transcripts may be used instead as clues for changes in enzyme concentrations. Here we investigate whether including mRNA data into an established model of yeast glycolysis allows to predict the steady state metabolic concentrations for different experimental conditions. To this end, we modified an established ODE model for the glycolytic pathway of yeast to include changes of enzyme concentrations. Presumable changes were inferred from mRNA transcript level measurement data. We investigate how this approach can be used to predict metabolite concentrations for steady-state yeast cultures at five different oxygen levels ranging from anaerobic to fully aerobic conditions. We were partly able to reproduce the experimental data and present a number of changes that were necessary to improve the modeling result.

Keywords: yeast; glycolysis; fermentation; respiration; kinetic modeling; metabolic regulation

1. Introduction

Cellular metabolism is one of the key components of living systems. Its most basic functions are to generate the energy and the building blocks necessary to sustain the cells' life. Elucidation of central carbon metabolism, the source of energy for all heterotrophic life, is one of the success stories of biochemistry: function and mechanism of most of its components are known in considerable detail. A large class of the regulatory mechanisms of metabolism is well understood: the catalytic function of many enzymes is influenced by metabolites present in the cell. This kind of interactions have been successfully

quantified in enzyme kinetic laws, which has led to ODE based models of metabolic pathways with considerable predicting power, as described in [4, 2, 7] and applied among others in [9, 5, 11].

However, metabolism is also regulated by other functional units of the cell, most importantly the transcriptional-regulatory system. It acts by changing the concentration of various enzymes via regulated production and degradation. This kind of regulation is necessary for the cell to steer its metabolism to meet its needs under various conditions. However, change in protein levels is usually not implemented in kinetic models: these typically adopt kinetic expressions for the included reactions with fixed maximal velocities, which amounts to the implicit assumption of constant enzyme concentrations. One of the possible reasons is that quantitative data on concentrations of single proteins in different experimental conditions are still lacking or rare.

A fundamental determinant of the concentration of an enzyme's active form, and hence, its activity, is the amount of mRNA transcripts presents in the cell. However, many other layers of regulation exist, e.g. at the level of translation and allosteric regulation of the final protein among many others. It is controversial to what extent the final enzyme activity is determined by or correlated to the concentrations of its mRNA components. While genome-wide comparisons between mRNA and enzyme concentrations exist [1, 3], the abundance of a given set of proteins and their corresponding transcription rates should be systematically compared in different cell states to obtain a clearer picture. To the authors' knowledge such studies are not yet available.

Based on an established ODE-based model of yeast glycolysis, we present an approach for modeling how metabolism is regulated by the transcriptional-regulatory system. In the model we include the change in enzyme concentrations in various experimental conditions. We used experimental data [12] from steady state yeast cultures with five different oxygen levels ranging from anaerobic to fully aerobic conditions. We implemented the change in enzyme concentrations by changing the maximal rates of the enzymatic reactions. For the above mentioned reasons, we determined these changes from mRNA concentration measurements, using them as inputs for the model. The model allows for computing metabolite concentrations and fluxes, which we compared to the corresponding experimental values. We performed parameter estimation to determine a set of parameters which best fit for the experimental data.

The main question posed is the following: to what extent can experimental data for different cell states be explained by including expression data in the model under the assumption that biochemical reaction rates obey rate laws known from enzyme kinetics?

2. Methods

2.1. Experimental data

We used metabolite concentration and flux data from Wiebe et al. [12] obtained from cultures of *Saccharomyces cerevisiae* CEN.PK113-1A grown in glucose-limited chemostat cultures (dilution rate $D=0.10/h$). External conditions in these cultures could be controlled to a high extent. Steady-state cultures were obtained under one anaerobic (0% oxygen) and four aerobic conditions (0.5%, 1%, 2.8%, 20.9% oxygen in the inlet gas) with all other external conditions being kept constant. Measured quantities included biomass, concentration of external metabolites^a (Glucose, Ethanol, Glycerol), of intermediate metabolites (G6P, F6P, F16P, PEP, PYR, ATP, ADP, AMP, and the sum of 3PG and 2PG concentrations), net fluxes (consumption rates of oxygen and glucose and exhaust rate of ethanol, glycerol and CO₂) per unit of biomass, and relative fold changes of the mRNA concentrations compared to the anaerobic cultures for 69 genes with functions in carbon metabolism.

2.2. Mathematical model

We constructed a mathematical model of central carbon metabolism in *S. cerevisiae* based on the glycolytic pathway model by Teusink *et al.* [11]. The original model was based on measurements on steady state cell cultures under anaerobic conditions by comparison of experimental data of concentrations and fluxes of intermediate and external metabolites.

The sum of the concentrations $[NAD^+]$ and $[NADH]$ is a conserved moiety of the model. The adenosine species $[ATP]$, $[ADP]$ and $[AMP]$ are not dynamical variables of the original model, instead, they were written as analytic expressions in term of the sum of high-energy phosphates. These were obtained under the assumptions that a) the sum of their concentrations is conserved, and b) the reaction catalyzed by adenosine kinase is fast in comparison to the other reactions, and hence in equilibrium. The metabolites GAP and DHAP are lumped to a single chemical species called “triose” reflecting the assumption that the transforming reaction between them (catalyzed by TPI) is also in equilibrium. The kinetic constants were largely obtained from experiments and fitted only to a minimal extent. The side branches of glycolysis contained in the model were

^aAbbreviations: G6P: Glucose-6-phosphate; F6P: Fructose-6-phosphate; F16P: Fructose-1,6-bisphosphate; Triose-P: sum of GAP: Glyceraldehyde-3-phosphate and DHAP: Dihydroxyacetone phosphate; BPG: 1,3-bisphosphoglycerate; 3PG and 2PG: 3- and 2-phosphoglycerate respectively; PG: sum of 3PG and 2PG; PEP: Phosphoenolpyruvate; ACA: Acetaldehyde; AMP, ADP, ATP: Adenosine-mono-, di-, and triphosphate, respectively. NAD^+ , $NADH$: oxidation states of Nicotinamide adenine dinucleotide.

Enzymes: ENO: Enolase; GAPDH: D-glyceraldehyde-3-phosphate dehydrogenase; ADH1, ADH2: Alcohol dehydrogenase 1 and 2, respectively; HK: Hexokinase; PGI: Phosphoglucose isomerase; PFK: Phosphofructokinase; ALD: Aldolase; G3PDH: Glycerol-3-phosphate-dehydrogenase; PGK: Phosphoglycerate kinase; PGM: Phosphoglycerate mutase; PYK: Pyruvate kinase; PDC: Pyruvate decarboxylase; FBPI: Fructose-1,6-bisphosphatase.

found to be crucial to reproduce the data. Glycerol producing branch was simplified to the reaction catalyzed by the enzyme G3PDH. The products ethanol and CO₂ were assumed to diffuse out of the cell quickly, thus their concentrations inside and outside the cell as equal in the steady state. We obtained the original model in SBML format from the JWS online database [14] (download on 26th May 2008). It is worth noting that the kinetic expression for PFK in the published SBML file differs from the one described in the article [11]; we adopted the latter version.

Table 1. List of the reactions which were added to the Teusink model. Numbers in brackets refer to reactions in Fig.1. Square brackets denote concentrations described by dynamic variables of the mathematical model. All other quantities are parameters of the model: their values are either adopted from [11], set to the measured values of external metabolites, or estimated.

Name	Reaction	Reaction rate expression
Adenosine kinase (19)	$\text{ATP} + \text{AMP} \rightleftharpoons 2 \text{ADP}$	$V_{\text{mAK}} \left([\text{ATP}] [\text{AMP}] - [\text{ADP}] [\text{ADP}] / K_{\text{eqAK}} \right)$
G6P consumption (3)	$\text{G6P} + \text{ATP} \rightarrow \text{ADP}$	$V_{\text{mG6P}} [\text{G6P}] [\text{ATP}]$
glycerol transport (9)	$\text{GLY} \rightleftharpoons \text{GLY}_{\text{out}}$	$V_{\text{mGLY}} \left([\text{GLY}] - \text{GLY}_{\text{out}} \right)$
TCA (16)	$4 \text{NAD} + \text{ADP} + \text{ACE} \rightleftharpoons 4 \text{NADH} + \text{ATP} + 2 \text{CO}_2$	$V_{\text{mTCA}} \left([\text{ACE}] [\text{NAD}] [\text{ADP}] - [\text{NADH}] [\text{ATP}] / K_{\text{eqTCA}} \right)$
respiration (18)	$0.5 \text{O}_2 + \text{NADH} + 2.5 \text{ADP} \rightleftharpoons \text{NAD} + 2.5 \text{ATP}$	$V_{\text{mRESP}} \left(\text{O}_2 [\text{NADH}] [\text{ADP}] - [\text{NAD}] [\text{ATP}] / K_{\text{eqRESP}} \right)$
ATP consumption (20)	$\text{ATP} \rightarrow \text{ADP}$	$V = V_{\text{mATPasc}} [\text{ATP}]$
PDC (15)	$\text{PYR} \rightleftharpoons \text{ACE} + \text{CO}_2$	$V_{\text{mpDC}} \frac{\left([\text{PYR}] / K_{\text{mPDC}}^{\text{PYR}} \right)^n - \left([\text{ACE}] / K_{\text{mPDC}}^{\text{ACE}} \right)^n}{\left(1 + [\text{PYR}] / K_{\text{mPDC}}^{\text{PYR}} \right)^n + \left([\text{ACE}] / K_{\text{mPDC}}^{\text{ACE}} \right)^n}$
FBP1 (6)	$\text{F16P} \rightarrow \text{F6P}$	$V_{\text{mFBP}} [\text{F16P}]$

We modified the original Teusink model in several details to fit our purposes. Reaction numbers refer to Fig. 1, for details of the stoichiometry and the kinetic expressions see Table 1.

1. We explicitly modeled the concentrations of AMP, ADP, and ATP as dynamic variables. The adenosine kinase reaction (reaction 19), modelled with reversible mass-action kinetics, was introduced to maintain the moiety conservation of the pool of these species.

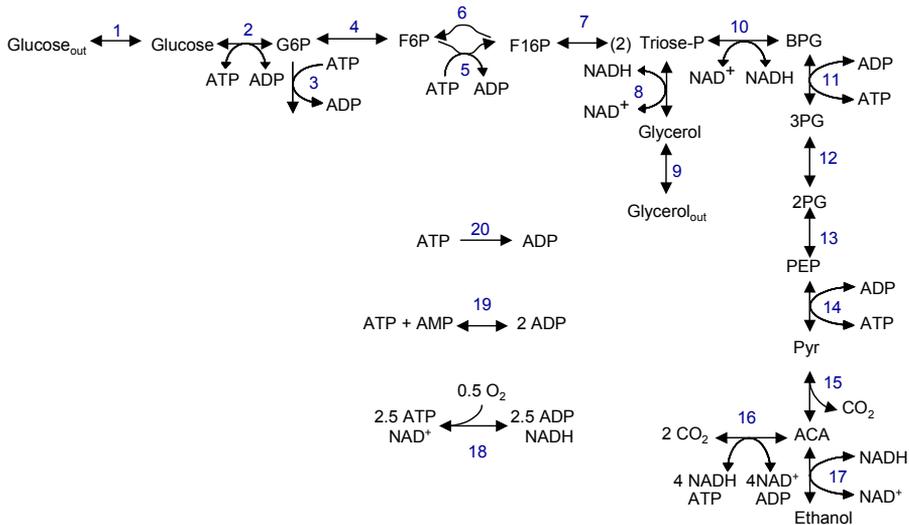


Fig. 1. Reaction scheme of the kinetic model of glycolysis. The numbers refer to the following reactions 1:glucose transport; 2:HK; 3:G6P consumption; 4:PGI; 5:PFK; 6:FBP1; 7:ALD; 8:G3PDH; 9:glycerol diffusion; 10:GAPDH; 11:PGK; 12:PGM; 13:ENO; 14:PYK; 15:PDC; 16:TCA; 17:ADH; 18:respiration; 19:adenosine kinase; 20:ATP consumption. Reaction 7 produces two Triose-P per F16P, as indicated. Subscript “out” refers to species outside the cell. Reactions which were added to the original model by Teusink et al. [11] are listed in Table 1.

2. Instead of considering two side chains with constant fluxes at G6P (leading to glycogen and trehalose), we replaced them by a single G6P-consuming process (reaction 3) with irreversible mass action kinetics. We did not distinguish between them since we do not have measurements for metabolites or fluxes of this branches that would allow for distinguishing one from the other.
3. At the end of the glycerol-producing branch, we included a diffusive transport reaction for glycerol out of the cell (reaction 9).
4. The original model contains the TCA cycle in the form of a succinate production branch. In this reaction, two molecules of acetaldehyde are consumed to produce one molecule of succinate. Since our model is aimed to describe respiration, we replaced this reaction by a simplified description of a running TCA cycle (reaction 16) and the respiratory chain (reaction 18): we consider two reactions which consume acetaldehyde and oxygen to produce energy in form of ATP and NADH as well as the by-product CO_2 [8]. We assumed that CO_2 concentration in the cell remains low due to rapid diffusion, therefore we did not include it in the backward rate expression of reaction 16.
5. The ATP-consuming reactions are summarized in one effective ATPase reaction (20). In the original model, this reaction had constant flux which we replaced by irreversible mass-action kinetics.

6. Reversibility of the main glycolytic chain is crucial to obtain qualitative agreement with the measured fluxes. Therefore, we changed the irreversible Hill kinetics of the PDC reaction (reaction 15) to a reversible kinetics by including an additional term with a parameter $K_{\text{mPDC}}^{\text{ACE}}$ in the original rate expression as shown in the table.
7. Also the reaction catalyzed by PFK is irreversible and modeled without product inhibition. To allow for a slowing down of the glycolytic flux at higher product concentrations, we included the reaction catalyzed by FBP1 into the model (reaction 6). In gluconeogenesis, this reverses the effect of PFK, but without involvement of ATP.

All other parts of the model including the values of the parameters which are not explicitly mentioned in this article were adopted from [11].

In contrast to glucose and glycerol, it was assumed that ethanol diffusion through the cell membrane is fast enough to keep the outer and inner concentrations close, therefore no distinction was made between extra- and intracellular ethanol. The resulting model has 20 reactions and 17 dynamic variables representing metabolite concentrations. It is available in SBML and text formats as supplementary material.

2.3. *Transcriptional regulation and external metabolites*

In order to include transcriptional regulation in our model, we write reaction rates for reaction i in the experimental condition j as

$$V_{ij} = E_{ij} R_i(C_j), \quad (1)$$

where E_{ij} denotes the concentration of the active form of the corresponding enzymes in the steady state cultures, R_i denotes the rest of the kinetic expression, and C_j denotes the vector of all metabolite concentrations at condition j .

We compared the four aerobic states to the anaerobic state^b. We indicate quantities belonging to this condition by the subscript $j=0$. Transcriptional regulation was accounted for in the following way: for each enzymatic reaction i and each aerobic condition j , we calculated E_{ij}/E_{i0} , the relative change of enzyme concentration of the four aerobic states from the transcription data by setting

$$\frac{E_{ij}}{E_{i0}} = g_{ij}^a \quad (2)$$

where the scaling exponent a is a constant and g_{ij} denotes the transcription fold change associated with reaction i in condition j . By definition, $g_{i0}=1$ for every reaction.

Assuming that the activity of an enzyme is proportional to its concentration, we describe the effect of transcriptional regulation on the reaction rate V_{ij} through replacing it by

$$V_{ij}^* = g_{ij}^a V_{ij} \quad (3)$$

for each reaction i and condition j . For most enzymatic reactions, we calculated g_{ij} as the arithmetic mean of the measured mRNA concentration fold change for the genes associated with reaction i . See the Appendix for the list of genes associated with each enzymatic reaction.

Since the transcriptional activities corresponding to Enolase and GAPDH were not measured, for these reactions we computed the value of g_{ij} by averaging the values for the next-neighbor reactions (PGM, PYK) and (ALD, PGK), respectively.

Also the reaction ADH was treated differently. The expression data for ADH1, together with ADH2, the isoenzyme responsible for converting ethanol to acetaldehyde, indicate that net Ethanol production is shut down with growing oxygen supply, reaching virtually zero in fully aerobic condition. The resulting ethanol flux also reflects this behavior (Fig. 2). For simplicity, instead of including ADH2, which would involve yet more unknown parameters, we only included the reaction for ADH1 and described its regulation, by setting g_{ij} to the values of the measured ethanol flux, normalized to the anaerobic condition. The resulting g_{ij} values for all experimental conditions are shown in Fig. 2.

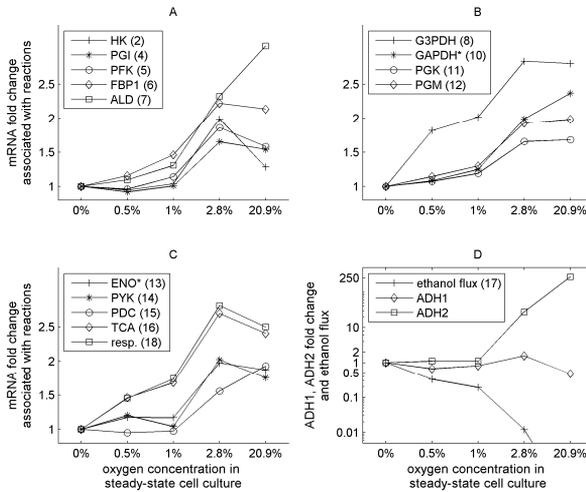


Fig. 2. **A,B,C:** fold change of mRNA concentration associated with reactions in the mathematical model, normalized to the anaerobic state (denoted by g_{ij} in the text). The values were calculated from the expression data of the genes associated with each reaction as given in the appendix. Numbers in brackets refer to reaction numbers in Fig.1. For reactions marked with (*), no transcript analysis was undertaken; the values were averaged from neighbors as described in the text. **D:** fold change of the genes ADH1 and ADH2 and the resulting ethanol flux. At the highest oxygen concentration the flux drops to zero (not shown in the logarithmic scale).

The external metabolites glucose, glycerol and ethanol were represented by the model species $\text{Glucose}_{\text{out}}$, $\text{Glycerol}_{\text{out}}$ and Ethanol (cf. Fig. 1). Their concentrations were set to constant values according to the experimental data: $\text{Glucose}_{\text{out}}$ was set to the

corresponding concentration in the inlet feed solution, 55.55 mmol/l, in all conditions. Measured glycerol concentrations was 8.90 mmol/l for the anaerobic condition, and zero for all aerobic conditions. Measured ethanol concentration was 75.37 mmol/l, 59.01 mmol/l, 47.56 mmol/l, 3.66 mmol/l, and 0 mmol/l for the conditions with 0%, 0.5%, 1%, 2.8%, and 20.9% oxygen, respectively.

2.4. Parameter estimation

We performed parameter estimation on a subset of the model parameters to achieve agreement with the data. Metabolite concentrations were compared with concentrations in the model. The measured fluxes for glucose, oxygen, ethanol, glycerol and CO₂ were each compared to the rates $0.5r_1$, r_{18} , r_{17} , r_8 , $r_{15} + 2r_{16}$, respectively, where r_i denotes the rate of the reaction i in Fig. 1.

We quantified goodness of fit for each possible set P of values for the estimated parameters by the following cost function:

$$\text{Cost}(P) = \sum_{k,j} \frac{(V_{kj}^{\text{sim}} - V_{kj}^{\text{exp}})^2}{\sigma_{kj}^2} + (\exp(\kappa) - 1) \quad (3)$$

where we denote the steady-state value of a metabolite concentration or flux k for the condition j by V_{kj}^{sim} and V_{kj}^{exp} for simulation results and experimental data values, respectively. V_{kj}^{sim} values were obtained by runs of 10000 seconds of simulation time. $1/\sigma_{kj}^2$ is a weight factor in which σ is often set to the value of the experimental error. However, this choice does not reflect an appropriate weight measure in our case, since we do not expect to be able to reproduce the experimental data within the errors. At the same time, small experimental error of a quantity does not necessarily correlate with higher importance of a good fit compared to other quantities with larger errors. To assign the same weight to all relative deviations, we set σ_{kj} to be proportional to V_{kj}^{sim} in the following way:

$$\begin{aligned} \sigma_{kj} &= 0.15 \cdot V_{kj}^{\text{exp}}, & \text{in case } V_{kj}^{\text{exp}} &\neq 0, \\ \sigma_{kj} &= 0.15 \cdot \min_j (V_{kj}^{\text{exp}}), & \text{in case } V_{kj}^{\text{exp}} &= 0, \end{aligned}$$

where V_{kj}^{exp} denotes all nonzero values for the concentration or flux k among all conditions. To avoid non-steady state solutions, we introduced a penalty term $(\exp(\kappa) - 1)$ in the cost function. The term κ quantifies the deviation of the solution from steady state. It is defined as

$$\kappa = \sum_{k=1}^{17} \sum_{l=1}^3 |x_k(t_l^{\text{comp}}) - x_k(t^{\text{last}})|,$$

where $x_k(t^{\text{last}})$ denotes the simulated value of the concentration x_k at the last time instance $t^{\text{last}}=10000$ sec, and $x_k(t_l^{\text{comp}})$ denotes its value at some earlier time instance t_l^{comp} . The values t_l^{comp} were chosen as $t_1^{\text{comp}} = 0.5 \cdot t^{\text{last}}$, $t_2^{\text{comp}} = 0.75 \cdot t^{\text{last}}$, and $t_3^{\text{comp}} = 0.8 \cdot t^{\text{last}}$.

We estimated a total of 31 parameters which was an acceptable number given a total number of data points of 70. The values of all other parameters were taken from [11]. We estimated the following groups of parameters:

1. Since the experiment by Wiebe *et al.* and the experiments underlying the Teusink model differ in the experimental conditions and the yeast strain, we could not rely on the absolute enzyme concentrations to be comparable. Therefore, we fitted all V_m values and the diffusion coefficient for reaction 9 (20 parameters).
2. We also fitted the new kinetic parameters of the reactions that were added to the original model (4 parameters, cf. Table 1.)
3. The sum of $[\text{NAD}^+]$ and $[\text{NADH}]$ is a conserved quantity of the model, determined by the initial concentrations of these species. Since experimental data were not available, we estimated this quantity for each condition separately (5 parameters).
4. We fitted the scaling exponent a from Eq. (2).
5. Concentration units: reaction rate expressions in our model are based on enzyme kinetics and hence the concentrations of the reactants need to be known. However, all metabolite concentrations and fluxes were measured in units per gram dry weight of biomass (gDW). The values were determined after collecting the cells from the culture by centrifugation, washing by distilled water, and drying to constant weight at 100°C . To determine the cytosol concentrations of the measured values, the net cytosol volume of the cells of 1gDW is needed. Although estimates for this number exist (amounting 1g dry weight to 2 ml cytosol, [13]), we preferred to fit this quantity along with the other parameters of the model.

2.5. Genetic algorithm and semiglobal search

We adopted the genetic algorithm Differential Evolution [16] to search for a parameter set with best fit to the experimental data. In a truly global search, parameters could assume any values between zero and infinity, with the aim to find a global optimum of fit. However, we found that this approach was not practicable since many parameter sets are, although in principle viable, not practical to work with. Some may not generate a steady state (for example due to accumulation of F16BP), others require long computation times.

Therefore we developed the following semi-global approach: at a given time, only a limited region of parameter space was screened. This was achieved by limiting each parameter to a certain range. If a parameter repeatedly (4 out of the last 5 times) produced values in the upper or lower 20% of its search range, the range was relocated such that the parameter value corresponding to the hitherto best result became the center of the search range of this parameter. If this process would have resulted in a negative value for the lower limit, the latter was set to zero. The width of the search range was

kept constant during the process and was determined at the beginning of the parameter estimation to be $\left[(1-r)p_0, (1+r)p_0 \right]$ where p_0 denotes the initial value of the parameter and r was set to 0.5.

Since evaluating the cost function (Eq. 3) involves integrating a system of 20 differential equations numerically, we used various software tools to convert the SBML model to an executable C-code for faster integration [6, 10]. The integrator used in the process was CVODE from Sundials [15].

3. Results

3.1. Parameter estimation

We ran four parameter estimation processes to find model parameter values which produce the best possible fit to the experimental data. Fig. 3. shows the evolution of the goodness of fit (as quantified by the cost function) and the value of five parameters during a parameter estimation process (data for all parameters published as supplementary file).

Most, but not all parameters converged to a certain value. However a unique

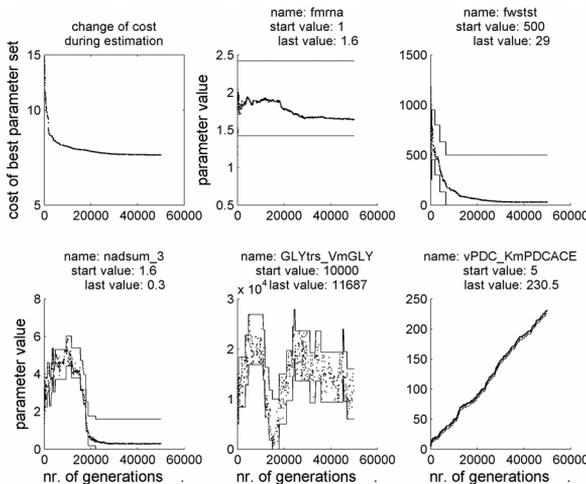


Fig. 3. Evolution of goodness of fit (cost) of the best parameter set (top left) and corresponding values of five of the 31 model parameters during a parameter estimation process of ca. 49000 generations. Shown are values corresponding to the parameter set with the best fit to data (as defined by the cost function, see text) after a certain number of generations. The momentary search range for each parameter (see text for description) is specified by upper and lower bounds (shown by lines). The parameters *fmrna* (called *a* in the text), *fwstst*, *nadsum_3* are explained in section 2.4 under points 4, 5 and 3, respectively. *GLYtrs_VmGLY* denotes the diffusion coefficient in reaction 9, and *vPDC_KmPDCACE* denotes the constant in reaction 15 (cf. Table1).

parameter set with best fit could not be determined within the available computing time (24 hours of computing time amounting to roughly 7000 generations on an AMD 3800+ processor), since a number of parameters did not converge to similar values during these parameter estimations (data not shown).

Notably, these parameter sets produced mostly similar simulation values. As shown in Fig. 4., the largest quantitative differences between the predictions generated by the four parameter sets can be observed in the simulation results for F16P concentration (0% oxygen) and of the O_2 .

Some of the parameters were seemingly not, or only weakly determined, i.e. their values did not matter for change in the cost function. This was to be expected, since only about two third of the dynamical variables of the model is measured. Since the number of data points (70) is more than twice the number of parameters (31), we do not expect

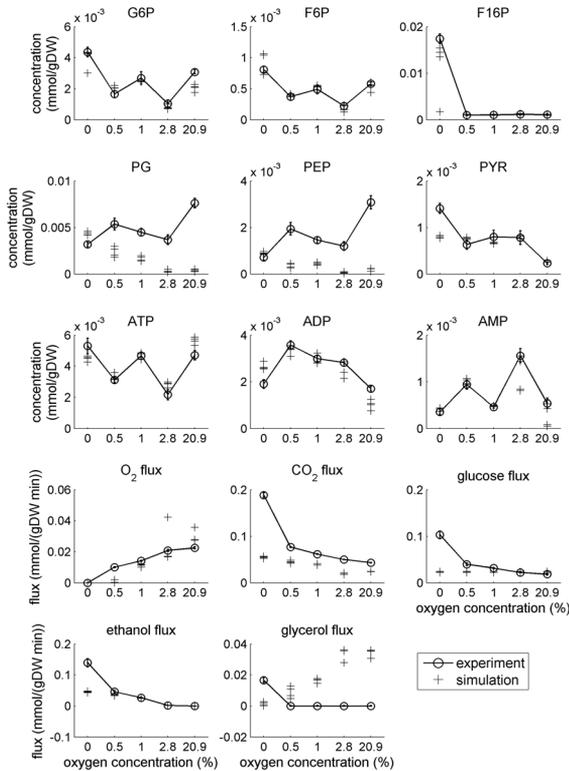


Fig. 4. Concentrations and fluxes of metabolites: comparison of experimental values and simulation results with parameter sets resulting from four different estimation runs (which we terminated after 49973, 49959, 12435, and 12678 generations of the Differential Evolution algorithm).

overfitting to occur.

3.2. Comparison to experimental data

Experimental data of metabolite concentrations and fluxes over the five experimental conditions and corresponding simulation results are shown in Fig. 4.

The mathematical model was able to reproduce the experimental data to varying extents. In general, the concentration values of the metabolites in upper glycolysis (G6P,

F6P, F16P) and of the adenosine species (ATP, ADP, AMP) were reproduced better than those of lower glycolysis (PG, PEP, PYR) and the metabolite fluxes. In the latter group, the decrease of pyruvate concentration and that of carbon dioxide and ethanol fluxes with higher oxygen concentrations was reproduced as a tendency, but neither the exact absolute values, nor the sharp difference between anaerobic and aerobic conditions was reproduced correctly by the model. Also the predicted increase of the O₂ flux with external oxygen concentration was qualitatively correct, but the experimental values for cultures with higher oxygen concentrations were not reproduced correctly. However, other measured quantities show distinctly different behavior from our simulation results: the model failed to reproduce the measured decrease of the glucose flux with increasing oxygen concentration predicting nearly constant simulation values instead, as well as the similar behavior of the glycerol flux for which it predicted an increase.

4. Discussion

We explored whether an established mathematical model of yeast glycolysis, created to describe one anaerobic condition, could be extended to describe different cell states corresponding to experimental conditions with various oxygen concentrations. To this end, regulation through enzyme concentration changes and a simple model of the TCA-cycle and respiratory chain were included in the model.

As enzyme data were not available, we assumed that differential enzyme concentrations and differential mRNA concentrations are related by a power law with a single exponent. This assumption is of course questionable: enzyme concentrations are also regulated posttranscriptionally, so changes in enzyme levels can, in principle, take place irrespective of differential expression and *vice versa*. However, a monotonous relationship between the two quantities holds, at least, on average; in a comparison of mRNA and protein levels for different genes, a scaling exponent of about 0.6 has been reported [1].

In our attempt to reproduce the experimental data, we were led to make a number of further changes in the original model. Most remarkably, we found that a number of reactions of the pathway (either by altering the kinetics as in PDC, or by including a reverse reaction such as FBP) needed a reversible description for the following reason: As we compared steady state cultures with higher concentration of oxygen, the data clearly showed that flux through glycolysis decreases in spite of upregulation of most enzymes in carbon metabolism (while the ethanol producing branch is simultaneously shut down). Although at first somewhat counterintuitive, this behavior can be reproduced without introducing posttranscriptional regulation into the model. In our case, the flux of the pathway is redirected from fermentation to respiration, i.e. to a branch with typically considerably lower reaction rates. This can result in a lower flux if concentrations of some metabolites rise enough to slow down the reactions producing them. Speaking in loose terms, the pipeline of the pathway becomes “jammed” which causes the flux slowing down. In contrast, fast diffusion of ethanol during fermentation may keep the lower glycolysis concentrations low, which speeds up the reactions. In principle, upregulating the enzyme production might even be an attempt of the cell to keep the flux through the pathway as high as possible. However, in a mathematical

model, this effect is only possible if the kinetics of each reaction is chosen such that the reaction rate is sufficiently slowed down by rising product concentration. This is true for reversible reactions, but not true for the irreversible kinetic expressions we replaced in the original model.

Although the introduced changes increased the agreement with the experimental data, at this stage the model did not agree with the data in a number of points. Probably most important is the measured decrease of glucose flux in spite of a general upregulation at higher oxygen concentrations. It is possible that further refinement of the model will lead to at least qualitative agreement with the experiment in this point.

There are a number of possible ways to refine the method presented here. Increasing the number of data points compared to the number of parameters to be estimated is, of course, desirable. An important special case would be the experimental determination of a remaining conserved quantity, the sum of $[NAD^+]$ and $[NADH]$. Lacking such data, we fitted this quantity for each condition separately. Regarding the model input, including appropriate measurement data on the transcriptional activity of GAPDH or Enolase would probably also improve the model. However, this can change its behavior only to the extent to which these values behave differently from their neighbors. The parameter estimation process was found very demanding in terms of computational power. Increasing its speed and finding well-defined parameter sets is a necessary technical step in further development.

Kinetic models have been successful in describing metabolism in cell states in which its explicit regulation through changing enzyme concentrations is negligible. Developing a class of models describing this additional layer regulation is a logical next step, and this might enable us to describe – or even predict - various states of the cell with one single model.

Acknowledgments

The authors thank to Marilyn Wiebe and Merja Penttilä of VTT Technical Research Centre, Helsinki, Finland for kindly providing experimental data and for stimulating discussions on the subject, as well as Preben G. Sørensen of the University of Copenhagen for drawing our attention to useful methods. The financial support of the Marie Curie EST project "Systems Biology" (EC contract number MEST-2-CT-2004-514169), of the Bacillus Systems Biology project (BaSysBio), and of the Yeast Systems Biology Network (YSBN) is gratefully acknowledged by the authors.

Appendix

Enzymatic model reactions and genes associated with them in brackets (for ENO, GAPDH, ADH, see text): HK(GLK1, HXK1), PGI(PGI1), PFK(PFK1, PFK2), ALD(FBA1), G3PDH(GPD1, GPD2), PGK(PGK1), PGM(PGM1), PYK(PYK1, PYK2); PDC(PDC1), TCA-Cycle(CIT3, KGD1, SDH1, SDH2, SDH3, SDH4, FUM1, LSC1,

LSC2, PDA1, PDB1, CIT2), Respiratory chain(CYB2, COX5a, COX5b, CYC1, CYC7, NDE1, NDE2), FBP1(FBP1) .

References

- [1] Beyer, A. et al., Post-transcriptional expression regulation in the Yeast S.c. on a genomic scale, *Molecular and cellular proteomics*, 3.11:1083-1092, 2004.
- [2] Fell, D., *Understanding the Control of Metabolism*, Portland Press, 1997
- [3] Greenbaum, D., et al., Comparing protein abundance and mRNA levels on a genomic scale, *Genome Biology*, 4:117, 2003
- [4] Heinrich, R.; Schuster, S. *The Regulation of Cellular Systems*. New York: Chapman and Hall, 1996.
- [5] Hynne, F., Danø, S. Sørensen, P.G., Full-scale model of glycolysis in *Saccharomyces cerevisiae*. *Biophys Chem.*, 94 (1-2):121-63, 2001
- [6] Keating, S.M, SBMLToolbox: an SBML toolbox for MATLAB users., *Bioinformatics*. 22(10):1275-7, 2006.
- [7] Klipp, E., Herwig, R., Kowald, A., Wierling, C. and Lehrach, H., *Systems Biology in Practice: Concepts, Implementation and Application*. Wiley-VCH, 2005.
- [8] Michal, G., *Biochemie-Atlas; Biochemical Pathways*, German ed., Spektrum Akademischer Verlag, 1999
- [9] Rizzi, M., Michael Baltés, Uwe Theobald and Matthias Reuss, In Vivo Analysis of Metabolic Dynamics in *Saccharomyces cerevisiae*: II. Mathematical Model, *Biotechnology and Bioengineering*, 55:592-608, 1997
- [10] Schmidt, Henning, Systems Biology Toolbox for MATLAB: A computational platform for research in Systems Biology, *Bioinformatics*, 22(4):514-515, 2006. (<http://www.sbtoolbox2.org>)
- [11] Teusink, B., et al.: Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *Eur J Biochem*. 267(17):5313-5329, 2000
- [12] Wiebe MG et al.: Central carbon metabolism of *Saccharomyces cerevisiae* in anaerobic, oxygen-limited and fully aerobic steady-state conditions and following a shift to anaerobic conditions *FEMS Yeast Research* 8(1):140-154, 2008
- [13] Wiebe, M.G., personal communication
- [14] <http://www.jjj.bio.vu.nl/>
- [15] <https://computation.llnl.gov/casc/sundials/description/description.html>
- [16] <http://www.icsi.berkeley.edu/~storn/code.html>