ARTICLES

Validation of a Metabolic Network for *Saccharomyces cerevisiae* Using Mixed Substrate Studies

Peter A. Vanrolleghem,*,† Patricia de Jong-Gubbels,‡ Walter M. van Gulik,§ Jack T. Pronk,‡ Johannes P. van Dijken,‡ and Sef Heijnen†

Departments of Bioprocess Engineering and Microbiology and Enzymology, Delft University of Technology, Julianalaan 67, 2628 BC Delft, The Netherlands, and EPFL-Institut de Génie Chimique, CH-Ecublens, 1015 Lausanne, Switzerland

Setting up a metabolic network model for respiratory growth of *Saccharomyces cerevisiae* requires the estimation of only two (energetic) stoichiometric parameters: (1) the operational *PO* ratio and (2) a growth-related maintenance factor *k*. It is shown, both theoretically and practically, how chemostat cultivations with different mixtures of two substrates allow unique values to be given to these unknowns of the proposed metabolic model. For the yeast and model considered, an effective *PO* ratio of 1.09 mol of ATP/mol of O (95% confidence interval 1.07−1.11) and a *k* factor of 0.415 mol of ATP/C-mol of biomass (0.385−0.445) were obtained from biomass substrate yield data on glucose/ethanol mixtures. Symbolic manipulation software proved very valuable in this study as it supported the proof of theoretical identifiability and significantly reduced the necessary computations for parameter estimation. In the transition from 100% glucose to 100% ethanol in the feed, four metabolic regimes occur. Switching between these regimes is determined by cessation of an irreversible reaction and initiation of an alternative reaction. Metabolic network predictions of these metabolic switches compared well with activity measurements of key enzymes. As a second validation of the network, the biomass yield of *S. cerevisiae* on acetate was also compared to the network prediction. An excellent agreement was found for a network in which acetate transport was modeled with a proton symport, while passive diffusion of acetate gave significantly higher yield predictions.

Introduction

For a long time, bioprocess intensification has relied on random mutagenesis of microbial strains or the nondirected improvement of the cultivation conditions. This labor-intensive approach is mainly taken because many potentially beneficial changes in the genetic makeup of an organism are not readily evident. Indeed, the enzyme systems governing the bottlenecks of a cell's metabolism may be situated far away from the enzyme system that finally synthesizes the desired biochemical compound. In many cases, fueling reactions of primary metabolism must be rerouted to achieve increased productivities (Vallino and Stephanopoulos, 1993).

In recent years, a more rational approach termed "metabolic engineering" emerged which aims to (1) identify the critical path (or the metabolic bottleneck) in a producing strains' metabolism, (2) direct the vast genetic engineering methodology to manipulate the identified enzyme systems, and in doing so, (3) minimize the efforts and corresponding expenses for process optimization. As Bailey (1991) points out, most often a new limitation arises after molecular modification so that the expected improvement is not completely attained. Hence, an iterative procedure of analysis, metabolic engineering, and evaluation is needed to obtain an optimized strain.

The identification of metabolic bottlenecks is based on a quantitative description of the metabolism of an organism. While kinetic constraints are sure to exist and methodologies for their identification have been proposed (e.g., biochemical systems theory: Savageau (1969) and Metabolic Control Theory: Kacser and Burns (1973), current knowledge of the dynamic behavior of complex enzyme systems is insufficient to allow application of the kinetic approach to metabolic bottleneck identification (Vallino and Stephanopoulos, 1993; Nielsen and Villadsen, 1994).

The insights in the stoichiometry of biochemical reaction networks on the other hand are more up to this task. With mass balances constructed over the enzyme reactions, a pseudo-steady-state assumption on the pools of metabolic intermediates and flux measurements of metabolites, significant information can be obtained on metabolic pathway utilization and potential limitations (Holms, 1986). However, while considerable knowledge has been gathered, some uncertainty remains on metabolic pathways, e.g on substrate transport mechanisms, compartmentation, and the stoichiometry of energy-

† Department of Bioprocess Engineering, Delft University of Technology.
‡ Department of Microbiology and Enzymology, Delft University of Technology.
§ EPFL-Institut de Génie Chimique.
* Author to whom all correspondence should be addressed. Current address: Department BIOMATH, University Gent, Coupure links 653, B-9000 Gent, Belgium.

S8756-7938(96)00022-7 CCC: $12.00

generating and -consuming processes. Therefore, the estimation of such unknown stoichiometric parameters and validation of the constructed metabolic network model (MNM) is an essential step before model predictions can be interpreted for metabolic engineering purposes or the design of optimal feeding strategies.

Until now, little effort has been spent on the calibration and validation of the stoichiometry of metabolic network models using experimental data. Most often, the models are constructed on the basis of a priori knowledge only, i.e. established biochemical reactions, thermodynamic irreversibilities, and energetic parameters obtained from the literature (Aiba and Matsuoko, 1979; Majewski and Domach, 1990; Vallino and Stephanopoulos, 1990; Goel et al., 1993; Varma et al., 1993). However, experimental data are a second source of information for model building (Vanrolleghem and Dochain, 1995) and have not been exploited sufficiently. For instance, isotope studies with radio- or mass-labeled substrates though hampered by high costs and experimental constraints (Bryers and Yeh, 1990; Zupke and Stephanopoulos, 1994) may provide highly informative data on intracellular flux distributions. Alternatively, careful measurement and evaluation of the component flows in and out of the cells and their dependence on the growth rate applied in chemostat cultivations (e.g., Varma and Palsson, 1994) can also serve model calibration and subsequent identification of metabolic constraints.

Moreover, the biomass yield of a microorganism on many substrates can be used very beneficially to obtain stoichiometric model parameters of metabolic networks (van Gulik and Heijnen, 1995). In this study, the dependence of biomass yields and the expression of key enzymes on the composition of a substrate mixture fed to chemostat cultivations is proposed as a new source of experimental data for network calibration. To validate the metabolic network, attention is focused on the capability of the network to predict biomass yields on different substrates and the correspondence between predicted metabolic fluxes and presence or absence of the corresponding enzyme activities in cell-free extracts.

Additionally, through the use of symbolic manipulation software, explicit expressions are derived which relate rates of reaction and biomass yields to the unknown stoichiometric parameters in the large metabolic network represented by a 88 × 84 stoichiometric matrix.

Throughout the paper, the uncertainty introduced by estimation of model parameters on the basis of noise-corrupted data was carefully followed. Relying to a large extent on evaluation of parameter sensitivities, the usefulness of symbolic manipulation software will be illustrated. This methodology proved also beneficial as it allowed considerable reduction in the computational burden associated with the parameter estimation task. Moreover, it was an invaluable tool to the theoretical identifiability study that was performed on this model to assess whether the unknowns could be given unique values at all, given the model and the proposed experiments.

Materials and Methods

Organism and Chemostat Cultivation. *Saccharomyces cerevisiae* T2-3D (Pronk et al., 1994) was cultivated in chemostats with a working volume of 1 L (Applikon, Schiedam, The Netherlands) at a dilution rate of 0.10 h^{-1}. Conditions applied were a temperature of 30 °C, pH controlled at 5.0, and dissolved oxygen level maintained above 25% of air saturation. Details can be found in de Jong-Gubbels et al. (1995). Medium composition contained a mineral and vitamin component as mentioned in de Jong-Gubbels et al. (1995), while carbon sources—glucose, ethanol, and acetic acid—were added to a final carbon concentration of 0.250 C-mol/L at the C-mol ratios reported in the text.

Analysis. Substrate and Metabolites. In the liquid phase, total organic carbon (TOC), glucose, ethanol, and acetate contents of the culture supernatant were assayed as described previously (de Jong-Gubbels et al., 1995).

In the gas phase, O_2 uptake and CO_2 production rates were quantified and calculated according to van Urk et al. (1988). The amount of CO_2 leaving the culture with the effluent was negligible due to the low pH.

Biomass Dry Mass and Composition. Methods used for determination of dry weights of washed culture samples, the elemental composition of biomass, and its protein content can be found elsewhere (de Jong-Gubbels et al., 1995).

Enzyme Assays. All assays were performed immediately after preparation of cell-free extracts using the methodology described in de Jong-Gubbels et al. (1995). Enzyme activities are expressed as micromoles of substrate converted per minute per milligram of protein. One unit is defined as the amount of enzyme catalyzing the conversion of 1 μmol of substrate/min.

Enzymes analyzed included fructose-1,6-bisphosphatase (EC 3.1.3.11), phosphofructokinase (EC 2.7.1.11), isocitrate lyase (EC 4.1.3.1), malate synthase (EC 4.1.3.2), phosphoenolpyruvate (PEP) carboxykinase (EC 4.1.1.32), pyruvate carboxylase (EC 6.4.1.1), pyruvate kinase (EC 2.7.1.40), hexokinase (EC 2.7.1.1), and glucose-6-P dehydrogenase (EC 1.1.1.49) (de Jong-Gubbels et al., 1995; Postma et al., 1988, 1989).

Metabolic Flux Analysis

Theory. The methodology used to study the metabolism of *S. cerevisiae* when grown on mixtures of glucose and ethanol is the metabolic flux analysis. It consists of quantifying the flow of metabolites through a network of reactions occurring in the cell. The first step in such analysis is to set up a stoichiometric model for these reactions. Considering the thousands of reactions taking place in the cell, it is obvious that not all of these can be included in the model. The biochemical reactions (n = 99) and considered metabolites (m = 98) on which this study is based are given in the Appendix. This basic information can be summarized in the following matrix notation:

$$\mathbf{S}\cdot C = \begin{bmatrix} S_1 & S_{12} & \cdots & S_{1m} \\ \vdots & \vdots & \vdots & \vdots \\ S_{n1} & S_{n2} & \cdots & S_{nm} \end{bmatrix} \begin{bmatrix} C_1 \\ \vdots \\ C_m \end{bmatrix} = 0 \qquad (1)$$

in which **S** is the stoichiometric matrix and C is the metabolite vector. Note that the consistency of such a network with respect to conservation laws for the elements (considered are C, H, O, N, S, and P) and charge l can be checked by applying the metabolite composition matrix **E**:

$$\mathbf{E} = \begin{bmatrix} C_1 & H_1 & O_1 & N_1 & S_1 & P_1 & l_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ C_m & H_m & O_m & N_m & S_m & P_m & l_m \end{bmatrix} \qquad (2)$$

and checking whether

$$\mathbf{S}\cdot\mathbf{E} = 0 \qquad (3)$$

The second step in metabolic model building consists of constructing a mass balance for each metabolite i in-

volved in a reactor system of volume V:

$$\frac{\mathrm{d}VC_i}{\mathrm{d}t} = V(r_i + \Phi_i) \tag{4}$$

where r_i is the net rate of conversion of metabolite i in the metabolic system and Φ_i denotes the net rate of transport of metabolite i over the system boundaries. Φ_i follows from flow and concentration measurements. The net conversion rate r_i is determined by the rates of all biochemical reactions v_j in which metabolite i is produced or consumed.

Two types of metabolites can be discerned: intracellular intermediate species and species which are transported across the cell membrane, e.g. nutrients, O_2, CO_2, and products. For the majority of metabolites that remain essentially intracellular, it holds that

$$r_i^{\text{intra}} = 0 \tag{5}$$

A second simplification results from the fact that only chemostat cultivation is considered in this study. Hence, the extracellular concentrations are constant in time and it follows that

$$r_i^{\text{extra}} = -\Phi_i \tag{6}$$

The rate vector R, composed of n reaction rates v_i and m net conversion rates r_i, is defined as

$$R = [v_1, v_2, ..., v_n, r_1, r_2, ..., r_m]^T \tag{7}$$

Note here that most net conversion rates r_i are known. Either they are zero when the component is intracellular or they can be obtained from relation 6. We also define an $m \times (n + m)$ coefficient matrix $\mathbf{A}$ composed of the previously defined reaction stoichiometry matrix $\mathbf{S}$ and an $(m \times m)$ identity matrix $-\mathbf{I}_m$ (Noorman et al., 1991):

$$\mathbf{A} = [\mathbf{S}_n - \mathbf{I}_m] \tag{9}$$

The balance equations for the m compounds of a system consisting of n biochemical reactions can now be denoted by

$$\mathbf{A}\cdot\mathbf{R} = 0 \tag{9}$$

This homogeneous system of m equations describes the metabolic network.

Solving the Metabolic Network. Use of the network involves the calculation of unknown rates, n reaction rates v_j and p net conversion rates r_i, where p is the number of metabolites with non-zero conversion rates. The rank of the $\mathbf{S}$ matrix determines the minimum number of rates that must be specified to provide unique values to the other rates, i.e. in total ($n + p -$ rank of $\mathbf{S}$) rates must be defined. The reduced row echelon form of the matrix $\mathbf{S}$, obtained using a standard software package (in this work Maple V, Math Soft, Inc., Cambridge, MA) provides the solution of the metabolic network, i.e. each unknown rate is given as a function of measured rates only.

Stoichiometric Metabolic Network Model (MNM). Metabolic networks of *S. cerevisiae* were constructed for aerobic growth on (mixtures of) glucose, ethanol, and acetate. The database of basic reactions used in our work on yeast metabolism is summarized in the Appendix. All decarboxylation and ATP-consuming reactions were considered unidirectional, while all others retained reversibility until proven otherwise. As this reaction database is also used for the description of other systems (e.g., see van Gulik and Heijnen, 1995), some reactions were not used in this *S. cerevisiae* network. These omitted reactions are indicated by "#" in the Appendix (for the argumentation, see below). The considered components are listed as well. In Figure 1 the central metabolic pathways are schematized.

For certain reactions multiple pathways have been described. To make a choice among these, a number of assumptions on the biochemistry and some design choices were made while constructing the network. These are reviewed below. First, the choice of the set of reactions is argued and, second, it is summarized how values were given to remaining unknown stoichiometric coefficients.

Reaction Network. In this work it is not attempted to incorporate the incomplete knowledge on cell compartmentation of *S. cerevisiae* in the metabolic network. The high uncertainty involved in situating particular reactions in one or the other compartment (mitochondria/cytoplasm) would require too many design choices, making the interpretation of the network predictions difficult. Moreover, compartmentation would introduce an additional level of complexity that was hoped not to be necessary for adequate description of the cell's metabolism.

Transport of solutes in and out of the cell was modeled according to the three possible transport mechanisms present in yeast (Cooper, 1982; Serrano, 1991): simple diffusion, facilitated diffusion, and active transport. The latter is based on proton symports for which the proton motive force is generated by a H^+-ATPase (r_{58}). Glucose uptake in *S. cerevisiae* occurs via facilitated diffusion so that the reaction for active transport (r_{50}) available in the reaction database can be omitted (see the Appendix). Ethanol on the other hand enters the cell by passive diffusion. With respect to acetate uptake, the picture is less clear and both possibilities were evaluated (see further). Evidently, for the glucose/ethanol network, reaction r_{52} was not taken into account. Uptake of phosphate and sulphate ions is based on proton symports with H^+/ion stoichiometries of 2 and 3, respectively. For ammonium uptake the model of Roon et al. (1977) was adopted, i.e. transport occurs via a proton symport with a H^+/NH_4^+ stoichiometry of 1.

Polymerization reactions of protein, polysaccharides, and RNA each involve the conversion of ATP to ADP. In the reaction network it was assumed that four ATPs are consumed per peptide bound, three ATPs are converted per nucleic acid monomer, and one ATP is used for extension of a polysaccharide with one carbohydrate unit (Stanier et al., 1987).

Some dehydrogenases exhibit a dual cofactor specificity: either NAD or NADP can be used as the electron acceptor. To prevent singularity of the metabolic network, a design choice had to be made for the reactions concerned (r_{13}/r_{14}; r_{17}/r_{18}). In a subsequent paper (Vanrolleghem et al., 1996) it will be shown how this choice can be argued and validated. At this stage, it is sufficient to state that the pentose phosphate pathway ($r_{26}-r_{31}$) was assumed to be the only source of NADPH, whereas the reactions catalyzed by acetaldehyde dehydrogenase and isocitrate dehydrogenase only had NADH as a cofactor (hence, $r_{14} = r_{18} = 0$ and both reactions are eliminated from the network, see the Appendix).

Since the experimental data of this study were collected for one growth rate (0.1 h^{-1}), no distinction could be made between growth- and non-growth-associated maintenance. Here, both processes were lumped and total maintenance energy requirements were modeled proportional to biomass synthesis, i.e. k mol of ATP is consumed per C-mol of biomass synthesized (r_{99}). Note that this single process now accounts not only for maintenance but also for futile cycles in the cell. Verduyn et al. (1991) introduced this concept, and it was

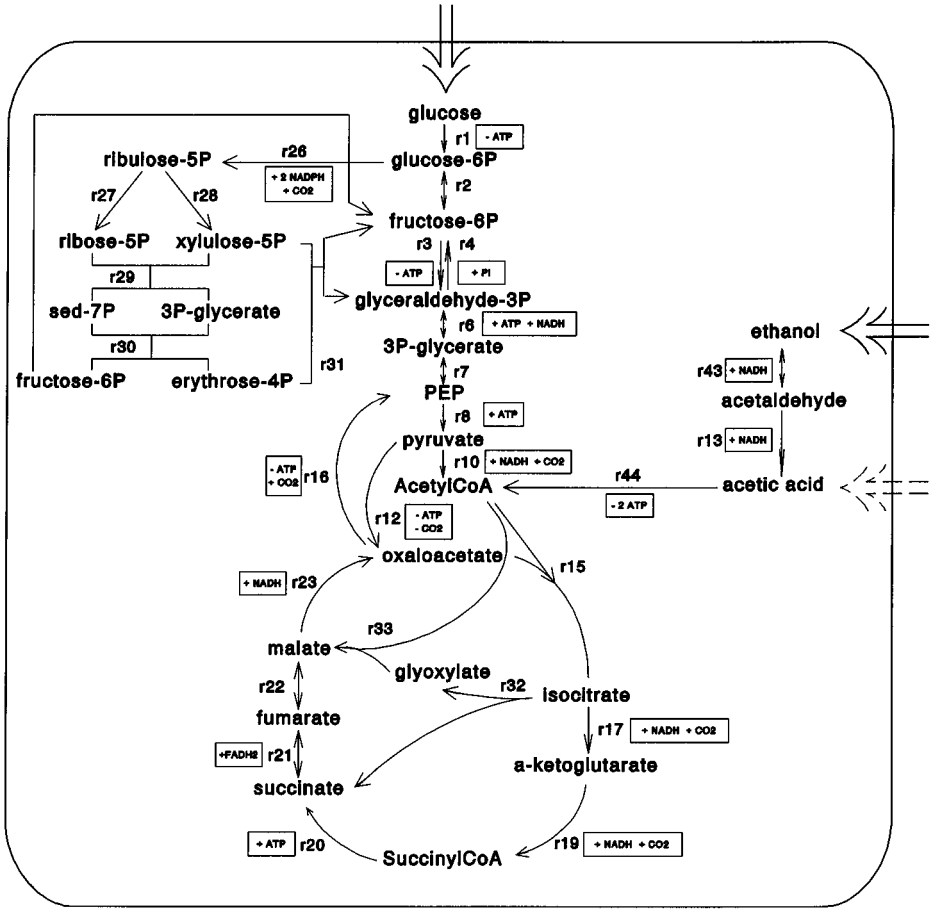

Figure 1. Central pathways of the metabolic network of *S. cerevisiae* growing on mixture of glucose and ethanol.

later also applied by Varma et al. (1993) and van Gulik and Heijnen (1995). Note that lumping these processes in k allows elimination of the futile cycle described by reaction r_{35} and the explicit ATP-hydrolysis reaction r_{95}.

Some reactions given in the Appendix could be eliminated because the compounds involved were not present in the studied process, i.e. citrate (r_{39}/r_{57}), lactate (r_{40}/r_{53}), gluconate (r_{42}/r_{55}), external pyruvate, and succinate (r_{54}/r_{56}). The alternative ATP-generation systems r_{37} and r_{38} were not considered either. Some reactions used in the study of *Corynebacterium glutamicum* (van Gulik and Heijnen, 1995) were obviously omitted in this investigation ($r_{24}/r_{25}/r_{65}$). Finally, the singularity induced by inclusion of the overflow mechanism for pyruvate to acetyl-CoA (via $r_{11}/r_{13}/r_{44}$) was broken by eliminating the first step.

In order to describe the metabolism for the whole range of glucose/ethanol mixtures, it was necessary to construct different network configurations. Thermodynamic constraints dictate that certain reactions are unidirectional, and it was observed that, as the ethanol fraction in the feed increased, these reactions were prone to reversal. In most cases another enzyme catalyzes the reversed reaction, albeit with different cofactors or the introduction of energy equivalents. Three such reaction pairs were identified to be critical, leading to four metabolic regimes denominated MNM I, II, III, and IV (Figure 2).

At high glucose fractions (MNM I), ethanol is only used as a source of acetyl-CoA (r_{44}), gradually replacing the acetyl-CoA synthesized from glucose via pyruvate (r_{10}). In this way ethanol is fueling the TCA cycle and the acetyl-CoA requiring pathways for fatty acid and amino acid synthesis. In MNM I the anaplerotic reaction to replenish carbon in the TCA cycle is reaction r_{12}.

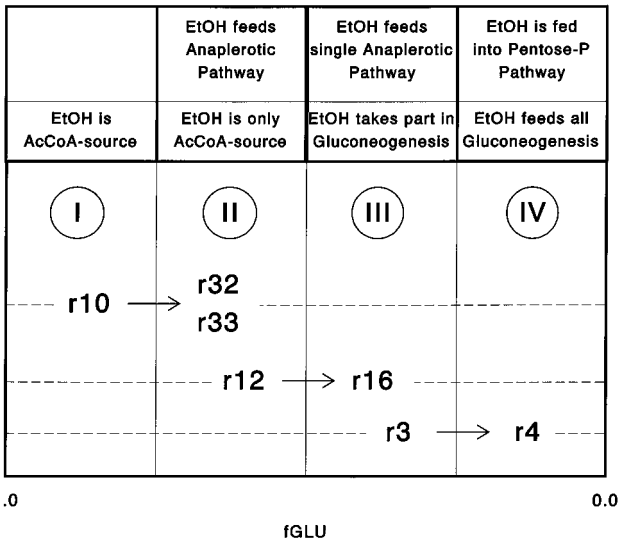

Figure 2. Overview of the four metabolic regimes, the three corresponding switchpoints, and the reactions involved for *S. cerevisiae* growing on all possible mixtures of glucose and ethanol.

At higher ethanol feeds (MNM II), all acetyl-CoA is synthesized from ethanol and two anaplerotic systems operate concurrently. As the ethanol fraction in the medium increases, reaction r_{12} is gradually replaced by the glyoxylate shunt, r_{32}/r_{33}. At the same time, to prevent reversal of the flux, reaction r_{10} (which is known to be irreversible) is eliminated from the network.

At still higher ethanol conversions, gluconeogenic reactions are incorporated to allow an increased portion of anabolism to be based on ethanol. First, reaction r_{16}

is incorporated and reaction r_{12} is eliminated to prevent a futile cycle ($r_8-r_{12}-r_{16}$) (MNM III).

Finally, reaction r_3 is replaced by reaction r_4 (MNM IV). It is evident that intermediate reactions r_6 and r_7 and reaction r_2 have reversed fluxes to participate in the gluconeogenesis.

Summarizing, switching between the different networks occurs at switch values of f_{GLU} ($f_{X\rightarrow Y}^{switch}$) where a specific irreversible reaction tends to reverse, i.e. when its flux becomes zero and a new reaction switches on. The reaction switches to examine are r_{10}/r_{32}-r_{33}, r_{12}/r_{16}, and r_3/r_4.

Stoichiometric Coefficients. Important stoichiometric coefficients that have to be given values include the elemental composition of biomass and the monomeric composition of the polymers protein, lipids, polysaccharides, and RNA. Polysaccharides were described as polyglucose, while lipids were considered to consist of two components, oleate and palmitoleate (Bruinenberg et al., 1983). The monomer composition of RNA was also obtained from Bruinenberg et al. (1983), while the amino acid content of protein was taken from Oura (1972). With a mean 42% protein, biomass was identical to the composition reported by Verduyn et al. (1991). Ash content was found to be 6%. The elemental composition of 1 mol (100 g) of biomass used is $C_{3.907}H_{6.290}O_{1.984}N_{0.586}P_{0.022}S_{0.0061}$ (Table 1, see supporting information). It must be noted that biomass composition was assumed to be independent of the different glucose/ethanol mixtures in the feed. This was verified by experimental analysis of protein, carbon, and nitrogen content of biomass samples taken from cultivations performed with various glucose-to-ethanol ratios. The following composition was measured at a 50% ethanol fraction in the feed: 40.2% protein, 46.7% C, and 7.82% N on a g/g basis. No significant trend in these values could be found as the composition of the reservoir medium changed (results not shown).

The only assumption made concerning the energetic efficiency of the oxidative phosphorylation concerns the equivalence of FADH and NADH (reactions r_{34} and r_{36}) in *S. cerevisiae* (Verduyn et al., 1991). As shown in the sequel, it is possible to estimate the operational *PO* ratio, i.e. the effective amount of ATP formed per mole of oxygen reduced, from well-designed experiments. Similarly, the value of the maintenance factor k is not fixed a priori but is estimated on the basis of experimental data as shown below.

Results

Estimation of Unknown Parameters in the MNM. ***Calculation of yields and switch functions.*** Given the assumptions and design choices made above, the model structure is defined completely, i.e. the set of components, reactions, and constraints (with their respective switching points). The task of validation of this metabolic network is now only to be preceded by the estimation of the two unknown stoichiometric parameters remaining in the network: (1) the operational *PO* ratio, i.e. the effective amount of ATP produced in the electron transport system per oxygen reduced, and (2) the maintenance energy factor k, which lumps the growth- and non-growth-associated maintenance requirements expressed as mol of ATP/C-mol of biomass produced.

In order to obtain numerical values for the reaction and net conversion rates, values for *PO* and k have to be inferred from a comparison between metabolic network predictions and experimental data. To this end, it is advantageous to write the model predictions as an explicit function of the unknown stoichiometric parameters since this allows direct evaluation of the discrepancy between predictions and data. This significantly decreases parameter estimation calculations because the network need not be solved (involving a computing-intensive calculation of the reduced row echelon form) separately for each *PO* and k value evaluated.

While it may be rather cumbersome to obtain solutions of the network for which the parameters are explicit (see, e.g., van Gulik and Heijnen, 1995), symbolic manipulation software is fit for this task. A typical result for a particular rate of reaction (r_3, phosphofructokinase in MNM III) is as follows (similar—nonlinear—relations are obtained for all other nonmeasured conversion rates):

$$v_3 = [-0.119(6PO - 1)r^{ethanol} + (0.230PO + 0.333k + 0.466)r^{glucose}]/(4.227PO + 2k + 4.221) \quad (10)$$

Hence, in this example, estimates of *PO* and k can be obtained, provided the reaction rate v_3 and the net conversion rates of glucose and ethanol are measured at different ratios of ethanol and glucose conversion. To this end, model predictions for v_3 are compared with the measured rates, typically using a sum of squared errors criterion that is minimized by variation of *PO* and k.

It is far easier to measure net conversion rates than (intracellular) reaction rates. Therefore, preference is given to estimation of the parameters on the basis of only conversion rates or, equivalently, apparent yield coefficients (ratios between conversion rates). It is important to note that not all yield coefficients are functions of *PO* and k. In this work conversion rates of glucose, ethanol, CO_2, H_2O, H^+(E), NH_4^+(E), O_2, P_i(E), SO_4^{2-}(E), and biomass are considered. Symbolic solution of the metabolic networks reveals that biomass yield coefficients involving H(E), NH_4(E), Pi(E), and SO_4(E) are independent of *PO* and k, while all other combinations can be used for parameter estimation. This is to be expected because the conversion rates of H(E), NH_4(E), Pi(E), and SO_4(E) are uniquely related to biomass on the basis of charge and, N, P, and S elemental composition.

Biomass yields on substrate and oxygen were chosen for the parameter estimation in this work. Substrate is defined as the sum of glucose and ethanol on a C-mole basis. The carbon fraction of glucose is given by the variable f_{GLU}. Defined in this way, the biomass yield on substrate from the metabolic network MNM III is found to be

$$Y_{sx}^{III} = \frac{(6PO - 1)(1 - f_{GLU}) + (4PO + 2)f_{GLU}}{4.227PO + 2k + 4.221} \quad (11)$$

and for the biomass yield on oxygen, one finds

$$Y_{ox}^{III} = \frac{(6PO - 1)(1 - f_{GLU}) + (4PO + 2)f_{GLU}}{(7.389 + 3k)(1 - f_{GLU}) + (2.108 + 2k)f_{GLU}} \quad (12)$$

For the other networks (MNM I, II, and IV), only the coefficients in these relations differ (Table 2). Note that an increase in k always results in a decreased biomass yield and that an increase in *PO* may compensate for this.

In the MNM, three critical reaction combinations have been identified. The combinations $r_{10} \rightarrow r_{32}/r_{33}$, $r_{12} \rightarrow r_{16}$, and $r_3 \rightarrow r_4$ each represent a switching point between the four networks constructed. The critical glucose fractions ($f_{X\rightarrow Y}^{switch}$) at which this switching occurs can be determined from the analytical result of the corresponding reaction rates. Taking again the example of the irreversible reaction catalyzed by phosphofructokinase (r_3) (eq 10), one can calculate the critical glucose fraction at

Table 2. Biomass Yield on Substrate and Oxygen for the Four Metabolic Networks Involved in the Description of *S. cerevisiae* Metabolism on Glucose/Ethanol Mixtures

network	Y_{sx} (C-mol/C-mol)	Y_{ox} (C-mol/mol)
MNM I	$\frac{(6PO-1)(1-f_{GLU})+\left(PO+\frac{4}{3}\right)f_{GLU}}{4.227PO+2k+3.681}$	$\frac{(6PO-1)(1-f_{GLU})+\left(PO+\frac{4}{3}\right)f_{GLU}}{(6.578+3k)(1-f_{GLU})+(2.272+2k)f_{GLU}}$
MNM II	$\frac{(6PO-1)(1-f_{GLU})+\left(PO+\frac{4}{3}\right)f_{GLU}}{4.227PO+2k+3.681}$	$\frac{(6PO-1)(1-f_{GLU})+\left(PO+\frac{4}{3}\right)f_{GLU}}{(6.578+3k)(1-f_{GLU})+(2.272+2k)f_{GLU}}$
MNM III	$\frac{(6PO-1)(1-f_{GLU})+(4PO+2)f_{GLU}}{4.227PO+2k+4.221}$	$\frac{(6PO-1)(1-f_{GLU})+(4PO+2)f_{GLU}}{(7.389+3k)(1-f_{GLU})+(2.108+2k)f_{GLU}}$
MNM IV	$\frac{(6PO-1)(1-f_{GLU})+\left(4PO+\frac{7}{3}\right)f_{GLU}}{4.227PO+2k+4.459}$	$\frac{(6PO-1)(1-f_{GLU})+\left(4PO+\frac{7}{3}\right)f_{GLU}}{(7.745+3k)(1-f_{GLU})+(1.996+2k)f_{GLU}}$

which the flux through this reaction is zero, i.e. $v_3 = 0$, given by eq 10:

$$f^{\text{switch}}_{\text{III}\rightarrow\text{IV}} = \frac{0.119(6PO-1)}{0.944PO+0.333k+0.347} \qquad (13)$$

It is evident that this critical feed composition is dependent on PO and k. Similar calculations for the other reactions lead to the ($f^{\text{switch}}_{X\rightarrow Y}$) values summarized in Table 3.

The dependence of the switch f_{GLU} fractions on PO and k complicates the estimation of these parameters. Indeed it means that it is necessary to evaluate for each PO,k combination proposed by the estimation algorithm, which experimental data are to be compared to the predictions by MNM I, II, III, and IV, respectively.

Evaluation of the Experimental Data Set. The experimental data set available for calibration of the metabolic network model consisted of conversion rates of biomass (cell dry weight), oxygen, ethanol, and glucose for 11 glucose fractions tested (de Jong-Gubbels et al., 1995). Before the estimation of the remaining unknown parameters PO and k in the MNM was initiated, the available experimental evidence was carefully analyzed with respect to consistency.

First, the carbon balance was checked. Typically about 4 mM of TOC was measured in the supernatant of the cultures. The residual concentrations of glucose and ethanol were below the detection limits of the analytical methods used, and no low-molecular-weight metabolites could be detected by HPLC. Probably, some high-molecular-weight compounds (e.g., extracellular protein) may have been present. Carbon recoveries were 97−102%.

Second, as the number of measured flows is such that the elemental balances present an overdetermined set of linear equations, the conservation laws can be applied to the data to increase the accuracy and credibility of the estimates of the conversion rates. This technique is called balancing or data reconciliation (van der Heijden et al., 1994). The necessary adjustments to the measured rates are related to the standard deviations of the measurements and the structure of the set of equations. The raw and reconciled data are summarized in Table 4 (see supporting information). Standard deviations for the raw data were obtained from repeat measurements and propagation of measurement variances. The chi-squared consistency test shows that all adjustments made were acceptable with respect to the observed measurement errors (Table 4).

Estimation of PO and k. With these reconciled data, the yield coefficients were calculated for the 11 mixtures of ethanol and glucose, evaluated, and subsequently used to estimate PO and k. A multiresponse sum of squared errors criterion was used:

$$J = \sum_{i=1}^{11} w_s(\hat{Y}_{sx}(i) - Y_{sx}(i))^2 + \sum_{i=1}^{11} w_o(\hat{Y}_{ox}(i) - Y_{ox}(i))^2 \qquad (14)$$

where the weights w_o and w_s were taken as approximately inversely proportional to the residual errors, i.e. $w_o = 1$ and $w_s = 10$.

Figure 3 shows the result of the parameter estimation. A satisfactory fit of the model to the data can be observed. In addition, approximate confidence regions ($\alpha = 0.05$) on the MNM predictions are indicated by dashed lines. These confidence regions were calculated as follows. First, a systematic exploration of the objective functional J for an extensive number of parameter combinations was performed. Figure 4a shows a contour plot of the objective functional J as calculated for a grid of 41 × 41 combinations of PO and k. One observes that the minimal value of the objective functional ($J_{opt} = 0.0189$) is found at a value of $PO = 1.090$ and $k = 0.415$. In Figure 4b sections through the objective functional are given for three k values in the neighborhood of the parameter estimates. One observes a sharp minimum in J as a function of PO for each particular k value, while the effect of k on the minimum value of the objective functional is somewhat less pronounced.

The $(1-\alpha)$ confidence region of the parameters consists of the set of parameter combinations, resulting in an objective functional less than the threshold value (Beale, 1960):

$$J_{\text{opt}}\left(1+\frac{p}{N-p}F_{\alpha,p,N-p}\right) \qquad (15)$$

where N and p are the number of measurements ($N = 22$) and parameters ($p = 2$), respectively, and $F_{\alpha,p,N-p}$ is the value of the F distribution with p and $N-p$ degrees of freedom and a confidence level α. This region is situated for values of PO between 1.07 and 1.11 and for k between 0.385 and 0.445.

Subsequently, biomass yields were predicted with an extensive set of parameter combinations taken from this 95% confidence region. These predictions allowed construction of the confidence regions of the model as given in Figure 3.

A Practical and Theoretical Check on Parameter Identifiability. The valley-like form of the objective functional gives rise to some concern and warrants additional attention because it may be indicative of problems with the estimation of PO and k. Indeed, in the range of $k = 0.385-0.445$ and $PO = 1.07-1.11$, it

Table 3. Switching Points for Energetic Parameters *PO* and *k* and Their Sensitivity to These Parameter Values[a]

switchpoint	$f_{X\to Y}^{switch}$ (C-mol of glucose/C-mol of carbon source)	$PO\,\partial f_{X\to Y}^{switch}/f_{X\to Y}^{switch}\,\partial PO$	$k\,\partial f_{X\to Y}^{switch}/f_{X\to Y}^{switch}\,\partial k$
I → II	$\frac{0.148(6PO-1)}{0.944PO+0.307k+0.218}$ (0.597)	(0.433)	(−0.092)
II → III	$\frac{0.131(6PO-1)}{0.944PO+0.323k+0.289}$ (0.499)	(0.472)	(−0.092)
III → IV	$\frac{0.119(6PO-1)}{0.944PO+0.333k+0.347}$ (0.434)	(0.502)	(−0.092)

[a] Values for the optimal values of *PO* (1.090) and *k* (0.415) are given in parentheses.

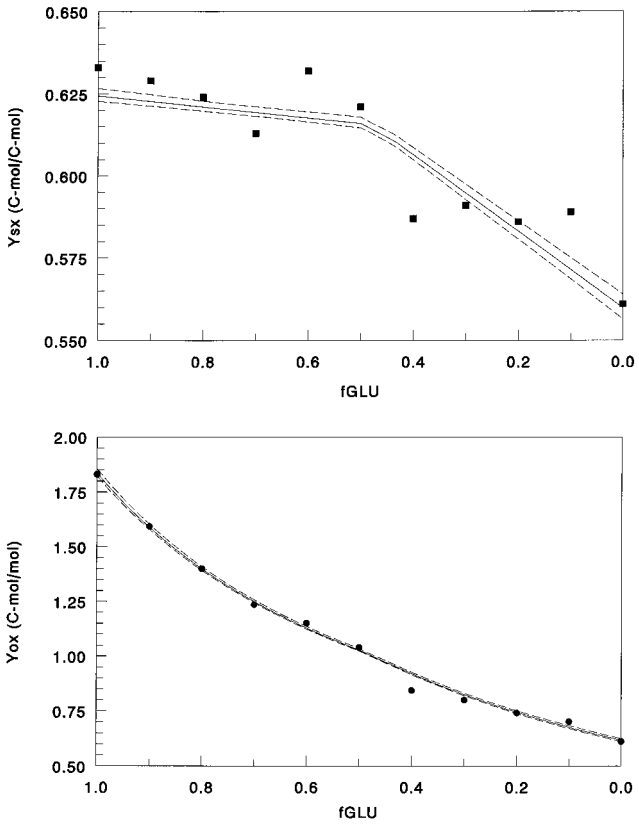

Figure 3. Reconciled biomass yield data and model predictions with error bounds. (a) biomass yield on substrate and (b) biomass yield on oxygen.

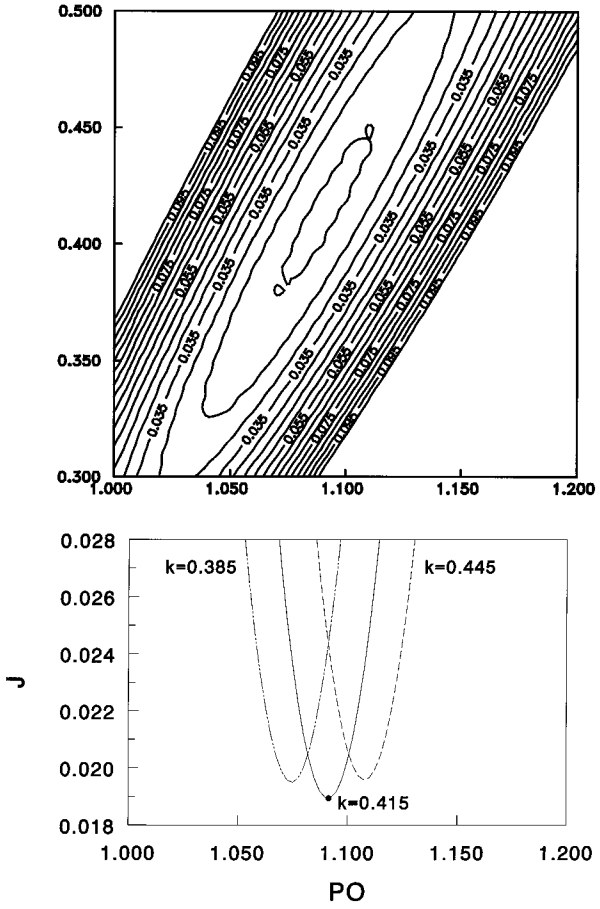

Figure 4. Multiresponse objective functional: (a) contour plot of *J* as function of *PO* and *k* and (b) sections across objective functional for different *k* values.

can be observed that parameter combinations satisfying

$$k - 0.415 = 1.77(PO - 1.090) \qquad (16)$$

will result in objective functional values that differ little from the minimal value. This makes it difficult to attribute unique values to *PO* and *k*. This problem is also known as the practical identifiability problem (Munack, 1991).

To make a theoretical decision on the identifiability problem, a standard technique consists of evaluating the output sensitivity functions, i.e. $\partial Y_{sx}/\partial PO$ and $\partial Y_{sx}/\partial k$, and similarly for the biomass yields on oxygen. Proportionality of these functions is a proof of the theoretical nonidentifiability of the considered parameters (Munack, 1991).

The output sensitivity functions of Y_{sx} for the parameters *k* and *PO* have been derived analytically for MNM III from (11):

$$\frac{\partial Y_{sx}^{III}}{\partial k} = -2\frac{(6-2f_{GLU})PO + 3f_{GLU} - 1}{(4.227PO + 2k + 4.221)^2} \qquad (17)$$

$$\frac{\partial Y_{sx}^{III}}{\partial PO} = 2.114\frac{\partial Y_{sx}}{\partial k} + \frac{6 = 2f_{GLU}}{4.227PO + 2k + 4.221} \qquad (18)$$

This result clearly shows the nonproportionality of the output sensitivity functions. Hence, it can be concluded that *PO* and *k* are theoretically identifiable from measurements of Y_{sx} at different substrate mixtures.

Similarly, the output sensitivities for Y_{ox} are deduced from (12):

$$\frac{\partial Y_{ox}^{III}}{\partial k} = -\frac{((6PO-1)(1-f_{GLU}) + (4PO+2)f_{GLU})(3-f_{GLU})}{((7.389+3k)(1-f_{GLU}) + (2.108+2k)f_{GLU})^2} \qquad (19)$$

$$\frac{\partial Y_{ox}^{III}}{\partial PO} = \frac{6-2f_{GLU}}{(7.389+3k)(1-f_{GLU}) + (2.108+2k)f_{GLU}} \qquad (20)$$

leading to the same conclusion.

We want to stress here that this theoretical exercise should be performed before parameter estimation starts and that the application of symbolic manipulation software is very useful.

Validation and Predictive Capacity of the MNM. Two approaches were used to validate the metabolic network model of *S. cerevisiae*. The first consists of confronting data and model predictions for the biomass yield on a different substrate than the ones used for calibration of the model. Second, predictions of the flux through identified critical reactions are compared with measured levels of the enzymes catalyzing these reactions.

Validation Based on Biomass on Acetic Acid Yield. *S. cerevisiae* can use glucose, ethanol, and acetic acid as a single carbon source. Verduyn (1991), using a closely related strain and the same cultivation conditions as this work, reported a biomass yield on acetic acid of 0.29 g/g or 0.34 C-mol/C-mol. A metabolic network was derived from the glucose/ethanol network valid for the lowest glucose fraction (MNM IV), by omission of r_{13} and r_{43} and setting $f_{GLU} = 0$. Except for the transport of acetic acid into the cell, this network is complete for prediction of metabolism in acetic acid cultivations. It is as yet not established whether acetic acid transport in *S. cerevisiae* is energy-dependent. Hence, two possibilities were evaluated. First, a proton symport with a H^+/acetic acid stoichiometry of 1 was implemented. Alternatively, passive or facilitated diffusion was incorporated in the metabolic network.

With the values for *PO* and *k* estimated from the glucose/ethanol experiments, biomass yields were calculated for both hypothetical acetic acid networks. For the network with facilitated diffusion, a yield coefficient of 0.44 C-mol/C-mol was predicted, while a biomass yield of 0.34 was calculated for the network in which a proton symport was incorporated. One observes an excellent agreement between the data and the network in which transport of acetic acid into the cell requires energy. Similarly the yields on oxygen, measured as 0.52 C-mol/mol, were found to be 0.82 and 0.53 C-mol/mol for the networks without and with transport-ATP demands. Hence, these results not only validate the metabolic network but also provide additional evidence for the existence of a proton symport for acetic acid. This is consistent with the results of Leao et al. (1986) and Cassio et al. (1987).

Validation Based on Metabolic Flux Analysis and in Vitro Enzyme Assays. Generally speaking, two control systems for enzyme activity are present in the cell. The first, operating at the level of enzyme synthesis is based on (1) the modulation of DNA expression into mRNA and translation of the latter in enzyme, (2) regulation of enzyme turnover, and (3) irreversible inactivation of the enzyme. The second control system acts directly on the enzyme activity by changing the kinetic properties of the biocatalysts, e.g. by allosteric modification or phosphorylation.

Enzyme levels (expressed as in vitro activities) were available for nine enzymes involved in the metabolism of glucose and ethanol: r_1, r_3, r_4, r_8, r_{12}, r_{16}, r_{26}, r_{32}, and r_{33} (de Jong-Gubbels et al., 1995). From the onset of the analysis, it must be stressed that these measurements may only be interpreted with respect to regulation at the level of transcription/translation since allosteric mechanisms cannot be assessed by in vitro techniques. Hence, only if an enzyme is absent where the model predicts activity may the validity of the model be questioned. Even then, the possibility still remains that isoenzymes exist in the cell that were not measured. Still one can confront the predicted fluxes with the measured enzyme levels and in this way obtain indications about the actual in vivo activity of these enzymes, pointing to the method of regulation.

A second precaution to take before conclusions are drawn from this type of metabolic network validation is due to the uncertainty of the parameters. Because the development of the metabolic network involves estimation of parameters on the basis of experimental data (see above), the inherent uncertainty of the measurements is propagated in the parameter estimates. The resulting model uncertainty can be evaluated, however, by checking the sensitivity of the predicted pathway rates and the switching points to the parameter values. If the switching points vary a lot with small changes in *PO* or *k*, it means that these switching points can be situated at almost any f_{GLU} without affecting the model fit to the yield data (see above). Hence, only if the sensitivities are sufficiently low can reliable conclusions be drawn on the validity of the network model structure.

Sensitivity of Switching Points. For the three switching point functions of Table 3, one can readily calculate the scaled sensitivity functions (we used the symbolic software tool again to minimize "human error"). For reaction r_3 (eq 13) one obtains

$$\frac{k\,\partial f^{\text{switch}}_{\text{III}\to\text{IV}}}{f^{\text{switch}}_{\text{III}\to\text{IV}}\,\partial k} = -0.333\frac{k}{(0.944PO + 0.333k + 0.347)} \quad (21)$$

$$\frac{PO\,\partial f^{\text{switch}}_{\text{III}\to\text{IV}}}{f^{\text{switch}}_{\text{III}\to\text{IV}}\,\partial PO} = \frac{2.835PO}{k}\frac{k\,\partial f^{\text{switch}}_{\text{III}\to\text{IV}}}{f^{\text{switch}}_{\text{III}\to\text{IV}}\,\partial k} + \frac{6PO}{6PO - 1} \quad (22)$$

It means that with an increased *k* value more ethanol will be converted before reaction r_3 tends to reverse and has to be replaced by the gluconeogenic analogon. Similarly, a higher efficiency of the oxidative phosphorylation will result in an opposite change of the switch glucose fraction at which reaction reversal would occur. In Table 3, numerical values are given for this scaled sensitivity function for the optimal parameter estimates obtained above. The sensitivity of the switch substrate mixtures for the other reactions are given as well. Similar conclusions as for $f^{\text{switch}}_{\text{III}\to\text{IV}}$ can be made. Note that the effect of changing *PO* and *k* values is rather low. For a relative change in the parameter value with 10%, a relative shift of the critical glucose fraction must be expected with only 5% or 1% for *PO* and *k*, respectively. Hence, the critical switching fractions are rather strictly controlled by the value of *PO* and *k*. This result implies that the experimental enzyme activity results may be interpreted to validate the metabolic network. Again, we want to stress the importance of this preliminary analysis for the validation task at hand.

Experimental Results. The predictions of the f^{switch} values, together with the pathway fluxes that can be calculated from the MNM solutions (e.g., eq 10), will now be compared to the experimental enzyme activity determinations (de Jong-Gubbels et al., 1995).

For the first switch in metabolism (from MNM I to MNM II), Figure 5a shows the measured activities of the enzymes involved in the glyoxylate shunt (r_{32}/r_{33}) and the corresponding flux predictions. From the stoichiometry of reactions r_{32} and r_{33}, it is evident that the metabolic network predicts identical reaction rates. The expression of isocitrate lyase (r_{32}) is initiated at lower ethanol conversions than predicted by the network. Assuming the metabolic network is correct, this may indicate that the synthesis of this enzyme is not regulated very strictly. Comparingly, malate synthase seems more tightly regu-

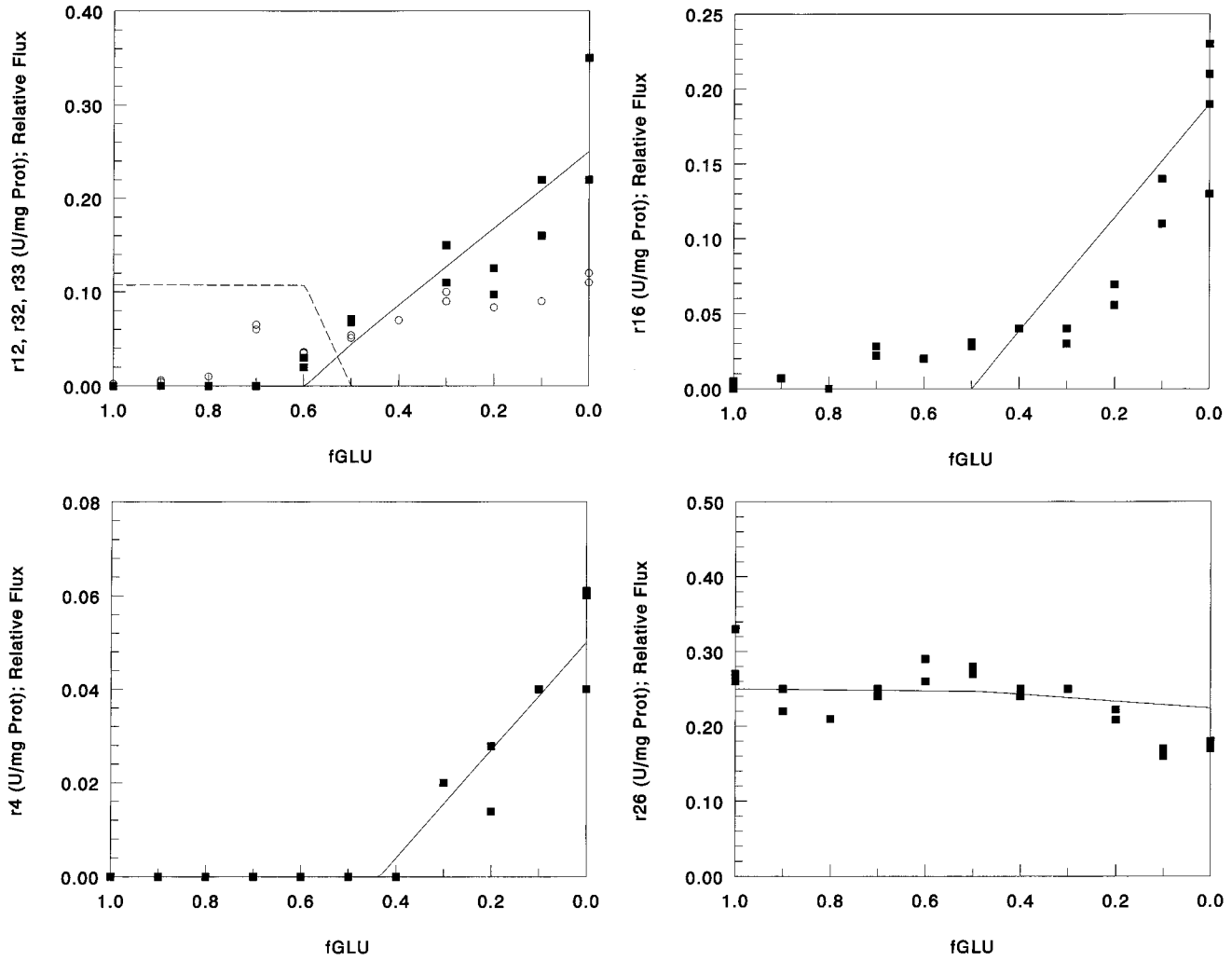

Figure 5. Measured enzyme activities and flux predictions from the metabolic network for (a) dashed line, pyruvate carboxylase (r_{12}); full line, glyoxylate shunt (r_{32}/r_{33}); isocitrate lyase (r_{32}, open symbols); malate synthase (r_{33}, closed symbols); (b) PEP carboxykinase (r_{16}); (c) fructose-1,6-bisphosphatase (r_4); and (d) glucose-6-P dehydrogenase (r_{26}).

lated and is only appearing at $f_{GLU} = 0.6$, which corresponds very well to $f_{I\rightarrow II}^{switch} = 0.597$ obtained from the network (Table 3, Figure 5a). Moreover, the measured enzyme activities are proportional to the required reaction rates, indicating that regulation seems to occur at the enzyme expression level. A different conclusion must be made for isocitrate lyase.

No enzyme activities were measured for reaction r_{10} (it is difficult to accurately measure activities of the pyruvate dehydrogenase complex).

While the model predicts that the reaction rate of the alternative glucose-related anaplerotic pathway (r_{12}, pyruvate carboxylase) drops quickly to zero between $f_{GLU} = 0.597$ and $f_{GLU} = 0.499$, the measured enzyme levels remained unchanged (data not shown to retain clarity of Figure 5a; the experimental results can be found in de Jong-Gubbels et al., 1995). Hence, no expression level regulation seems active for this reaction, and control based on enzymic modification must play a role. Furthermore, the measured pyruvate carboxylase activities remained at a constant level even at lower glucose feeds and down to growth on ethanol only. Another remarkable result is that it is predicted that both anaplerotic reactions coexist in the cell (between $f_{GLU} = 0.597$ and $f_{GLU} = 0.499$, Figure 5a), a result that is not contradicted by experimental evidence.

At glucose fractions lower than $f_{II\rightarrow III}^{switch} = 0.499$, the model predicts that the gluconeogenic enzyme catalyzing reaction r_{16} (PEP carboxykinase) must be present, and this is confirmed by the experimental results (Figure 5b). However, the simultaneous presence of both PEP carboxykinase and pyruvate carboxylase could lead to a futile cycle ($r_{16}-r_8-r_{12}$), and this is probably prohibited in vivo by allosteric regulation mechanisms. Equivalently, in the network, reaction r_{12} is eliminated for f_{GLU} lower than 0.499. One should note that PEP carboxykinase was also present at significant levels under cultivation conditions where the glyoxylate shunt was not yet active, indicating either another role for the enzyme than gluconeogenesis or the presence of regulation based on enzyme modification. Again, predicted reaction rates corresponded fairly well with the measured activities in the f_{GLU} range (0.3−0.0). Activities of pyruvate kinase (r_8) showed a linear decrease to 50% of the maximum level as the ethanol fraction in the feed increased (see also de Jong-Gubbels et al., 1995). The model predictions did not correspond to this behavior. Under the assumption that the proposed model describes cellular metabolism, allosteric modification of the synthesized enzymes must be acting to adjust the corresponding flux.

The switch from MNM III to IV involves phosphofructokinase (r_3) and fructose-1,6-bisphosphatase (r_4). It is predicted to occur at a critical glucose fraction of 0.434. The enzyme activities of FBPase summarized in Figure 5c are in excellent agreement with this prediction. Moreover, as in the case of malate synthase (r_{33}), the

reaction rate is proportional to the enzyme activity measured, indicating control at the expression level under the conditions tested.

It is well-known that, in addition to regulation at the level of enzyme synthesis, the activity of the enzymes of the glyoxylate cycle and gluconeogenesis may be influenced by posttranslational modification and allosteric regulatory mechanisms. However, the examples discussed above show that, under the carbon-limited steady-state conditions investigated in this study, regulation occurs mainly at the level of enzyme synthesis. For the enzymes which are not controlled at the expression level, additional regulation systems must be active, like allosteric systems or kinases. In this respect it is striking to have a closer look at an enzyme involved in glycolysis, phosphofructokinase (r_3). While the carbon flux through this reaction alters drastically as the ethanol fraction in the feed increases, constant enzyme levels are found through the whole range of glucose/ethanol mixtures tested (data not shown, see de Jong-Gubbels et al., 1995). From this result and the evidence reported on the other reactions, one could conclude that the enzymes involved in gluconeogenesis are more tightly controlled at the transcription/translation expression level than the glycolytic enzymes which seem largely constitutive for the conditions tested.

Finally, it is interesting to pay some attention to the results of the flux analysis of reaction r_{26} catalyzed by glucose-6-P dehydrogenase. Figure 5d illustrates that the synthesis of this enzyme is rather constant for all feed compositions tested. Although this does not necessarily mean that the fluxes are actually constant, it is interesting to observe that the predicted reaction rates (Figure 5d) correspond to the enzyme levels found.

Discussion

A metabolic network was constructed for growth of *S. cerevisiae* on mixtures of glucose and ethanol. The main problem that had to be dealt with after inclusion of all biochemical a priori knowledge (in the form of the reaction network) was the estimation of two unknown energetic coefficients: (1) the operational *PO* ratio describing the efficiency of ATP formation in the electron transport chain and (2) a growth-related maintenance factor *k* which in this work was a fixed amount of ATP dissipated per unit of biomass formed.

In many applications, this estimation problem is circumvented (1) by accepting a (most often theoretical and not operational) *PO* ratio from the literature (Aiba and Matsuoko, 1979; Majewski and Domach, 1990; Vallino and Stephanopoulos, 1990, 1993; Goel et al., 1993; Varma and Palsson, 1995) and (2) by applying a two-parameter (growth-associated and non-growth-associated) maintenance reaction which provides sufficient parameter flexibility to the reaction network to produce acceptable model predictions (Varma et al., 1993; Vallino and Stephanopoulos, 1993; Varma and Palsson, 1995). Since most applications of the metabolic modeling methodology have been theoretical in nature (Majewski and Domach, 1990; Vallino and Stephanopoulos, 1993; Varma et al., 1993; Varma and Palsson, 1994a), the present models are not sufficiently validated with "rich" experimental data and therefore possess limited predictive capacities.

Tackling this estimation problem by experimental methods was one of the main goals of our work. Van Gulik and Heijnen (1995) have presented a first approach to estimate these energetic parameters in metabolic networks of *S. cerevisiae* and *Candida utilis*: it consists of presenting a multitude of substrates one-by-one to the organism. From the yields, it was shown to be possible to obtain unique values for the unknown parameters *PO* and *k*.

In the same study it proved also possible in the case of lysine production by *C. glutamicum* to use the changing biomass and product yield data obtained in a chemostat operated at a series of dilution rates for estimation of the operational *PO* ratio and maintenance coefficient.

Our work evaluated whether yield data on a number of different mixtures of substrates provided sufficiently rich data for estimation of the unknowns. The approach taken was more rigorous compared to the previous work as the identifiability properties of the estimation problem were thoroughly tested, both theoretically and practically.

For the theoretical identifiability check, the usefulness of symbolic manipulation software was clearly shown as it eased the derivation of the proof of nonproportionality of the sensitivity functions. Its application has been advocated previously for theoretical identifiability studies of such large systems for it prevents human error (Raksanyi et al., 1985; Posten and Munack, 1990).

In addition, this software also helped reducing the necessary computations for the parameter estimation as the four metabolic networks operational over the whole range of substrate mixtures do not have to be solved for each new (*PO*, *k*) parameter set tested.

Finally, the sensitivity functions which were calculated during the theoretical identifiability study generated a lot of insight on the dependence of reaction and conversion rates on the energetic properties of the cell. Also, this clearly showed the interrelation between the two parameters, pointing to the multivariable nature of this parameter estimation problem. It must be stressed here that one-dimensional sensitivity analysis, as performed in many instances, may shed an illusion of identifiability since the multivariable character is not addressed.

For the practical identifiability, the numerical properties of the nonlinear parameter estimation problem were evaluated. It was observed that the objective functional shape was sufficiently conditioned for reliable parameter estimation. Using a multiresponse objective functional involving biomass yields both on substrate and on oxygen and the postulated metabolic network, operational *PO* and *k* values were obtained together with their 95% confidence intervals ($1.07 < PO < 1.11$ mol of ATP/mol of O; $0.385 < k < 0.445$ mol of ATP/C-mol of biomass).

The *PO* value obtained is substantially lower than the overall mechanistic *PO* ratio for *S. cerevisiae* of 2 (Verduyn et al., 1991). In their discussion on the subject, Verduyn et al. (1991) mention, next to their own, several other approaches used in the past to describe this discrepancy (von Meyenburg, 1969; Harder and van Dijken, 1976; Harder et al., 1981) and conclude that an effective *PO* ratio for baker's yeast is expected to be close to 1.

The value of 1.20 found by van Gulik and Heijnen (1995) is somewhat further distant than the one estimated here, but this can be explained by the approach taken: They attempted to combine the yield data of both *C. utilis* and *S. cerevisiae* to obtain two *PO* values but a common *k* value for the respective yeasts. Hence, compromise estimates may have been obtained which may be different from the best estimates for the separate yeasts. Comforting, however, is the finding that the (*PO*, *k*) parameter set found in their work (1.20, 0.644) is lying in the objective functional valley obtained in our work (see Figure 4) and is almost fulfilling equation 16.

With respect to the *k* value, it is good to recall the two concepts found in the literature to accommodate for the

Table 5. Comparison of *PO* and *k* Values

organism	*PO* estd (mol of ATP/ mol of O)	*PO* assumed (mol of ATP/ mol of O)	*k*(biomass) (mol of ATP/ C-mol of biomass) at $D = 0.1\ h^{-1}$	*k*(protein) (mol of ATP/C-mol of protein) at $D = 0.1\ h^{-1}$	ref
S. cerevisiae (aerobic)	1.09		0.415	0.88	this work
S. cerevisiae (aerobic)	1.20		0.644	1.37	van Gulik and Heijnen (1995)
C. utilis	1.53		0.712	1.37	van Gulik and Heijnen (1995)
S. cerevisiae (anaerobic)			0.675	1.43	Verduyn et al. (1991)
E. coli		1.30	2.04	2.91[a]	Varma et al. (1993)
E. coli		1.30	2.22	3.17[a]	Varma and Palsson (1995)
P. chrysogenum		2.60	1.27	2.31	Jorgensen et al. (1995)

[a] Assuming 70% protein.

difference in theoretical and experimental ATP requirements for growth (Verduyn et al., 1991): In the first concept, it is assumed that a fixed amount *k* of ATP is dissipated and this is, consequently, independent of the growth substrate. An alternative concept is that a fixed fraction of all produced ATP, which will thus depend on the substrate used, is no longer available for biomass synthesis but is used in such processes as maintenance of gradients, futile cycles, proofreading in protein, and RNA/DNA synthesis. Verduyn et al. (1991) and van Gulik and Heijnen (1995) collected ample evidence that the first concept is the more likely because the alternative leads to very different *PO* values for the same organism growing on different substrates, which is unlikely.

In our metabolic network model, the approach of a fixed amount of *k* ATP/C-mol of biomass was adopted. The obtained *k* value can be compared to the *k* values reported in the literature (Table 5). The higher *k* values for the non-yeast growth systems can partially be explained by the lower maintenance requirements of yeasts compared to other organisms (Verduyn et al., 1991). More important, however, may be that the *PO* ratios used in the other studies were not fitted to experimental data and were probably too high. As mentioned previously, this would result in a too efficient ATP generation and thus, to account for the excess ATP, a substantially higher *k* value is calculated. Finally, also the protein content of the cell may play a role in the *k* value (see below). Indeed, van Gulik and Heijnen (1995) express the *k* value per C-mol of protein to account for a substantial difference in protein content between *C. utilis* and *S. cerevisiae*. In this manner a single maintenance factor could be used for both organisms. For the literature data presented in Table 5, protein contents are indeed significantly (up to 40%) higher than the 42% protein of *S. cerevisiae*.

In this context, the question can be addressed as to what the effect on the *k* factor would be if six ATPs are used for elongation of the peptide chain instead of the four included in our metabolic network. Indeed, this stoichiometry is not established completely yet. One observes in the reaction network used (see the Appendix) that the ATP conversion for protein elongation and biomass formation reactions (r_{85} and r_{99}) are exchangeable. It can be calculated that an increase to six ATPs per peptide bond would decrease *k* with 0.25 mol of ATP/C-mol of biomass. Taking the uncertainty of assumed *PO* values into account, it appears that a *k* value of approximately 1 mol of ATP/C-mol of biomass is a reasonable value. However, a more extensive analysis of growth yields for different types of organisms giving independent values of *PO* and *k* is required to establish a reliable value for *k*.

An attempt was made to validate the fully calibrated metabolic network model. Special attention was focused on the use of rather easy to measure variables of the cell's metabolism (yields and in vitro enzyme activities) to ensure ease of application for other biological reaction systems. To the authors' knowledge this exercise is the first of its kind for metabolic networks.

As a first validation test, a new substrate (acetate) was administered to *S. cerevisiae* and the observed yield corresponded excellently with the model prediction. As a spinoff, new evidence was supplied that acetate uptake requires an active transport system, indicating again that metabolic network analysis is an excellent tool for increasing insight in the cell's metabolism.

The second validation method was based on a more biochemical evaluation of the model's predictive capabilities. In the applied test, a number of measured activities of enzymes of central metabolism were compared with predicted reaction rates. None of the enzymes were absent when its corresponding reaction was predicted to be active in the cell. Moreover, the onset of expression of a number of key enzymes was found to coincide very well with the switching between different networks as the glucose fraction in the feed decreased. This result is a very powerful validation of the metabolic network.

It was also observed that for a number of enzyme-catalyzed reactions (especially gluconeogenesis) enzyme levels were proportional to the predicted reaction rates, while for others the in vitro activities diverged from the model responses. This information was interpreted in the frame of enzyme regulation, i.e. the former enzymes are subjected to regulation at the expression level whereas the latter's activity is regulated by enzymic modification.

Finally, in the work of de Jong-Gubbels et al. (1995), switching of the metabolic network reactions was predicted on the basis of an analysis of carbon fluxes only (Table 6). The three switching f_{GLU} fractions are clearly lower than the fractions calculated here with the overall metabolic network where cofactors are also taken into account. For the glucose fraction for which all acetyl-CoA is produced from ethanol ($f_{I\rightarrow II}^{switch}$), the predictions correspond fairly well. For the gluconeogenic pathways, however, the initiation of PEP carboxykinase (r_{16}) and FBPase activity (r_4) based on the method of carbon-flux analysis predicts a much higher critical glucose fraction than the one given by the overall network, taking carbon and cofactors into account. However, given the experimental error involved in enzyme activity measurements, the evidence is not totally conclusive to decide on one or the other approach as the better one. Still, the overall metabolic network approach seems more appropriate.

The results suggest that the cofactor balances, especially NAD(P)H, may play an important role in flux distribution. The importance of the bifunctional nature of the pentose phosphate cycle is highlighted in this system as the flux through this pathway is determined by the need for biosynthetic reducing power, apparently exceeding the cell requirement for C_4 and C_5 units.

Table 6. Comparison of Switching f_{GLU} Fractions Based on an Analysis Of Carbon Balances Only Versus an Analysis Including Cofactors

switch-point	carbon balancing (C-mol of glucose/ C-mol of carbon source)	carbon and cofactor balancing (C-mol of glucose/ C-mol of carbon source)
I → II	0.57	0.597
II → III	0.29	0.499
III → IV	0.26	0.434

Concluding Remarks

This work showed that yield data for a number of mixtures of two substrates provide sufficiently rich information for unique estimation of the two energetic parameters *PO* and *k*, the only unknowns remaining in a proposed stoichiometric description of *S. cerevisiae* metabolism.

Symbolic manipulation was very helpful in reducing the computing time for parameter estimation, the evaluation of the theoretical identifiability of the estimation problem, and the generation of more insight in the dependency of the reaction and conversion rates on these energetic parameters.

The validation results—both the prediction of the biomass yield on acetic acid and the predictions of enzymatic activities and switching points—provide extensive evidence that the proposed metabolic network is a good representation of the metabolism of *S. cerevisiae*. The extrapolative power of the network seems considerable and may lead to the generation of interesting hypotheses on the physiology of *S. cerevisiae*, the design of optimal substrate feeding strategies, or the identification of metabolic bottlenecks to be expected under certain process intensifications.

Notation

f_{GLU}	glucose fraction in the feed (C-mol/C-mol)
$f_{III \rightarrow IV}^{switch}$	f_{GLU} where switch in metabolism occurs (C-mol/C-mol)
J	objective function (weighted sum of squared errors)
J_{opt}	optimal value of J
k	maintenance factor (mol of ATP/C-mol of biomass)
MNM(I, II, III, IV)	metabolic network models
PO	effective P/O ratio (mol of ATP/mol of O)
r_i	reaction i
v_i	rate of reaction r_i (mol/h)
Y_{sx}	biomass yield on substrate (C-mol/C-mol)
Y_{ox}	biomass yield on oxygen (C-mol/mol)

Acknowledgment

Financial support for this work was partly provided by the Belgian National Fund for Scientific Research (N.F.W.O.) through a research grant.

Supporting Information Available: Table 1 (biomass composition) and Table 4 (conversion rate data) (4 pages). Ordering information is given on any current masthead page.

Appendix

Vector of considered Metabolites

(1)	AC	Acetate
(2)	ACCOA	Acetyl coenzyme A
(3)	ACET	Acetaldehyde
(4)	ADP	Adenosine-5'-diphosphate
(5)	AKG	a-Ketoglutarate
(6)	AKI	a-Ketoisovalerate
(7)	ALA	Alanine
(8)	AM	Amino acids
(9)	AMP	Adenosine-5'-monophosphate
(10)	ARG	Arginine
(11)	ASN	Asparagine
(12)	ASP	Aspartate
(13)	ATP	Adenosine-5'-triphosphate
(14)	BIOM	Biomass
(15)	CARP	Carbamyl phosphate
(16)	CHO	Chorismate
(17)	CIT	Citrate
(18)	CMP	Cytidine-5'-monophosphate
(19)	CO2	Carbon dioxide
(20)	COA	Coenzyme A
(21)	CTP	Cytidine-5'-triphosphate
(22)	CYS	Cysteine
(23)	E4P	Erythrose-4-phosphate
(24)	ETOH	Ethanol
(25)	FAD	Flavin adenine dinucleotide
(26)	FADH2	Flavin adenine dinucleotide, reduced
(27)	FTHF	Formyl-tetrahydrofolate
(28)	FRUC6P	Fructose-6-phosphate
(29)	FUM	Fumarate
(30)	G3P	3-Phosphoglycerate
(31)	GAP	Glyceraldehyde-3-phosphate
(32)	GLUC6P	Glucose-6-phosphate
(33)	GLUC(E)	Glucose, extracellular
(34)	GLUC	Glucose
(35)	GLUCON	Gluconic acid
(36)	GLUM	Glutamine
(37)	GLUT	Glutamate
(38)	GLY	Glycine
(39)	GLYO	Glyoxylate
(40)	GMP	Guanosine-5'-monophosphate
(41)	GOH	Glycerol
(42)	H2O	Water
(43)	HIS	Histidine
(44)	HOM	Homoserine
(45)	H(E)	Proton, extracellular
(46)	H	Proton
(47)	ILEU	Isoleucine
(48)	IMP	Inosine-5'-monophosphate
(49)	ISOCIT	Isocitrate
(50)	LAC	Lactate
(51)	LEU	Leucine
(52)	LYS	Lysine
(53)	MAL	Malate
(54)	MET	Methionine
(55)	METHF	Methylene tetrahydrofolate
(56)	MYTHF	Methyl tetrahydrofolate
(57)	NAD	Nicotinamide adenine dinucleotide
(58)	NADH	Nicotinamide adenine dinucleotide, reduced
(59)	NADP	Nicotinamide adenine dinucleotide phosphate
(60)	NADPH	Nicotinamide adenine dinucleotide phosphate, reduced
(61)	NH4(E)	Ammonium ion, extracellular
(62)	NH4	Ammonium ion
(63)	O2	Oxygen
(64)	OAC	Oxaloacetate
(65)	OL	Oleate
(66)	PAL	Palmitoleate
(67)	PEP	Phosphoenolpyruvate
(68)	PHEN	Phenylalanine
(69)	Pi	Phosphate ion
(70)	Pi(E)	Phosphate ion, extracellular
(71)	PRO	Proline
(72)	PROT	Protein
(73)	PRPP	5-Phosphoribosyl-1-pyrophosphate
(74)	PSACCH	Polysaccharides
(75)	PYR	Pyruvate
(76)	RIB5P	Ribose-5-phosphate
(77)	RIBU5P	Ribulose-5-phosphate
(78)	RNA	Ribonucleic acid
(79)	SED7P	Sedoheptulose-7-phosphate
(80)	SER	Serine
(81)	SO4(E)	Sulphate ion, extracellular
(82)	SO4	Sulphate ion
(83)	SUC	Succinate
(84)	SUCCOA	Succinyl coenzyme A
(85)	THF	Tetrahydrofolate
(86)	THR	Threonine
(87)	TRYP	Tryptophan
(88)	TYR	Tyrosine
(89)	UMP	Uridine-5'-monophosphate
(90)	UTP	Uridine-5'-triphosphate
(91)	VAL	Valine
(92)	XYL5P	Xylulose-5-phosphate
(93)	AC(E)	Acetate, extracellular
(94)	CIT(E)	Citrate, extracellular
(95)	GLUCON(E)	Gluconic acid, extracellular
(96)	LAC(E)	Lactate, extracellular
(97)	PYR(E)	Pyruvate, extracellular
(98)	SUC(E)	Succinate, extracellular

Library of biochemical reactions (reactions indexed with '#' are omitted in the applied metabolic network)

Glycolysis and citric acid cycle

(1)	1 GLUC + 1 ATP	--> 1 GLUC6P + 1 ADP + 1 H
(2)	1 GLUC6P	<-> 1 FRUC6P
(3)	1 FRUC6P + 1 ATP	--> 2 GAP + 1 ADP + 1 H
(4)	2 GAP + 1 H2O	--> 1 FRUC6P + 1 Pi
(5)	1 GAP + 1 NADH + 1 H2O + 1 H	--> 1 GOH + 1 NAD + 1 Pi
(6)	1 GAP + 1 NAD + 1 Pi + 1 ADP	<-> 1 G3P + 1 ATP + 1 NADH + 1 H
(7)	1 G3P	<-> 1 PEP + 1 H2O
(8)	1 PEP + 1 ADP + 1 H	--> 1 PYR + 1 ATP
(9)#	1 PYR + 2 ATP + 1 H2O	--> 1 PEP + 2 ADP + 1 Pi + 2 H
(10)	1 PYR + 1 NAD + 1 COA	--> 1 ACCOA + 1 NADH + 1 CO2
(11)#	1 PYR + 1 H	--> 1 ACET + 1 CO2
(12)	1 PYR + 1 ATP + 1 H2O + 1 CO2	--> 1 OAC + 1 ADP + 1 Pi + 2 H
(13)	1 ACET + 1 NAD + 1 H2O	<-> 1 AC + 1 NADH + 2 H
(14)#	1 ACET + 1 NADP + 1 H2O	<-> 1 AC + 1 NADPH + 2 H
(15)	1 OAC + 1 ACCOA + 1 H2O	--> 1 ISOCIT + 1 COA + 1 H
(16)	1 OAC + 1 ATP	--> 1 PEP + 1 CO2 + 1 ADP
(17)	1 ISOCIT + 1 NAD	--> 1 AKG + 1 NADH + 1 CO2
(18)#	1 ISOCIT + 1 NADP	--> 1 AKG + 1 NADPH + 1 CO2
(19)	1 AKG + 1 COA + 1 NAD	--> 1 SUCCOA + 1 NADH + 1 CO2
(20)	1 SUCCOA + 1 ADP + 1 Pi	--> 1 SUC + 1 ATP + 1 COA
(21)	1 SUC + 1 FAD	<-> 1 FUM + 1 FADH2
(22)	1 FUM + 1 H2O	<-> 1 MAL
(23)	1 MAL + 1 NAD	--> 1 OAC + 1 NADH + 1 H
(24)#	1 FADH2 + 1 NAD	<-> 1 NADH + 1 H + 1 FAD
(25)#	1 GLUC + 1 PEP	<-> 1 GLUC6P + 1 PYR

Pentose phosphate cycle

(26)	1 GLUC6P + 2 NADP + 1 H2O	--> 1 RIBU5P + 2 NADPH + 1 CO2 + 2 H
(27)	1 RIBU5P	<-> 1 RIB5P
(28)	1 RIBU5P	<-> 1 XYL5P
(29)	1 RIB5P + 1 XYL5P	<-> 1 SED7P + 1 GAP
(30)	1 SED7P + 1 GAP	<-> 1 FRUC6P + 1 E4P
(31)	1 XYL5P + 1 E4P	<-> 1 FRUC6P + 1 GAP

Glyoxylate shunt

(32)	1 ISOCIT	<-> 1 GLYO + 1 SUC
(33)	1 GLYO + 1 ACCOA + 1 H2O	--> 1 MAL + 1 COA + 1 H

Oxidative phosphorylation

(34)	2 NADH + 1 O2 + 2 *PO* ADP + 2 *PO* Pi + (2 + 2 *PO*) H	--> (2 + 2 *PO*) H2O + 2 NAD + 2 *PO* ATP
(35)#	2 NADH + 1 O2 + 2 H	--> 2 NAD + 2 H2O
(36)	2 FADH2 + 1 O2 + 2 *PO* ADP + 2 *PO* Pi + 2 *PO* H	--> (2 + 2 *PO*) H2O + 2 FAD + 2 *PO* ATP
(37)#	2 FADH2 + 1 O2 + 6/5 *PO* ADP + 6/5 *PO* Pi + 6/5 *PO* H	--> (2 + 6/5 *PO*) H2O + 2 FAD + 6/5 *PO* ATP
(38)#	2 NADPH + 1 O2 + 2 *PO* ADP + 2 *PO* Pi + (2 + 2 *PO*) H	--> (2 + 2 *PO*) H2O + 2 NADP + 2 *PO* ATP

Carbon substrates other than glucose

(39)# 1 CIT <-> 1 ISOCIT
(40)# 1 LAC + 1 FAD <-> 1 PYR + 1 FADH2
(41)# 1 GOH + 1 ATP + 1 FAD --> 1 GAP + 1 FADH2 + 1 ADP + 1 H
(42)# 1 GLUCON + 1 ATP --> 1 PYR + 1 GAP + 1 ADP + 1 H2O + 1 H
(43) 1 ETOH + 1 NAD <-> 1 ACET + 1 NADH + 1 H
(44) 1 AC + 1 COA + 2 ATP + 1 H2O --> 1 ACCOA + 2 ADP + 2 Pi + 1 H

Transfer of 1-carbon compounds

(45) 1 THF + 1 ATP + 1 NADH + 1 CO2 <-> 1 FTHF + 1 ADP + 1 Pi + 1 NAD
(46) 1 THF + 1 CO2 + 3 NADH + 3 H <-> 1 MYTHF + 3 NAD + 2 H2O
(47) 1 THF + 1 CO2 + 2 NADH + 2 H <-> 1 METHF + 2 NAD + 2 H2O

Transport

(48) 1 Pi(E) + 2 H(E) <-> 1 Pi + 2 H
(49) 1 NH4(E) + 1 H(E) <-> 1 NH4 + 1 H
(50)# 1 GLUC(E) + 1 H(E) <-> 1 GLUC + 1 H
(51) 1 SO4(E) + 3 H(E) <-> 1 SO4 + 3 H
(52)# 1 AC(E) + 1 H(E) <-> 1 AC + 1 H
(53)# 1 LAC(E) + 1 H(E) <-> 1 LAC + 1 H
(54)# 1 PYR(E) + 1 H(E) <-> 1 PYR + 1 H
(55)# 1 GLUCON(E) + 1 H(E) <-> 1 GLUCON + 1 H
(56)# 1 SUC(E) + 2 H(E) <-> 1 SUC + 2 H
(57)# 1 CIT(E) + 3 H(E) <-> 1 CIT + 3 H

H+ ATP-ase

(58) 1 ATP + 1 H2O --> 1 ADP + 1 H(E) + 1 Pi

Amino acid synthesis

(59) 1 AKG + 1 NH4 + 1 NADPH + 1 H <-> 1 GLUT + 1 NADP + 1 H2O
(60) 1 GLUT + 1 NH4 + 1 ATP --> 1 GLUM + 1 ADP + 1 Pi + 1 H
(61) 1 GLUT + 1 ATP + 2 NADPH + 2 H --> 1 PRO + 1 ADP + 1 Pi + 1 H2O + 2 NADP
(62) 1 ATP + 1 NH4 + 1 CO2 --> 1 CARP + 1 ADP + 2 H
(63) 2 GLUT + 1 ACCOA + 4 ATP + 1 NADPH + 1 CARP + 1 ASP + 3 H2O --> 1 ARG + 1 COA + 1 AKG + 1 AC + 4 ADP + 1 FUM + 5 Pi + 1 NADP + 4 H
(64) 2 GLUT + 1 ACCOA + 3 ATP + 2 NADPH + 2 NAD + 3 H2O --> 1 LYS + 1 COA + 1 AKG + 1 CO2 + 3 ADP + 3 Pi + 2 NADP + 2 H + 2 NADH
(65)# 1 ASP + 1 PYR + 2 NADPH + 1 SUCCOA + 1 GLUT + 1 ATP + 2 H --> 1 LYS + 1 SUC + 1 AKG + 1 CO2 + 2 NADP + 1 COA + 1 ADP + 1 Pi
(66) 1 G3P + 1 GLUT + 1 NAD + 1 H2O --> 1 SER + 1 AKG + 1 Pi + 1 H + 1 NADH
(67) 1 SER + 1 THF <-> 1 GLY + 1 METHF + 1 H2O
(68) 1 SER + 1 ACCOA + 1 SO4 + 4 NADPH + 4 H + 1 ATP --> 1 CYS + 1 AC + 1 COA + 4 NADP + 1 ADP + 3 H2O + 1 Pi
(69) 1 OAC + 1 GLUT <-> 1 ASP + 1 AKG
(70) 1 ASP + 1 NH4 + 2 ATP + 1 H2O --> 1 ASN + 2 H + 2 ADP + 2 Pi
(71) 1 ASP + 1 ATP + 2 NADPH + 2 H --> 1 HOM + 1 ADP + 1 Pi + 2 NADP
(72) 1 HOM + 1 ATP + 1 H2O --> 1 THR + 1 ADP + 1 Pi + 1 H
(73) 1 HOM + 1 SUCCOA + 1 CYS + 1 MYTHF + 2 H2O + 1 ATP --> 1 MET + 1 COA + 1 PYR + 1 SUC + 1 NH4 + 2 H + 1 ADP + 1 Pi + 1 THF
(74) 1 THR + 1 PYR + 1 NADPH + 1 GLUT + 2 H --> 1 ILEU + 1 NH4 + 1 NADP + 1 H2O + 1 CO2 + 1 AKG
(75) 1 PYR + 1 GLUT <-> 1 ALA + 1 AKG
(76) 2 PYR + 1 NADPH + 2 H --> 1 AKI + 1 CO2 + 1 NADP + 1 H2O
(77) 1 AKI + 1 GLUT <-> 1 VAL + 1 AKG
(78) 1 AKI + 1 ACCOA + 1 GLUT + 1 NAD + 2 H2O + 1 ATP --> 1 LEU + 1 AKG + 1 COA + 1 CO2 + 2 H + 1 Pi + 1 NADH + 1 ADP
(79) 2 PEP + 1 E4P + 1 NADPH + 1 ATP --> 1 CHO + 1 ADP + 4 Pi + 1 NADP
(80) 1 CHO + 1 GLUT + 1 H --> 1 PHEN + 1 AKG + 1 CO2 + 1 H2O
(81) 1 CHO + 1 GLUT + 1 NAD --> 1 TYR + 1 AKG + 1 CO2 + 1 NADH
(82) 1 CHO + 1 GLUM + 1 PRPP + 1 SER --> 1 TRYP + 2 Pi + 1 CO2 + 1 GAP + 1 GLUT + 1 H + 1 PYR + 1 H2O
(83) 1 RIBU5P + 2 ATP --> 1 PRPP + 2 ADP + 1 H
(84) 1 PRPP + 3 ATP + 3 H2O + 1 NH4 + 1 GLUM + 2 NAD + 1 NADPH + 1 CO2 --> 1 HIS + 6 Pi + 2 NADH + 1 NADP + 3 ADP + 1 AKG + 8 H

Amino acid polymerisation

(85) 820 GLUT + 285 GLUM + 448 PRO + 437 ARG + 776 LYS
+ 502 SER + 787 GLY + 19 CYS + 806 ASP + 277 ASN
+ 518 THR + 138 MET + 524 ILEU + 1246 ALA + 719 VAL
+ 803 LEU + 364 PHEN + 277 TYR + 76 TRYP + 179 HIS <-> 10000 AM
(86) 10000 AM + 40000 ATP + 30000 H2O --> 48248 PROTEIN + 40000 ADP+ 40000 H + 40000 Pi

Nucleotide synthesis

(87) 1 PRPP + 2 GLUM + 1 GLY + 4 ATP + 1 ASP + 2 H2O + 2 FTHF + 1 CO2 --> 1 IMP + 4 ADP + 6 Pi + 2 GLUT + 2 THF + 1 FUM + 8 H
(88) 1 IMP + 1 ASP + 1 ATP --> 1 AMP + 1 ADP + 1 Pi + 1 FUM + 2 H
(89) 1 IMP + 1 NAD + 2 ATP + 1 GLUM + 3 H2O --> 1 GMP + 2 ADP + 2 Pi + 1 GLUT + 1 NADH + 4 H
(90) 1 GLUM + 1 PRPP + 2 ATP + 1 ASP + 2 H2O + 1 NAD --> 1 UMP + 2 ADP + 4 Pi + 1 GLUT + 1 NADH + 4 H
(91) 1 UMP + 2 ATP <-> 1 UTP + 2 ADP
(92) 1 UTP + 1 GLUM + 1 ATP + 1 H2O --> 1 CTP + 1 ADP + 1 Pi + 2 H + 1 GLUT
(93) 1 CTP + 2 ADP <-> 1 CMP + 2 ATP

RNA synthesis

(94) 2330 AMP + 2330 GMP + 3060 UMP + 2279 CMP + 32279 ATP + 22279 H2O --> 94660 RNA + 32279 ADP + 32279 Pi + 32279 H

ATP hydrolysis

(95)# 1 ATP + 1 H2O --> 1 ADP + 1 Pi + 1 H

Synthesis of fatty acids

(96) 8 ACCOA + 15 ATP + 13 NADPH + 9 H2O --> 1 PAL + 8 COA + 15 ADP + 13 NADP + 3 H + 15 Pi
(97) 9 ACCOA + 17 ATP + 15 NADPH + 10 H2O --> 1 OL + 9 COA + 17 ADP + 15 NADP + 3 H + 17 Pi

Synthesis of polysaccharides

(98) 1 GLUC6P + 1 ATP + 1 H2O --> 6 PSACCH + 1 ADP + 2 Pi + 1 H

Biomass formation

(99) 47003 PROTEIN + 35376 PSACH + 5234 RNA
+ 344 PAL + 344 OL + 226 GOH + 100000 *k* ATP + 100000 *k* H2O --> 100000 BIOM + 100000 *k* ADP + 100000 *k* Pi + 100000 *k* H

Literature Cited

Aiba, S.; Matsuoka, M. Identification of metabolic model: Citrate production from glucose by *Candida lipolytica*. *Biotechnol. Bioeng.* **1979**, *21*, 1373−1386.

Bailey, J. E. Toward a science of metabolic engineering. *Science* **1991**, *252*, 1668−1675.

Beale, E. M. L. Confidence regions in non-linear estimation. *J. R. Stat. Soc.* **1960**, *B22*, 41−88.

Bruinenberg, P. M.; van Dijken, J. P.; Scheffers, W. A. A theoretical analysis of NADPH production and consumption in yeasts. *J. Gen. Microbiol.* **1983**, 129, 953−964.

Bryers, J. D.; Yeh, T. Diauxic metabolism of *Hansenula polymorpha*—Steady and unsteady-state considerations. *Ann. N.Y. Acad. Sci.* 1990, 589, 314−332.

Cassio, F.; Leao, C.; van Uden, N. Transport of lactate and other short-chain monocarboxylates in the yeast *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **1987**, *53*, 509−513.

Cooper, T. G. Transport in *Saccharomyces cerevisiae*. In *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression*; Strathern, J. N., Ed.; Cold Spring Harbor Laboratory: Plainview, NY, 1982; pp 399−461.

de Jong-Gubbels, Vanrolleghem, P.; Heijnen, J. J.; van Dijken, J. P.; Pronk, J. T. Metabolic fluxes in chemostat cultures of *Saccharomyces cerevisiae* grown on mixtures of glucose and ethanol. *Yeast* **1995**, *11*, 407−418.

Goel, A.; Ferrance, J.; Jeong, J.; Ataai, M. M. Analysis of metabolic fluxes in batch and continuous cultures of *Bacillus subtilis*. *Biotechnol. Bioeng.* **1993**, *42*, 686−696.

Harder, W.; van Dijken, J. P. Theoretical considerations on the relation between energy production and growth of methane-utilizing bacteria. In *Microbial Production and Utilization of Gases (H_2, CH_4, CO)*; Schlegel, H. G., Gottschalk, G., Pfennig, N., Eds.; E. Goltze KG: Gottingen, Germany, 1976; pp 403−418.

Harder, W.; van Dijken, J. P.; Roels, J. A. Utilization of energy in methylotrophs. In *Microbial Growth on C1 Compounds*; Dalton, H., Ed.; Heyden: London, 1981; pp 258−269.

Holms, W. H. The central metabolic pathways of *Escherichia coli*: Relationship between flux and control at a branch point, efficiency of conversion to biomass, and excretion of acetate. In *Current Topics in Cellular Regulation*; Academic Press: New York, 1986; Vol. 28, pp 69−105.

Jorgensen, H.; Nielsen, J.; Villadsen, J.; Mollgaard, H. Metabolic flux distribution in *Penicillium chrysogenum* during fed-batch cultivations. *Biotechnol. Bioeng.* **1995**, *46*, 117−131.

Kacser, H.; Burns, J. A. The control of flux. *Symp. Soc. Exp. Biol.* **1973**, *27*, 65−104.

Leao, C.; van Uden, N. Transport of lactate and other short-chain monocarboxylates in the yeast *Candida utilis*. *Appl. Microbiol. Biotechnol.* **1986**, *23*, 389−393.

Majewski, R. A.; Domach, M. M. Simple constrained-optimization view of acetate overflow in E. coli. *Biotechnol. Bioeng.* **1990**, *35*, 732−738.

Munack, A. Optimization of sampling. In *Biotechnology, a Multivolume Comprehensive Treatise*, Vol. 4: Measuring, Modelling and Control; Schügerl, K., Ed.; VCH: Weinheim, Germany, 1991; pp 251−264.

Nielsen, J.; Villadsen, J. *Bioreaction Engineering Principles*; Plenum Press: New York, 1994.

Noorman, H. J.; Heijnen, J. J.; Luyben, K. Ch.A. M. Linear relations in microbial reaction systems: A general overview of their origin, form, and use. *Biotechnol. Bioeng.* **1991**, *38*, 603−618.

Oura, E. The effect of aeration on the growth energetics and biochemical composition of baker's yeast. Ph.D. Thesis, University of Helsinki, Finland, 1972.

Posten, C.; Munack, A. On-line application of parameter estimation accuracy to biotechnical processes. In *Proceedings American Control Conference*; Book, W. J., Ed.; IEEE Publ. Services: New Jersey, 1990; pp 2181−2186.

Raksanyi, A.; Lecourtier, Y.; Walter, E.; Venot, A. Identifiability and distinguishability testing via computer algebra. *Math. Biosci.* **1985**, *77*, 245−266.

Roon, R. J.; Levy, J. S.; Larimore, F. Negative interactions between amino acid and methylamine/ammonia transport systems of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **1977**, 252, 3599−3604.

Savageau, M. A. Biochemical systems analysis. II. The steady state solutions for an n-pool system using a power-law approximation. *J. Theor. Biol.* **1969**, *25*, 370−379.

Serrano, R. Transport across yeast vacuolar and plasma membranes. In *The Molecular Biology of the Yeast Saccharomyces: Genome Dynamics Protein Synthesis and Energetics*; Broach, J. R., Ed.; Cold Spring Harbor Lab.: Plainview, NY, 1991; p 523.

Stanier, R. Y.; Ingraham, J. L.; Wheelis, M. L.; Painter, P. R. *General Microbiology*; MacMillan Education Ltd.: Basingstoke, 1987.

Vallino, J. J.; Stephanopoulos, G. Flux determination in cellular bioreaction networks: Applications to lysine fermentations. In *Frontiers in Bioprocessing*; Sikdar, S. K., Bier, M., Todd, P., Eds.; CRC Press: Boca Raton, FL, 1990; pp 205−219.

Vallino, J. J.; Stephanopoulos, G. Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. *Biotechnol. Bioeng.* **1993**, *41*, 633−646.

van der Heijden, R. T. J. M.; Heijnen, J. J.; Hellinga, C.; Romein, B.; Luyben, K. Ch.A. M. Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates. *Biotechnol. Bioeng.* **1994**, *43*, 3−10.

van Gulik, W. M.; Heijnen, J. J. A metabolic network stoichiometry analysis of microbial growth and product formation. *Biotechnol. Bioeng.* **1995**, *48*, 681−698.

Vanrolleghem, P. A.; Dochain, D. Model Identification. In *Advanced Instrumentation, Data Interpretation and Control of Biotechnological Processes*; Van Impe, J., Vanrolleghem, P., Iserentant, D., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1995, in press.

Vanrolleghem, P. A.; de Jong-Gubbels, P.; Pronk, J. T.; van Dijken, J. P.; Heijnen, J. J. The source of reducing equivalents in *Saccharomyces cerevisiae* during growth on mixed substrates. Manuscript in preparation.

Varma, A.; Palsson, B. O. Predictions of oxygen supply control to enhance population stability of engineered production strains. *Biotechnol. Bioeng.* **1994a**, *43*, 275−285.

Varma, A.; Palsson, B. O. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* **1994**b, *60*, 3724−3731.

Varma, A.; Palsson, B. O. Parametric sensitivity of stoichiometric flux balance models applied to wild-type *Escherichia coli* metabolism. *Biotechnol. Bioeng.* **1995**, *45*, 69−79.

Varma, A.; Boesch, B. W.; Palsson, B. O. Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl. Environ. Microbiol.* **1993**, *59*, 2465−2473.

Verduyn, C.; Stouthamer, A. H.; Scheffers, W. A.; Van Dijken, J. P. A theoretical evaluation of growth yields of yeasts. *Antonie van Leeuwenhoek* **1991**, *59*, 49−63.

von Meyenburg, H. K. Energetics of the budding cycle of *Saccharomyces cerevisiae* during glucose limited aerobic growth. *Arch. Mikrobiol.* **1969**, *66*, 289−303.

Zupke, C.; Stephanopoulos, G. Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrices. *Biotechnol. Prog.* **1994**, *10*, 489−498.

Accepted February 28, 1996.[⊗]

BP960022I

[⊗] Abstract published in *Advance ACS Abstracts,* June 1, 1996.