A Structured Approach for Selection Among Candidate Metabolic Network Models and Estimation of Unknown Stoichiometric Coefficients

P. A. Vanrolleghem,[1] J. J. Heijnen[2]

[1]BIOMATH, University Gent, Coupure Links 653, B-9000 Gent, Belgium; telephone: 32-9-2645932; fax: 32-9-2234941; e-mail: Peter.Vanrolleghem@rug.ac.be
[2]Department of Biochemical Engineering, Delft University of Technology, Julianalaan 67 2628 BC Delft, The Netherlands

Received 7 July 1997; accepted 28 July 1997

Abstract: A metabolic network model is one of the cornerstones of the emerging Metabolic Engineering methodology. In this article, special attention is therefore, given to the phase of model building. A five-stage structured approach to metabolic network modeling is introduced. The basic steps are: (1) to collect a priori knowledge on the reaction network and to build candidate network models, (2) to perform an a priori check of the model, (3) to estimate the unknown parameters in the model, (4) to check the identified model for acceptability from a biological and thermodynamic point of view, and (5) to validate the model with new data. The approach is illustrated with a growth system involving baker's yeast growing on mixtures of substrates. Special attention is given to the central uncertainties in metabolic network modeling, i.e., estimation of energetic parameters in the network and the choice of the source of anabolic reducing equivalents NADPH. © 1998 John Wiley & Sons, Inc. *Biotechnol Bioeng* **58:** 133–138, 1998.
Keywords: metabolic modeling; model selection; parameter estimation; identification; yeast; stoichiometry

INTRODUCTION

During a study that aims to maximize the profits of a fermentation process, one is often confronted with the question whether (a) one should optimize the rate of product formation or the concentration of product in the fermentation broth, or (b) one should try to maximize the product yield (i.e., the amount of product per amount of supplied substrates/precursors).

The conclusion will depend on the chosen production process (including the necessary downstream processing), and a key role in the overall evaluation procedure will be played by the assessment of bioprocess performance. A modeling methodology that can be helpful in this quantification is the subject of this article. Moreover, the methodology also provides quantitative tools that may help in identifying possibilities for process intensification.

Correspondence to: P. A. Vanrolleghem
Contract grant sponsor: Belgian National Fund for Scientific Research

However, let us first formalize the above optimization goals in variables that are more closely related to the production process. To increase the volumetric productivity of a fermentation process (the amount of product formed per unit time and volume), one must increase the rate of product formed per amount of biomass grown (the specific productivity q_p) and/or the biomass concentration in the bioreactor (C_x). On the other hand, if maximum product selectivity is pursued, i.e., the amount of product formed per amount of substrate(s) added (Y_{sp}) is to be maximized, then the specific product formation (q_p) should be maximized while minimizing specific substrate consumption (q_s).

Key to the understanding of the introduced methodology is the general description of the distribution of substrate consumption over the processes of growth, product formation and maintenance.

$$q_s(C - \text{mol substrate}/C - \text{mol biomass} \cdot h) = \frac{1}{Y_{sx}^{max}}\mu + \frac{1}{Y_{sp}^{max}}q_p + m_s \quad (1)$$

Considering this relationship, one observes for the first optimization goal (maximizing $q_p \cdot C_x$) that the specific product formation rate, q_p, is function of the growth rate. Indeed, solving the above equation for q_p yields:

$$q_p(\mu) = Y_{sp}^{max}\left(q_s(\mu) - \frac{1}{Y_{sx}^{max}}\mu - m_s\right) \quad (2)$$

Moreover, it is well known that the biomass concentration attained in a bioreactor also depends on the implemented growth rate (e.g., in a continuous- or fed-batch fermentation). Hence, a rather complex optimization problem may result that has been the subject of important research efforts (Van Impe et al., 1993).

For the second optimization goal (maximizing the product selectivity), it is obvious from the expansion

 CCC 0006-3592/98/020133-06

$$Y_{sp} = \frac{q_p}{q_s} = \frac{q_p}{\frac{1}{Y_{sx}^{max}}\mu + \frac{1}{Y_{sp}^{max}} q_p + m_s} \quad (3)$$

that product selectivity is also growth rate dependent, but in a different way. Here too, the optimization of the growth rate (profile) becomes a complex study.

At this stage, it is important to realize that optimization of the growth pattern requires the knowledge of the $q_p(\mu)$-relationship and the parameters involved in the relationships deduced above (Y_{sx}^{max}, Y_{sp}^{max}, m_s). These parameters are the result of a complex intracellular metabolic network. One way to obtain their values is to perform a range of well designed labor-intensive experiments.

However, an analysis of the stoichiometry of the network reactions can provide their values as well. Moreover, it makes better use of the available biochemical knowledge. In recent years, metabolic network stoichiometry analysis has become a basic tool to understand how these parameters and the $Y_{sp}(\mu)$ functions may be manipulated by revealing metabolic bottlenecks (Bailey, 1991; van Gulik and Heijnen, 1995; Vallino and Stephanopoulos, 1993; Varma et al., 1993). These can be overcome by rational strain manipulation, e.g., mutation/selection, amplification of key enzymes using genetic manipulation, or by directed improvement of cultivation conditions, e.g., specific substrates or particular precursors. This rational approach of process intensification has been termed Metabolic Engineering (Bailey, 1991).

The metabolic model is the cornerstone of the presented methodology and its expected results. It is therefore, worthwhile to pay special attention to the phase of model building preceding the metabolic network application for process optimization. In this contribution, a structured approach to metabolic network modeling is introduced and illustrated experimentally with a growth system involving yeast growing on mixtures of substrates. Special attention will be focused on the following central problems in the stoichiometric model building process:

1. The stoichiometric coefficient *k* for the maintenance demands of the cell's metabolism;
2. The stoichiometric coefficient *PO* describing the effective amount of ATP produced per oxygen reduced in aerobic respiration;
3. The reaction(s) that are the source of anabolic reducing equivalents NADPH.

STRUCTURED APPROACH TO METABOLIC NETWORK MODELING

The structured modeling approach involves five stages.

1. Collection of a priori knowledge on the metabolism and construction of candidate models.
2. A priori check of the proposed metabolic network model.
3. Estimation of unknown parameters in the model.
4. A posteriori check of the identified model.
5. Validation of the model with a new dataset.

Step 1

Collection of A Priori Knowledge and Construction of the Metabolic Network Model

All available a priori knowledge on the biochemistry of the studied cell system is collected and stored in a database of *m* components considered relevant for the description together with *n* relevant biochemical reactions between these components. Among the components, a number *p* are exchanged over the cell membrane (substrates, CO_2), while the remaining *m-p* chemicals are intracellular intermediates (e.g., ATP, NADH). Reactions to be considered are the transport of the *p* solutes over the cell membrane, anabolic reactions (involved in biosynthesis), and catabolic reactions (involved in energy metabolism). The resulting network may be summarized as partially done in Figure 1 for the central metabolism of *Saccharomyces cerevisiae*. Writing the stoichiometry of the reactions is a central task of this first step. For the first reaction of the pentose phosphate cycle (r_{26} in Fig. 1), stoichiometry is as follows:

$$\text{Reaction 26: 1 Glucose-6-P + 2 } NADP^+ \text{ + 1 } H_2O \xrightarrow{v_{26}} \text{1 Ribulose-5-P + 2 NADPH + 1 } CO_2 \text{ + 2 } H^+ \quad (4)$$

For each of the components involved in the reaction network a mass balance can be constructed next. This mass balance consists of all reaction rates, *j*, in which the component, *i* is consumed or produced (e.g., v_{26} is one of the reaction rates to be considered for NADP), its stoichiometry (s_{ij}), and a net conversion rate r_i (for NADP:r_{NADP}). For NADP, the (constant volume) mass balance looks like:

$$r_{NADP} = s_{NADP,1} \cdot v_1 + \ldots + s_{NADP,26} \cdot v_{26} + \ldots + s_{NADP,n} \cdot v_n \quad (5)$$

Combining the mass balances for all *m* components, a matrix formulation of the metabolic network is obtained:

$$\begin{pmatrix} S_{11} & S_{12} & \ldots & S_{1n} & -1 & 0 & \ldots & 0 \\ S_{21} & S_{22} & \ldots & S_{2n} & 0 & -1 & \ldots & 0 \\ \vdots & \vdots & \text{Stoichiometry} & \vdots & \vdots & \vdots & \text{–Identity} & \vdots \\ S_{m1} & S_{m2} & \ldots & S_{mn} & 0 & 0 & \ldots & -1 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_n \\ r_1 \\ \vdots \\ r_m \end{pmatrix} = (S_{mn} - I_m) \cdot R_{n+m} = 0 \quad (6)$$

Note, that in case the pseudo-steady state assumption is accepted for the *m-p* intracellular intermediates, no net conversion takes place so that the corresponding r_i are zero. The net conversion rates of the components transported over the cell membrane are either measured (e.g., CO_2 evolution rate), or can be calculated from the metabolic network. Re-

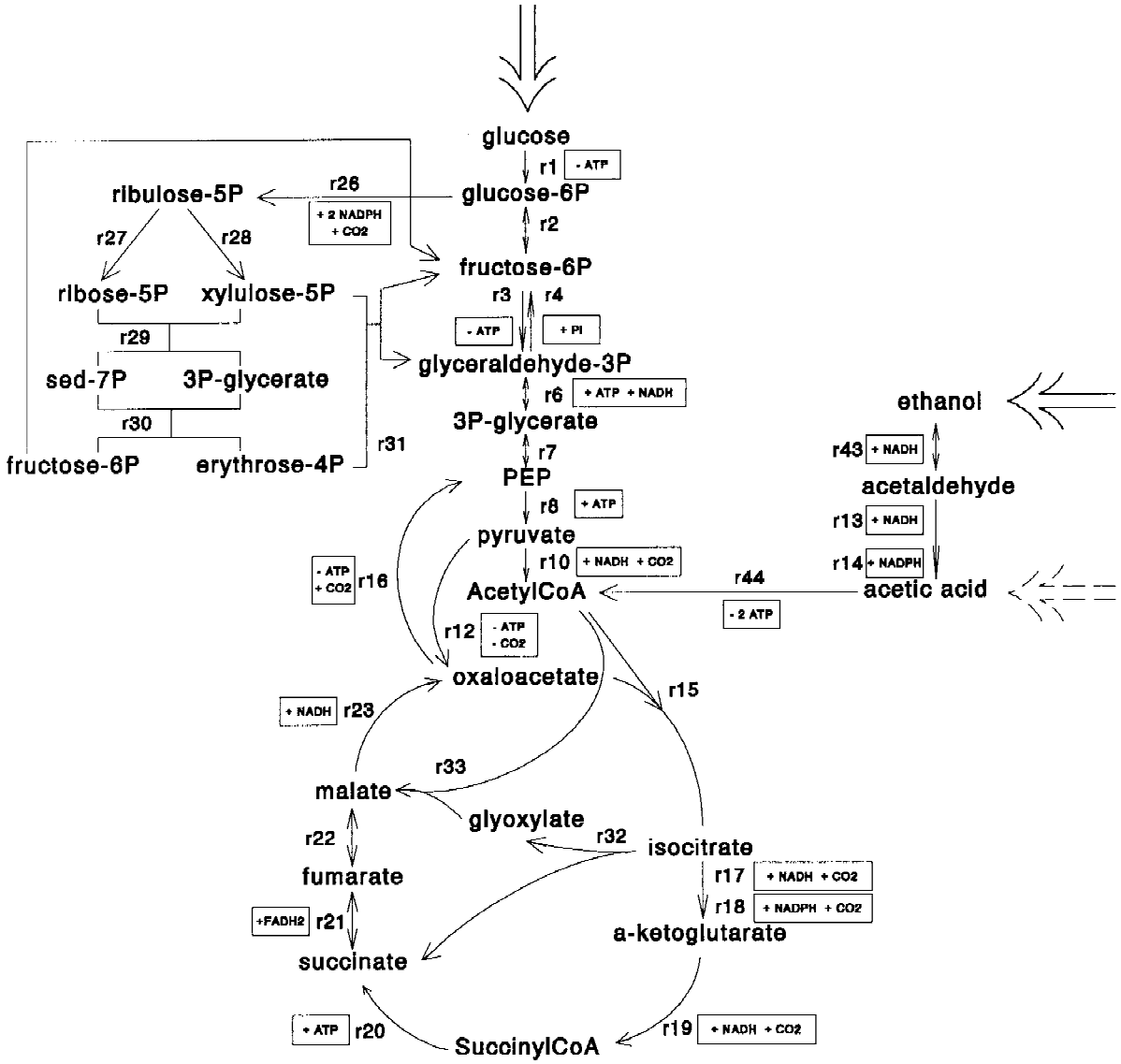

Figure 1. Central metabolic pathways in *Saccharomyces cerevisiae* growing on mixtures of glucose and ethanol (optionally acetic acid).

action rates, v_j can be calculated as well. However, it is the rank of the S matrix that determines the minimum number of rates that must be specified to provide unique values to the other rates. In total $x = (n + p - Rank\ S)$ rates must be defined. Then the reduced row echelon form (RREF, a standard linear algebra matrix operation) of the matrix ($S_{mn} - I_m$) provides the solution of the metabolic network:

$$(v_1\ v_2 \ldots v_n\ r_1\ r_2 \ldots r_{m-x})^T = \text{RREF} \cdot (r_{m-x+1}\ r_{m-x+2} \ldots r_m)^T \quad (7)$$

i.e., each unknown rate (left hand side) is a function of measured rates only (right hand side).

Eight candidate metabolic network models were built for the aerobic growth of *Saccharomyces cerevisiae* on mixtures of glucose and ethanol consisting of a set of 88 or 89 reactions (depending on the candidate model) between 84 components (Vanrolleghem et al., 1996). The differences among the models lie in the choice of the source of reducing equivalents (Table I). They are glucose-6-P-dehydrogenase, acetaldehyde dehydrogenase, and isocitrate dehydrogenase.

Step 2

A Priori Check of the Proposed Metabolic Network Model

Once the metabolic network is built, an a priori check of the metabolic network is performed. First, it is ascertained that the model is non-singular for the system chosen, i.e., the metabolic network and the defined rates. If singularity is found, it means that the solution cannot be obtained without introduction of new information. Hence, in its current form, this model can be discarded from the candidate model set. For the system under study, it was found that solution of the eight networks requires the measurement of only two net conversion rates. In the present work, these were fluxes for

Table I. Sources of reducing equivalents in the set of eight candidate models evaluated for the description of *Saccharomyces cerevisiae* grown aerobically on mixtures of glucose and ethanol.

Candidate Model	Glucose-6-P-DH	Acetaldehyde-DH		Isocitrate-DH	
	NADPH (r_{26})	NADH (r_{13})	NADPH (r_{14})	NADH (r_{17})	NADPH (r_{18})
1	+	+		+	
2	+	+			+
3	+		+	+	
4	+		+		+
5	+	+		+	+
6	+	+	+	+	
7	+		+	+	+
8	+	+	+		+

Reaction numbers given between brackets refer to Figure 1.

glucose and ethanol. However, for the candidate models 5–8 (see Table I) it was necessary to assume that the flux through the pentose phosphate pathway was only determined by the required production of carbon skeletons such as erythrose-4-P (i.e., reaction r_{31} in Figure 1 was not present in these four candidate networks).

At this stage of the model-building process, the singularity evaluation step requires the use of symbolic manipulation software because some parameters in the metabolic model may still remain without actual numerical value. Among these unknown coefficients one always finds the energetic parameters of the cell's metabolism. One is related to the maintenance energy demands and another is, in the case of respiratory growth, the P/O-ratio, i.e., the effective amount of ATP produced in the electron transport chain per oxygen reduced. In the network developed for *S. cerevisae* these two parameters are called *k* and *PO* (Vanrolleghem et al., 1996)

After non-singularity is ascertained, a check is made on whether all reactions are consistent with thermodynamic constraints: Irreversible reactions should have a flux direction consistent with the thermodynamic laws for all conditions prevailing. This check can be performed by evaluating the sign of all reaction rates calculated as the (symbolic) solution of the metabolic network. For the illustrative case study, the rate of the irreversible reaction 3 of Figure 1 is obtained as a function of the ethanol and glucose conversion rates:

$$v_3 = \frac{-0.119(6PO-1)}{4.227PO+2k+4.221} r^{\text{ethanol}} + \frac{(0.230PO+0.333k+0.466)}{4.227PO+2k+4.221} r^{\text{glucose}} \quad (8)$$

PO and *k* are the only two (positive) parameters remaining as unknown in the network. It can be seen that this reaction rate is not consistently negative and, therefore, no decision can be made at this stage whether this model is acceptable or not. Only after values are given to *PO* and *k*, i.e., in step 4, are such decisions are allowed.

Step 3

Estimation of Unknown Parameters in the Model

As mentioned above, some stoichiometric coefficients in the network remain unknown after consideration of all available biochemical knowledge. The task to be fulfilled in step 3 of the network modeling is to attribute values to these parameters by fitting model predictions to experimental data. Data that can be used are, for instance:

1. Growth of an organism on a number of single substrates;
2. Growth/product formation of an organism at different growth rates;
3. Growth of an organism on a number of mixtures of two substrates.

While examples of the former two options were given in van Gulik and Heijnen (1995), the latter option was used in the illustrative case study presented here. Biomass yields on oxygen and carbon source (expressed as C-mol of two-substrate mixtures) were used as measurements. Carbon source-limited chemostat experiments (de Jong-Gubbels et al., 1995) were performed at a dilution rate of 0.1 h^{-1} over a range of mixtures of the two carbon sources ethanol and glucose. The different mixtures were characterized by the glucose fraction (f_{GLU}) in the feed.

For the first candidate model (Table I), the sum of weighted-squared errors (J) is plotted in Figure 2 as a function of the two parameters to be estimated. Errors were considered between measurements and model predictions of biomass yield on carbon Y_{sx} and oxygen Y_{ox} (Fig. 3). Van-

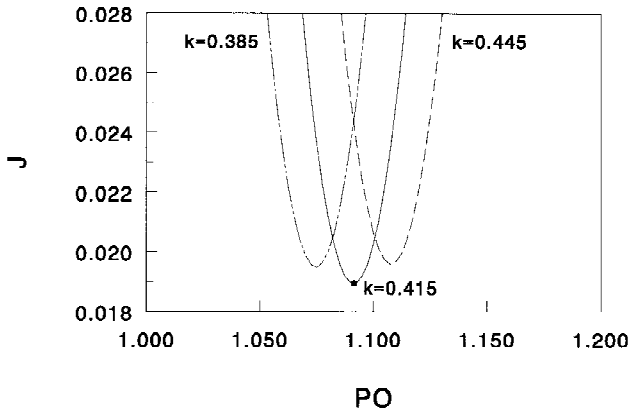

Figure 2. Evolution of the sum of squared errors objective function for different values of *PO* and *k* for candidate network model 1.

rolleghem et al. (1996) performed a thorough analysis of the (theoretical and practical) identifiability problems associated with this estimation problem and concluded that unique values could be given to the parameters with rather small confidence regions (1 to 4% coefficient of variation).

Step 4

A Posteriori Check of the Identified Model

In the fourth phase, an a posteriori check of the metabolic network is performed. First, the biochemical acceptability of the estimated coefficients is evaluated. For instance, it is unacceptable that one obtains negative P/O-ratios or maintenance values *k.* Although different *PO* and *k* values were obtained for the eight different candidate models of this study, none of them were biologically unacceptable and hence, no candidate model was discarded on the basis of this criterion.

Second, the non-violation of thermodynamic constraints is ascertained again. However, now the model is complete, and explicit calculations can be made on the range of metabolic conditions for which the metabolic network model is thermodynamically acceptable. The relations obtained in step 2 can be used, but at this stage the parameters can be replaced by the estimated values. Now clear regions of acceptability as function of the conversion rates can be made as illustrated next for the *S. cerevisiae* system.

While all eight candidate networks originally proposed in the illustrative case study passed the a priori check of step 2, seven of them could be discarded in the thermodynamic check performed here. First, if the flux through glucose-6-P-dehydrogenase (reaction 26 in Fig.1) in the pentose phosphate pathway was considered, three reaction networks (networks 2, 3, and 4, see Fig. 4) showed reversed fluxes (negative sign) for part of the range of glucose/ethanol ratios studied. Hence, these three networks violated the thermodynamic irreversibility of this reaction. On the basis of this finding, they could be discarded as possible network models. Additional evidence was given by determinations of the in vitro activity (a measure of the level of expression of a particular enzyme in the cell) of glucose-6-P-dehydrogenase as it coincided particularly well with the fluxes predicted by the metabolic networks 1 and 5–8.

Candidate networks 5 to 8 have two NADPH-mediated dehydrogenase reactions of which one is always combined with a NADH-dependent dehydrogenase reaction in a single oxidation step (in other words, isozymes which can either use NADH or NADPH are assumed to exist for that reaction). At certain glucose/ethanol mixtures, these reaction networks always resulted in conversions between NADH and NADPH caused by simultaneous oxidation and reduction reactions (acetate <–> acetaldehyde or α-ketoglutarate <–> isocitrate). As the reduction reactions are thermodynamically impossible, it could be concluded that only one proposed network was acceptable for the system under study, i.e., the network whose only source of anabolic-reducing equivalents NADPH is the glucose-6-P-dehydro-

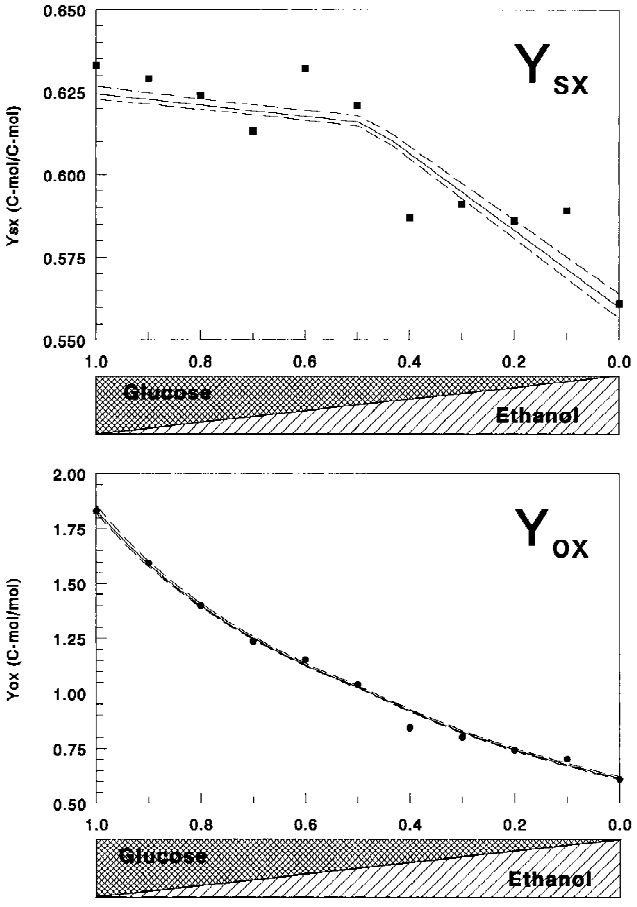

Figure 3. Yield measurements and model predictions for growth of *S. cerevisiae* on mixtures of glucose and ethanol for candidate network model 1.

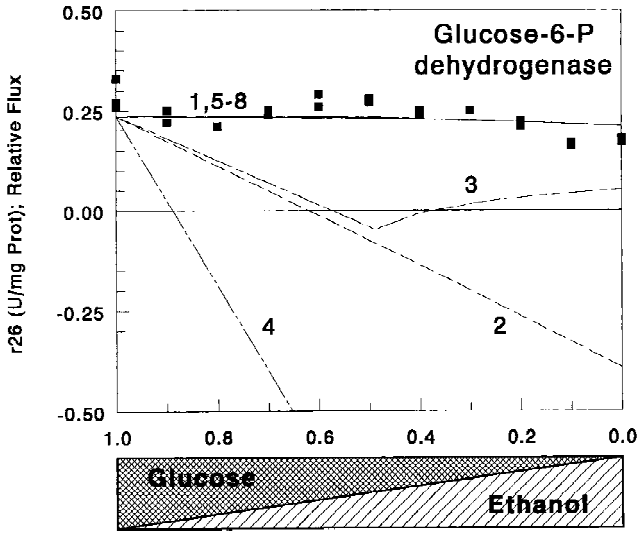

Figure 4. Comparison between the measured activity of the enzyme catalyzing reaction 26 (Fig. 1) and predictions by eight different metabolic networks.

genase-catalyzed reaction and in which acetaldehyde and isocitrate oxidation are catalyzed by NADH-dependent dehydrogenases.

Step 5

Validation of the Model with a New Dataset

In the last model-building step, the metabolic network has to be validated by confronting model predictions with experimental data not used during parameter estimation (step 3). For the case study, additional data were available that allowed for a thorough validation of the model.

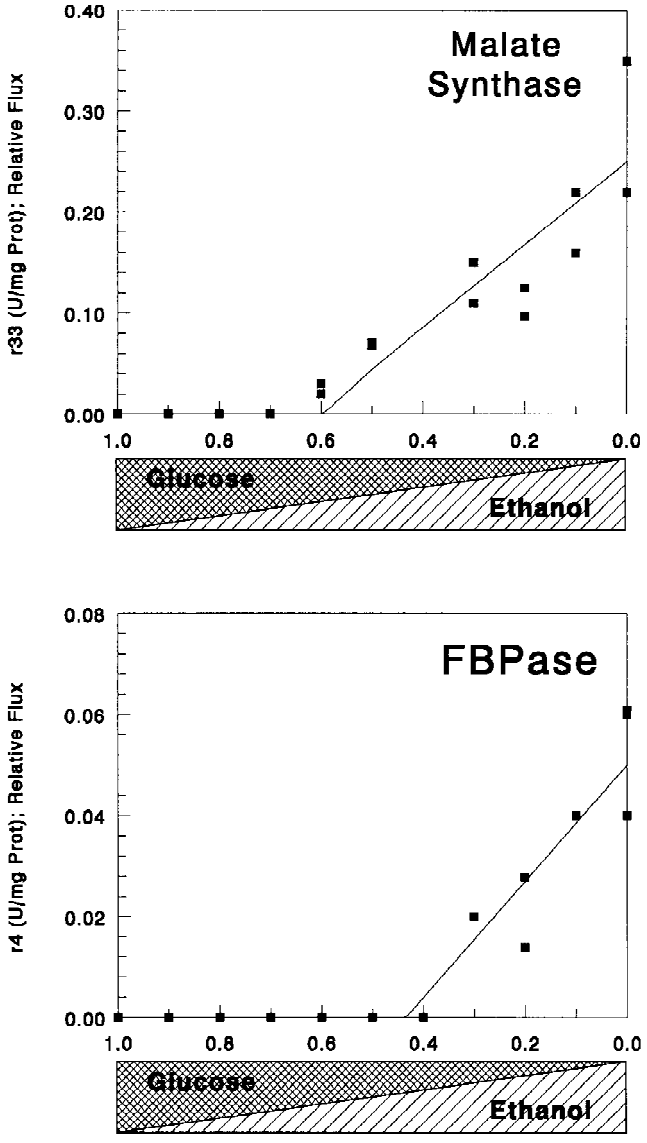

Figure 5. Measured and predicted enzyme activities for *S. cerevisiae* grown on glucose/ethanol mixtures for the selected metabolic network model.

First, a new substrate (in casu acetic acid) was administered to the continuous culture, and the biomass yield predicted by the metabolic model (Y_{sx} = 0.34 C-mol/C-mol) corresponded very well with the measured yield (Vanrolleghem et al., 1996)

Second, activities of key enzymes in the central pathways were measured and could be confronted with predicted reaction rates. These gave excellent agreement as exemplified in Figure 5. Vanrolleghem et al. (1996) deduced from these comparisons that certain pathways were regulated at the enzyme expression level, whereas the in vivo activity of others was apparently mediated by allosteric modification.

CONCLUSION

A structured modeling methodology for metabolic network analysis was proposed. It was applied to the metabolic description of aerobic growth of *S. cerevisiae* on mixtures of glucose and ethanol. Ample experimental evidence was collected and used to obtain reliable and biologically acceptable estimates of the energetic parameters. A number of candidate network models were discarded because they violated thermodynamic constraints. However, this check could only be performed after the energetic parameters were estimated. A priori evaluation of thermodynamic consistency of the network models remained inconclusive. Hence, a fully calibrated description of the yeast's metabolism over the whole range of mixtures of both carbon sources was obtained. Additional data allowed for the thorough validation of the model.

It may be concluded that the given modeling methodology supports metabolic modeling and is therefore, a useful ingredient of process intensification studies.

References

Bailey J. E. (1991). Toward a science of metabolic engineering. Science **252**: 1668–1675.

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

Vallino J. J., Stephenopoulos G. (1993) Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. Biotechnol. Bioeng. **41**: 633–646.

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

Van Impe J. F., Nicolaï B. M., Vanrolleghem P. A., Spriet J. A., De Moor B., Vandewalle J. (1993). Optimal control of the penicillin G fed-batch fermentation: An analysis of the model of Heijnen et al. Optimal Control Applications and Methods **15**: 13–34.

Vanrolleghem P. A., de Jong-Gubbels P., van Gulik W. M., Pronk J. T., van Dijken J. P., Heijnen J. J. (1996). Validation of a metabolic network for *Saccharomyces cerevisiae* using mixed substrate studies. Biotechnol. Prog. **12**: 434–448.

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