Metabolic Network Modelling: Improving Predictions of Microbial Metabolism by Maximal Incorporation of Knowledge on Biochemical Reaction Stoichiometry

P.A. Vanrolleghem[*] *and J.J. Heijnen*[#]

[*]BIOMATH, University Gent, Coupure Links 653, B-9000 Gent, Belgium
[#]Department of Bioprocess Engineering, Delft University of Technology,
Julianalaan 67 2628 BC Delft, The Netherlands

ABSTRACT

To study a cell's metabolism quantitatively and (re)direct it in the desired way, a new methodology termed "Metabolic Network Stoichiometry Analysis" has recently evolved. The approach may provide: 1) maximum biomass and product yields for single and mixed substrates, 2) limits to product yields due to biochemical constraints and 3) product yields as function of the biomass' growth rate and it can indicate metabolic bottlenecks that can be the subject of directed manipulation of strains or cultivation conditions, known as "Metabolic Engineering".
A model of the metabolic network is the corner stone of this methodology and the results one can expect of it. In this paper attention is therefore given to the phase of model building. A structured approach to metabolic network modelling is introduced and is illustrated with a growth system involving yeast growing on mixtures of substrates.

INTRODUCTION

During a study that aims to maximize the profits of a fermentation process, one is often confronted with the question whether 1) one should optimize the concentration of product in the fermentation broth or 2) 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 modelling methodology that can be helpful in this quantification is the subject of this paper. Moreover, the methodology also provides with quantitative tools that may help in identifying possibilities for process intensification.

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 amount of product formed per amount of biomass grown (the specific productivity q_p) and/or the biomass concentration in the bioreactor (C_x). If, on the other hand, 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.

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

Considering this relationship, one observes for the first optimization goal, i.e. maximizing $q_p.C_x$,

that the specific product formation rate q_p is function of the growth rate: indeed, solving (1) for q_p results in:

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

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, i.e. maximizing the product selectivity, it is obvious from the expansion

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

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 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 labour-intensive experiments.
However, an analysis of the stoichiometry of the network reactions may provide their values as well and makes better use of the available biochemical knowledge. Moreover, 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; Vallino and Stephanopoulos, 1993; Varma et al., 1993; van Gulik and Heijnen, 1995). 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 corner stone of the presented methodology and the results one can expect of it. 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 modelling is introduced and is illustrated with a growth system involving yeast growing on mixtures of substrates.

STRUCTURED APPROACH TO METABOLIC NETWORK MODELLING

The structured modelling approach involves five stages.

- Collection of a priori knowledge on the metabolism and construction of the model
- A priori check of the proposed metabolic network model
- Estimation of unknown parameters in the model
- A posteriori check of the identified model
- Validation of the model with a new dataset

In the sequel more details and illustrations are given for the different model building steps.

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. Among these *p* compounds are exchanged over the cell membrane (substrates, CO_2), while the remaining *m-p* chemicals are intracellular intermediates (e.g. ATP, NADH).

Next, *n* relevant biochemical reactions between these components are described. Reactions 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 Fig. 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:

$$\mathbf{r_{26}:\ 1\ Glucose\text{-}6\text{-}P + 2\ NADP + 1\ H_2O \xrightarrow{v_{26}} 1\ Ribulose\text{-}5\text{-}P + 2\ NADPH + 1\ CO_2 + 2\ H}$$

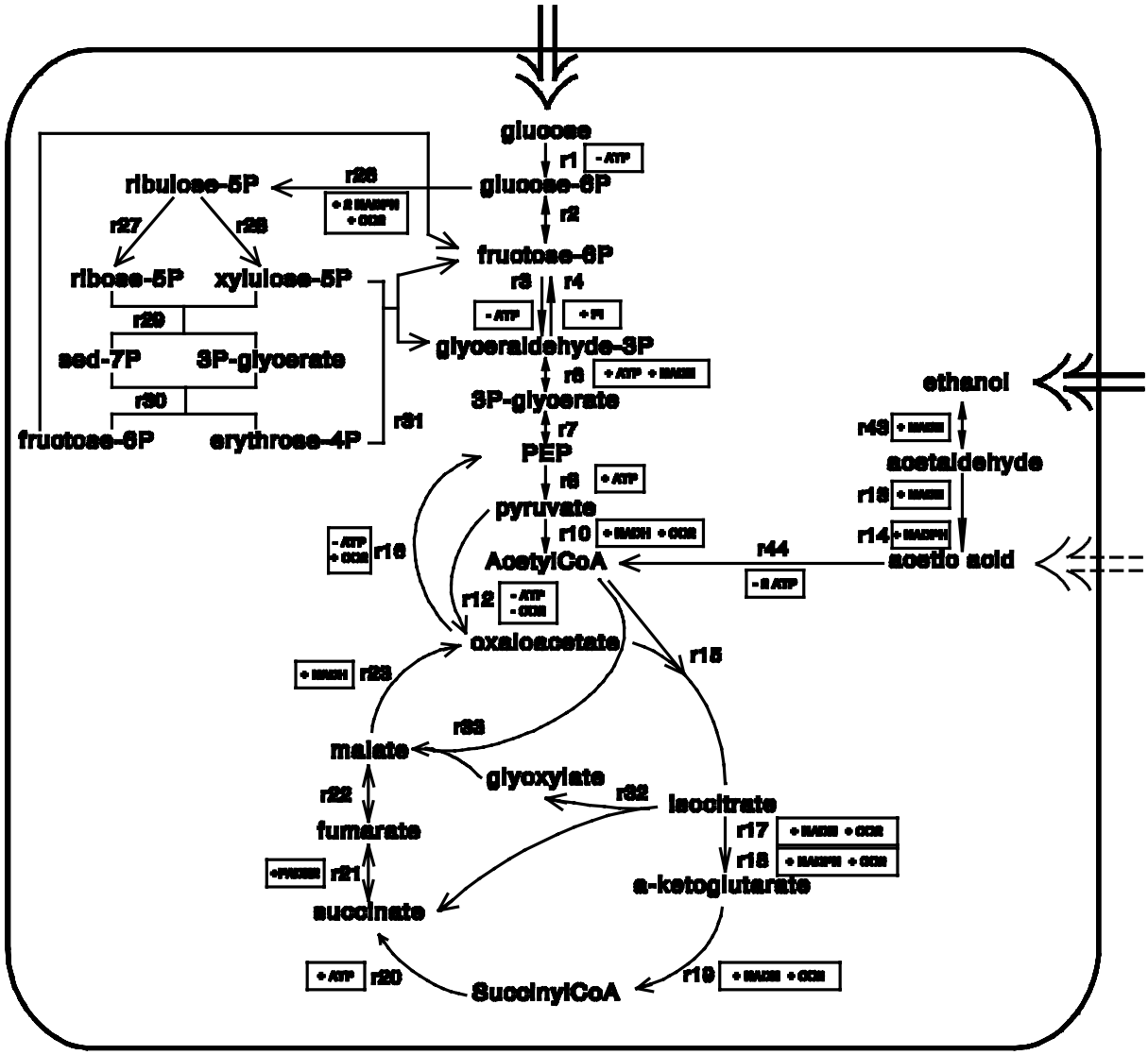


Figure 1*. Central Metabolic Pathways in Saccharomyces cerevisiae growing on mixtures of glucose and ethanol (optionally acetic acid).*

For each of the components involved in the reaction network a mass balance is constructed. This mass balance consists of all reaction rates in which the component 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 (for NADP: r_{NADP}). For NADP, the mass balance looks like:

$$\frac{dC_{NADP}}{dt} = s_{NADP,1}.v_1 + \cdots + s_{NADP,26}.v_{26} + \cdots + s_{NADP,n}.v_n - r_{NADP} \qquad \textbf{(4)}$$

In case continuous cultures are considered, the concentration of the component in the fermenter remains constant and, hence, the derivative vanishes.

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

$$\begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} & -1 & 0 & \cdots & 0 \\ s_{21} & s_{22} & \cdots & s_{2n} & 0 & -1 & \cdots & 0 \\ . & . & \textit{Stoichiometry} & \vdots & \vdots & \vdots & -\textit{Identity} & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mn} & 0 & 0 & \cdots & -1 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_n \\ r_1 \\ \vdots \\ r_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_{mn} & -\boldsymbol{I}_m \end{pmatrix} . \boldsymbol{R}_{n+m} = \boldsymbol{0} \qquad \textbf{(5)}$$

Note that 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, O_2 uptake rate or biomass production rate) or can be calculated from the metabolic network. All reaction 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 $x = (n + p - Rank\ \boldsymbol{S})$ rates must be defined. Then the reduced row echelon form of the matrix (obtained using standard linear algebra) provides the solution of the metabolic network:

$$\begin{pmatrix} v_1 & v_2 & \cdots & v_n & r_1 & r_2 & \cdots & r_{m-x} \end{pmatrix}^T = \boldsymbol{RREF} \cdot \begin{pmatrix} r_{m-x+1} & r_{m-x+2} & \cdots & r_m \end{pmatrix}^T \qquad \textbf{(6)}$$

i.e. each unknown rate (left hand side) is function of measured rates only (right hand side).
The metabolic network model built for the aerobic growth of *S. cerevisiae* on mixtures of glucose and ethanol is based on a set of 88 reactions between 84 components (for details see Vanrolleghem et al., 1995). Solution of the network requires the measurement of two net conversion rates, in the work these were fluxes for glucose and ethanol.

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 reduced row echelon form cannot be obtained and therefore the network cannot be solved without introduction of new information. This step requires the use of symbolic manipulation software because some parameters in the metabolic model may still remain without actual numerical value at this stage of the model building process.

Second, a check is made 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 Fig. 1 is obtained as a function of the ethanol and glucose conversion rates:

$$v_3 = \frac{-0.119(6PO - 1)r^{ethanol} + (0.230PO + 0.333k + 0.466)r^{glucose}}{4.227PO + 2k + 4.221} \quad (7)$$

PO and *k* are the only two (positive) parameters remaining as unknown in the network (see Vanrolleghem et al.. 1995). 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, such decisions are allowed.

Step 3. Estimation of unknown parameters in the model

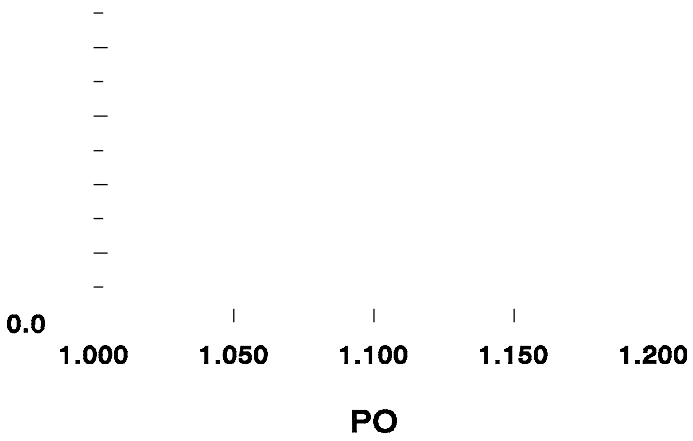
As mentioned above, some stoichiometric coefficients in the network remain unknown after consideration of all available biochemical knowledge. The task to be fulfilled in step 3 of the network modelling is to attribute values to these parameters by fitting model predictions to experimental data. Among these unknown coefficients one always finds the energetic parameters of the cell. One is related to the maintenance energy demands and another is, in 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. cerevisiae* these two parameters are called *k* and *PO*. Experimental data that can be used are for instance conversion rates of different components or ratios thereof, e.g. biomass yields for growth on a number of substrates or mixtures thereof. In the case study, biomass yields on oxygen and carbon source (expressed as C-mol) were used as measurements. 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 expressed as the glucose fraction (f_{GLU}) in the feed.

In Fig. 2 the sum of squared errors for different values of the two parameters to be estimated is given. Errors were considered between measurements and model predictions of biomass yield on carbon Y_{sx} and biomass yield on oxygen Y_{ox} (Fig. 3). Vanrolleghem et al. (1995) performed a thorough analysis of the identifiability problems associated with this parameter estimation problem and concluded that unique values could be given to the parameters with rather small confidence regions (1 to 4 percent 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 negative P/O-ratios would be obtained.
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 their estimated values. Now, clear regions of acceptability as function of the conversion rates can be made as illustrated next for the *S. cerevisiae* system.


0.0
1.000
1.050
1.100
1.150
1.200
PO