Computer-Aided Design of Metabolic Networks

Klaus Mauch, Stefan Buziol, Joachim Schmid and Matthias Reuss *
Institute of Biochemical Engineering
University of Stuttgart, Germany

Abstract

This contribution presents results obtained from a model based design of metabolic networks. In the first part of the paper, topological analysis is used for exploring the metabolic architecture. These investigations—also called pathway analysis or flux space analysis—are aimed at detecting the metabolic routes that lead from anyone starting point to some products. The technique is applied for the computation of maximal yields for amino acids and, for the first time, also for the analysis of metabolic networks in context with the formation of biomass. The latter study leads to an array of mutants with different biomass yields, for which the name "Phenome" has been coined.

In the second part of the contribution, a strategy for the optimization of product formation rates is presented by means of the ethanol formation rate in *Saccharomyces cerevisiae*. A dynamic model based on experimental observations at defined anaerobic conditions serves as a starting point. Non-linear optimization of the distribution of enzyme activities results in a substantial increase of ethanol formation rate. The optimum is mainly constrained by homeostasis and can be characterized by higher activities of strongly rate limiting steps. However, some enzymes exerting almost no control on ethanol flux (e.g. triose phosphate isomerase) are found at higher activities as well. This finding can be explained by the enzyme's ability of counteracting an increase of pool concentrations effectively.

Keywords

Metabolic networks, Flux analysis, Flux optimization

Introduction

One of the fascinating challenges in mastering biosystems is to interpret living processes in a quantitative manner. As such, mathematical modeling is of central importance. At a time of the ballooning amount of data generated by the high throughput technology in genome, transcriptome, proteome and metabolome research, one of the important issues for modeling is to bridge the gap between data and an integrated understanding of the complex functionality of biosystems. Moreover, there is an urgent need for new approaches to strengthen the model based design of biosystems. This activity is of increasing relevance with respect to the optimization of yields, selectivities and productivities in industrial bioprocesses. Furthermore, mathematical modeling will lead to significant insight through an integrative analysis of diseases and support target identification for drug discovery.

While the potential and promise of biological systems modeling is substantial, also several obstacles are encountered. Before benefits can be gathered from biological systems analysis, issues like, for example, the appropriate balance between significance, complexity and availability of quantitative experimental observations need to be addressed. Furthermore, in many cases insufficient emphasis has been placed on fundamental questions of purpose, intended application (Bailey, 1998), predictive power and relevance concerning significant contributions to the solution of the aforementioned problems.

This contribution aims at the design of biosystems within industrial applications. Therefore, modeling is part of a well known engineering cycle depicted in Figure 1.

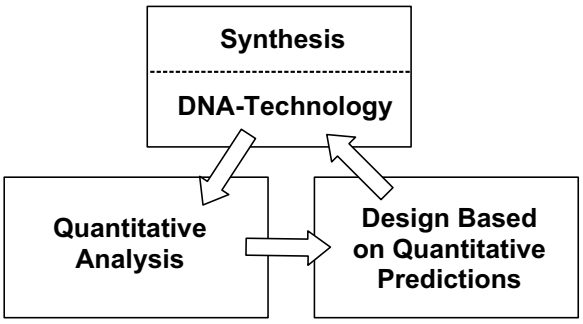

Figure 1: Optimization of bioprocesses through metabolic engineering.

Central to the general modeling framework is its predictive strength. A sound prediction, in turn, must rest upon reliable experimental data. In the majority of cases, implementation of the model-based suggestions for genetic reprogramming is performed with the aid of recombinant DNA-technology. Unfortunately, only few models qualify for the design of metabolic networks. When extrapolation beyond the horizon of experimental observations is required, the missing link mostly is predictive strength. A critical assessment of the state of the art and meaningful discussion of the present limits would require a comprehensive evaluation of several fundamental issues of modeling biosystems. Only two of these issues will be addressed in this contribution: (1) Flux analysis based on topological properties and (2) Flux optimization based both on toplogical and kinetic

*reuss@ibvt.uni-stuttgart.de

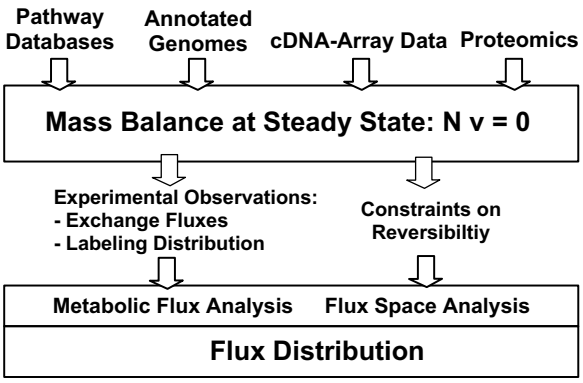

Figure 2: Quantification of flux distributions.

properties of metabolic networks.

Flux Analysis

According to the source of information, balance equations for the metabolic system of interest can be created (see Figure 2).

When balance equations are derived from annotated sequence data, the resulting stoichiometric matrix reflects the metabolic capabilities of the genotype. While depending on physiological conditions, a more specific phenotype is considered when data from cDNA chips or proteome analysis are used. Generally, balance equations of structured metabolic models can be written as

$$\frac{d}{dt}\mathbf{n} = \mathbf{N}\,\mathbf{v}, \tag{1}$$

where matrix $\mathbf{N}\,(m \times n)$ contains the stoichiometric coefficients $\nu_{i,j}$ of the n biochemical reactions. m denotes the number of metabolites, $\mathbf{v}$ the vector of reaction rates, whereas the vector of metabolites is denoted by $\mathbf{n}$. Note that in Equation 1 the transport rates across the various membranes of the cell as well as the dilution of metabolites due to growth are included in the state vector of the reaction rates $\mathbf{v}$. For steady state conditions Equation 1 reads:

$$\mathbf{N}\,\mathbf{v} = \mathbf{0}. \tag{2}$$

Metabolic Flux Analysis

Depending on the data available and the area of application, Equation 2 can be used for the computation of flux distributions in metabolic networks. The first route delivers solutions for flux distributions through experimentally determined exchange fluxes (Stephanopoulos et al., 1998; Mauch et al., 2000).

For the estimation of unknown metabolic fluxes from experimentally determined fluxes, we rewrite Equation 2 as

$$[\,\mathbf{N}_m|\,\mathbf{N}_c]\left[\frac{\mathbf{v}_m}{\mathbf{v}_c}\right] = \mathbf{0}, \tag{3}$$

where vector $\mathbf{v}_m$ consists of q experimentally determined fluxes. The remaining $(q-n)$ unknown fluxes are gathered in vector $\mathbf{v}_c$. Partitioning of $\mathbf{v}$ into the known fluxes $\mathbf{v}_m$ and unknown fluxes $\mathbf{v}_c$ then yields

$$\mathbf{N}_m\mathbf{v}_m + \mathbf{N}_c\mathbf{v}_c = \mathbf{0}, \tag{4}$$

with matrices $\mathbf{N}_m\,(m \times q)$ and $\mathbf{N}_c\,(m \times n-q)$ corresponding to $\mathbf{v}_m$ and $\mathbf{v}_c$, respectively. From Equation 4 we obtain

$$\mathbf{N}_c\mathbf{v}_c = -\mathbf{N}_m\mathbf{v}_m, \tag{5}$$

and upon multiplying Equation 5 by the transposed of matrix $\mathbf{N}_c$, the solution for unknown fluxes $\mathbf{N}_c$ is obtained according to

$$\mathbf{v}_c = -\left(\mathbf{N}_c^T\mathbf{N}_c\right)^{-1}\mathbf{N}_c^T\mathbf{N}_m\mathbf{v}_m, \tag{6}$$

where the superscript -1 indicates matrix inversion. By defining the pseudoinverse $\mathbf{N}_c^{\#}$

$$\mathbf{N}_c^{\#} = \left(\mathbf{N}_c^T\mathbf{N}_c\right)^{-1}\mathbf{N}_c^T, \tag{7}$$

Equation 6 may be rewritten as

$$\mathbf{v}_c = -\mathbf{N}_c^{\#}\mathbf{N}_m\mathbf{v}_m. \tag{8}$$

Mathematically, a determined metabolic system is defined by

$$dim\,(\mathbf{v}_m) = n - rank\,(\mathbf{N})\,. \tag{9}$$

That is, the amount of experimentally determined fluxes q in $\mathbf{v}_m$ equals the degree of freedom of the metabolic network. If no conservation relations are present in a determined system, the pseudoinverse $\mathbf{N}_c^{\#}$ coincides with the inverse of $\mathbf{N}_c$, thus

$$\left(\mathbf{N}_c^T\mathbf{N}_c\right)^{-1}\mathbf{N}_c^T = (\mathbf{N}_c)^{-1}\,. \tag{10}$$

Hence, the fluxes $\mathbf{v}_c$ of a determined system without conservation relations may be obtained by

$$\mathbf{v}_c = -\mathbf{N}_c^{-1}\mathbf{N}_m\mathbf{v}_m. \tag{11}$$

Importantly, a solution for $\mathbf{v}_c$ only exists for a nonsingular matrix $\mathbf{N}_c$, that is

$$det\,(\mathbf{N}_c) \neq 0, \tag{12}$$

and for a determined metabolic system with $0 \leq z < m$ conservation relations:

$$det\left(\mathbf{N}_c^T\mathbf{N}_c\right) \neq 0. \tag{13}$$

Provided the existence of a unique solution for $\mathbf{v}_c$, the metabolic system is called observable. For an underdetermined metabolic network we have

$$dim\,(\mathbf{v}_m) < n - rank\,(\mathbf{N})\,, \tag{14}$$

and since a unique solution for $\mathbf{v}_c$ cannot be obtained, we always find

$$det\left(\mathbf{N}_c^T\mathbf{N}_c\right) = 0. \tag{15}$$

Hence, the dimension l of the solution space for an underdetermined system is given by

$$l = n - rank\,(\mathbf{N}_c) - dim\,(\mathbf{v}_m)\,, \tag{16}$$

whereas an overdetermined (i.e. redundant) metabolic system is characterized by

$$dim\,(\mathbf{v}_m) > n - rank\,(\mathbf{N})\,. \tag{17}$$

Similarly to a determined system, in overdetermined systems a unique solution for the unknown fluxes $\mathbf{v}_c$ can only be derived if Equation 13 holds. Note that an overdetermined system is not necessarily observable. Frequently we meet the situation where experimental information on some fluxes is redundant while part of the metabolic network still cannot be observed. The problem of non-observable fluxes can often be bypassed when experimental data derived from tracer experiments are available. Those measurements of labeled substrates are either performed with NMR (Marx et al., 1996; Szyperski et al., 1999) or gas chromatography/mass-spectroscopy (GC-MS) (Christensen and Nielsen, 2000; Dauner and Sauer, 2000). When metabolite balance equations are extended with balance equations for the metabolite labeling distributions, a system of non-linear equations has to be solved to estimate the associated steady state flux distributions (Wiechert and de Graaf, 1997; Wiechert et al., 1997; Schmidt et al., 1997, 1999).

Flux Space Analysis

Flux space analysis (see Figure 2) summarizes strategies for topological analysis leading to meaningful information on the flux space in metabolic networks. This is an area of increasing importance in integrative functional genomics aimed at a better understanding of the complex relation between genotype and phenotype (Schuster et al., 1999; Schilling et al., 1999, 2000; Schuster et al., 2000a; Edwards and Palsson, 1999; Palsson, 2000). This area—in terms of the "omic" revolution sometimes named "phenomics"—becomes urgent for developing engineering methods to deal with the massive amounts of genetic and expression data in an integrative and holistic way.

Flux space analysis (sometimes also called pathway analysis) aims at detecting the metabolic routes that lead from anyone starting point to the products of interest steady state condtitions. The two strategies applied for this analysis are the null space and the convex cone. While linear algebra is used to determine the null space of the homogeneous system of linear equations, methods of convex analysis (Vanderbei, 1998) are applied to compute the convex cone.

The null space of matrix $\mathbf{N}$ is the subspace spanned by $k = rank\,(\mathbf{N}) - dim\,(\mathbf{v})$ linearly independent vectors $\mathbf{v}$ satisfying Equation 2. Accordingly, every linear independent base vector forms a column of the null space matrix $\mathbf{K}\,(n \times k)$, thus

$$\mathbf{N}\,\mathbf{K} = \mathbf{0}. \tag{18}$$

Within the null space lie all of the flux distributions under which the system can operate at steady state. Thus, the null space allows to describe any flux distribution of a genotype (by superposing it's base vectors). The vectors spanning the nullspace are, however, non-unique solutions of Equation 2. Moreover, the base vectors of the nullspace do not necessarily fulfill reversibility criteria of individual reactions.

A different approach rests upon convex analysis and leads to unique sets of vectors spanning the space of admissible fluxes, for which names like, for example, elementary flux modes (Heinrich and Schuster, 1996) or extreme pathways (Schilling et al., 2000) have been coined for. In contrast to the base vectors of the nullspace, flux vectors obtained by convex analysis always obey sign restrictions of practically irreversible reaction steps. The investigations presented in the following are based on the concepts of the elementary flux modes.

Elementary Flux Modes

Elementary flux modes are non-decomposable flux distributions admissible in steady state, including reaction cycles (Heinrich and Schuster, 1996). Normally, elementary flux modes comprise a set of non-zero fluxes and a set of zero fluxes, with the latter pointing to enzymes which are not used to implement a specific function. An example for an elementary mode which frequently occurs in cellular systems is the complete oxidation of substrate in the respiratory chain. To sustain respiration, enzymes catalyzing anabolic reactions obviously become dispensable. Another example might be an elementary flux mode leading to the formation of an amino acid where, again, larger parts of the respiratory chain are nonessential. Thus, by computing elementary flux modes, the metabolic capacity of a given metabolic network is unitized. In other words, the phenotype of a certain genotype may be characterized by the complete set of elementary flux modes.

The motivation for the study of elementary flux modes arises from various potential applications. In biotechnology, an important objective is to increase the yield of biosynthetic processes where a desired product can often be synthesized by various different routes. It is then of interest to detect and subsequently implement the route on which the product/substrate ratio is maximum. Generally, a flux pattern that uses only the optimal route cannot be obtained in practice. Nevertheless, it is helpful to compute the upper limits for the molar yield from a given network topology.

It has turned out that the problem of maximizing the yield of a biotransformation can be solved by detecting all elementary modes in the system and choosing the mode giving the best yield (Schuster et al., 1999). Al-

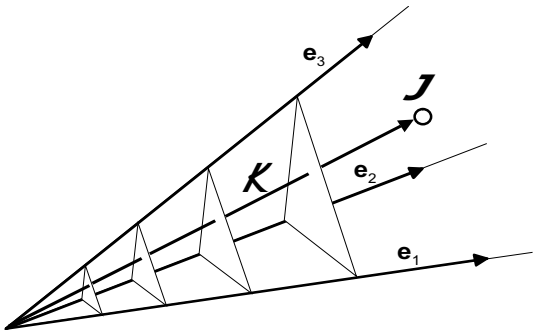

Figure 3: Convex polyhedral cone $\mathcal{K}$ spanned by three generating vectors $\mathbf{e_1}$ to $\mathbf{e_3}$. Flux distribution **J** results from a linear combination of the generating vectors $\mathbf{e_i}$.

ternatively, such optimization problems were tackled by linear programming (Fell and Small, 1986; Savinell and Palsson, 1992; van Gulik and Heijnen, 1995).

In many cases, the net direction of a reaction is known. Therefore, we decompose the flux vector into two subvectors, $\mathbf{v}^{irr}$ and $\mathbf{v}^{rev}$, corresponding to what will be called the irreversible and reversible reactions, respectively. So we have

$$\mathbf{v}^{irr} \geq 0. \tag{19}$$

Equation 2 in conjunction with inequality(19) determines what is called a convex polyhedral cone. The edges of the convex cone are established by the elementary flux modes, and all the points on the interior of the cone can be represented as positive combinations of these fundamental pathways (see Figure 3). From there, the convex cone enfolds all potential stationary flux distributions of a metabolic system.

Pursuing the goal to find basic pathways in biochemical reaction networks, Schuster and Schuster (1993) have earlier developed a method for detecting the simplest flux vectors **v** fulfilling relations (2) and (19) with all reactions assumed to be irreversible. Generalizing the approach in that both reversible and irreversible reactions are allowed for, this has led to the concept of elementary flux modes (Schuster and Hilgetag, 1994).

Elementary flux modes are defined as follows (Heinrich and Schuster, 1996, cf.): An elementary flux mode, $\mathcal{M}$, is defined as the set

$$\mathcal{M} = \{\mathbf{v} \in R^n | \, \mathbf{v} = \lambda \, \mathbf{v}^*, \lambda > 0\}, \tag{20}$$

where $\mathbf{v}^*$ is a n- dimensional vector (unequal to the null vector) fulfilling the following two conditions.

(a) Steady-state condition. $\mathbf{N}\,\mathbf{v}^* = \mathbf{0}$.

(b) Sign restriction.

If the system involves irreversible reactions, then the corresponding subvector $\mathbf{v}^{irr}$ of $\mathbf{v}^*$ fulfils inequality (19). For any couple of vectors and (unequal to the null vector) with the following properties:

- $\mathbf{v}^{'}$ and $\mathbf{v}^{''}$ obey restrictions (a) and (b),
- both $\mathbf{v}^{'}$ and $\mathbf{v}^{''}$ contain zero elements wherever $\mathbf{v}^*$ does, and they include at least one additional zero component each, $\mathbf{v}^*$ is not a nonnegative linear combination of $\mathbf{v}^{'}$ and $\mathbf{v}^{''}$,$\mathbf{v}^* \neq \lambda_1\mathbf{v}^{'} + \lambda_2\mathbf{v}^{''}$, $\lambda_1, \lambda_2 > 0$.

The last condition formalizes the concept of genetic independence introduced by Seressiotis and Bailey (1988). The condition says that a decomposition into two other modes should not involve additional enzymes.

Elementary modes have been determined in a number of biochemical networks, such as the synthesis of precursors of aromatic amino acids (Liao et al., 1996), the tricarboxylic acid cycle and adjacent pathways (Schuster et al., 1999) and glycolysis and alternative pathways in bacteria (Dandekar et al., 1999).

While the promise is substantial, the value of the topological analysis will not be fully utilized until algorithms are developed capable of tackling large metabolic, signaling or gene networks efficiently. What are the limits of the known algorithms in the calculation of elementary flux modes when applied to large systems?

Most of the algorithms applied to biological systems so far are bottom up approaches (Schuster et al., 2000b; Schilling et al., 2000). Initially, the stoichiometric matrix is augmented with the identity matrix. Next, a consecutive computation of matrices through combination of rows is performed until the stoichiometric matrix only contains zero elements. Obviously, this is a consecutive approach where all elementary modes are created at the final step. In addition, a large number of interim solutions are computed which finally disappear. By consequence, since even modest network sizes showing a larger degree of freedom might feature thousands of elementary flux modes, the problem easily becomes computationally intractable. Nevertheless, (Schilling and Palsson, 2000) have recently tackled the problem of prediction of the so called extreme pathways for *Haemophilus influenza*, represented by a network of 461 reactions. The strategy applied rests upon a decomposition of the network into subsystems to get a picture of the structural information for the entire system.

Alternative approaches make use of top down strategies considering the entire system in a more direct way (Happel and Sellers, 1989). As a result, successive solutions containing the complete network information are created. Mauch (to be published) developed an algorithm performing combinations of the base vectors of the null space in pairs. Figure 4 shows an example of the application of this algorithm for a system with 141 reactions in which the polymerization reactions for formation of biomass—and therefore growth—has been taken into account for the prediction of the elementary flux modes.

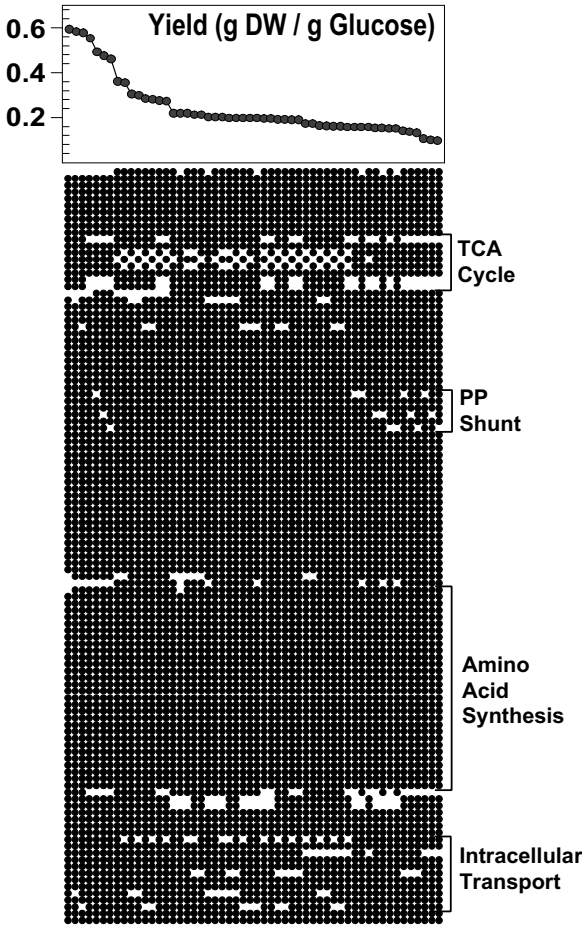

Figure 4: Elementary flux modes in *S. cerevisiae*. Explanation see text.

Each column shown in Figure 4 represents an elementary flux mode utilizing glucose as the sole carbon and energy source while producing biomass. Open symbols represent deleted genes for various reactions in the network. Obviously, these different mutants—or different phenotypes—are still able to grow albeit with varying yield coefficients. Elementary modes given in Figure 4 are ordered in descending sequence with respect to the yield of biomass on glucose. The maximal value of the biomass yield is in agreement to what has been measured for *S. cerevisae* growing with glucose under aerobic conditions. For mutants with defects in the transport systems via the mitochondrial membrane, for example, the biomass yield decreases to a value observed under anaerobic conditions. Thus, structural properties of the network clearly define the upper limit of product yields. Interestingly, these findings are independent of the kinetic properties of the network. Absolute values of fluxes as well as a dynamic response to system perturbations, however, can only be described when kinetic rate equations have been assigned.

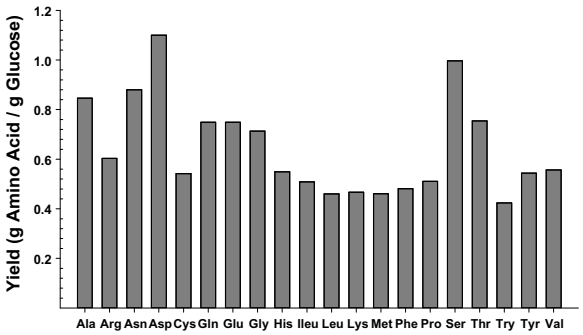

Figure 5: Maximal yields of amino acids on glucose in *S. cerevisiae*.

By a closer inspection of the solution shown in Figure 4, it is possible to separate more flexible regions of the network from rigid parts. For instance, we cannot delete any reaction in the pathways leading to amino acids. This is a reasonable finding for growth in the presence of synthetic media. In contrast, greater flexibility is observed in the citric acid cycle (TCA), pentose phosphate (PP) shunt and intracellular transport. It is interesting to recognize that these variations also have been identified as key modulations during the process of evolution.

The design aspect of these predictions may be better elaborated by examples of biotechnological relevant product formation. A desired product can often be synthesized by various different routes or pathways within the network. It is then of interest to detect the route on which the product/substrate ratio is maximum. Figure 5 illustrates an example showing the maximal yield of the individual 20 amino acids on substrate glucose in *S. cerevisiae*.

The results shown in Figure 5 have been obtained by detecting all elementary modes in the system and then choosing the mode giving the best yield. Sometimes, however, implementation of an elementary mode leading to a slightly non-optimal yield might be easier from a practical point of view. In contrast to linear optimization, the complete set of elementary flux modes immediately provides the complete spectrum of alternative implementations.

Flux Optimization

The driving force for selection of an optimal pathway is the maximization of the yield of the product. However, economic considerations also require optimization of the product formation rate (productivity). This problem leads to the question of an optimal modulation of enzyme activities in metabolic networks.

At first glance, the example chosen for this discussion—ethanol production with the yeast *S.*

cerevisiae—appears to be rather boring because it has been tackled so many times by yeast geneticists. The primary aim of these empirical attempts to modulate (mostly amplify) the key enzymes within glycolysis is the maximization of ethanol production rate which correlates with carbon dioxide evolution and, in turn, with the baking power of yeast. While addressing this problem it should be emphasized that the interest in answering the question of an optimal redistribution of enzyme activities is much broader. Generally, knowledge of the rate controlling steps in the central metabolism (glycolysis and PP shunt) is of central importance for cell cultures used for producing proteins as well as for the analysis of potential targets in cancer cells. Another field of interest is the identification of potential targets for antitrypanosomal drugs, important for treatment of the african sleeping sickness (Bakker et al., 1999). The last named authors concluded that "Despite the great interest, it is not yet known completely for any organism how the control of the glycolytic flux is distributed".

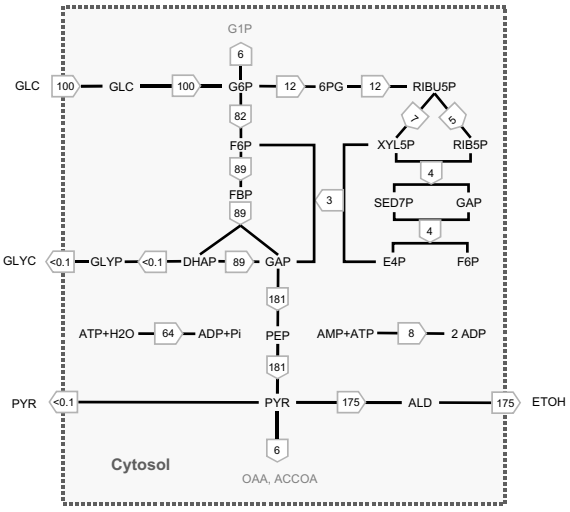

Figure 7: Flux distribution within central metabolic pathways of *S. cerevisiae* under anaerobic conditions at a specific growth rate of $\mu = 0.1\,h^{-1}$. All fluxes are related to the influx of glucose.

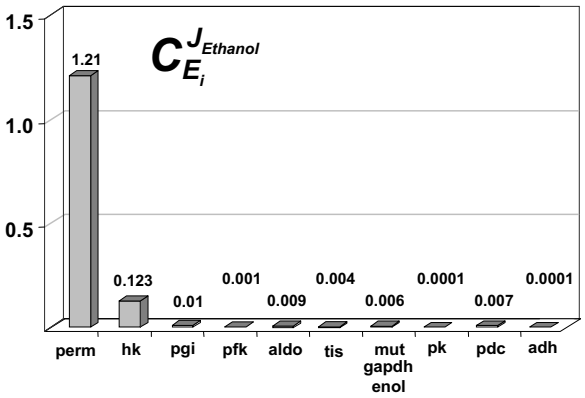

Figure 8: Distribution of flux control coefficients on ethanol formation route. Enzymes involved: Hexose transporter (perm), hexokinase(hk), phosphoglucose isomerase (pgi), phosphofructokinase (pfk), aldolase (aldo), triose phosphate isomerase (tis), gap dehydrogenase (gapdh), phospoglyceromutase (mut), enolase (enol), pyruvate kinase (pk), pyruvatedecarboxylase (pdc) and alcohol dehydrogenase (adh).

Dynamic Model

Similar to the strategy of Bakker et al. (1999), the problem can be approached from the basis of experimentally determined kinetic properties of the key enzymes which are then aggregated to a dynamic model. Individual rate expressions including their kinetic parameter have been identified *in vivo* by a stimulus response methodology (Theobald et al., 1997; Rizzi et al., 1997; Vaseghi et al., 1999): A pulse of glucose or alternative stimuli are introduced into a continuous culture operating at steady state and the transient response of several intracellular metabolite and cometabolite pools is experimentally determined in time spans of seconds or, recently, also milliseconds (Buziol et al., to be published). Within these relatively short time spans, enzyme concentrations are considered to be in a "frozen" state. Figure 6 summarizes examples for some of the experimental observations of metabolites and cometabolites from the yeast *S. cerevisiae* growing under anaerobic conditions.

The metabolome's response due to dynamic system excitation has been used to identify the dynamic system behavior by a stepwise internalization of metabolites similar to the method proposed by Rizzi et al. (1997). To describe the dynamic system behavior, deterministic kinetic rate equations for the pathways for the reactions have been formulated.

The general form of this rate equations can be written as

$$r_i = r_{max,i}\, f\left(\mathbf{c}, \mathbf{p}\right), \tag{21}$$

where the maximal rate (capacity) $r_{max,i}$ is obtained from the vector of model parameters $\mathbf{p}$, the vector comprising metabolite, cometabolite and effector concentrations $\mathbf{c}$ and the flux distribution $\mathbf{J}$ at the systems's steady state (e.g. $\mu = 0.1\,h^{-1}$); accordingly

$$r_{max,i} = 1/r_i^{steady\ state}\ f\left(\mathbf{c}^{steady\ state}, \mathbf{p}\right). \tag{22}$$

Intracellular flux distribution has been estimated by experimentally determined uptake and excretion rates of glucose, carbon dioxide, ethanol, glycerol and biomass. The results are documented in Figure 7.

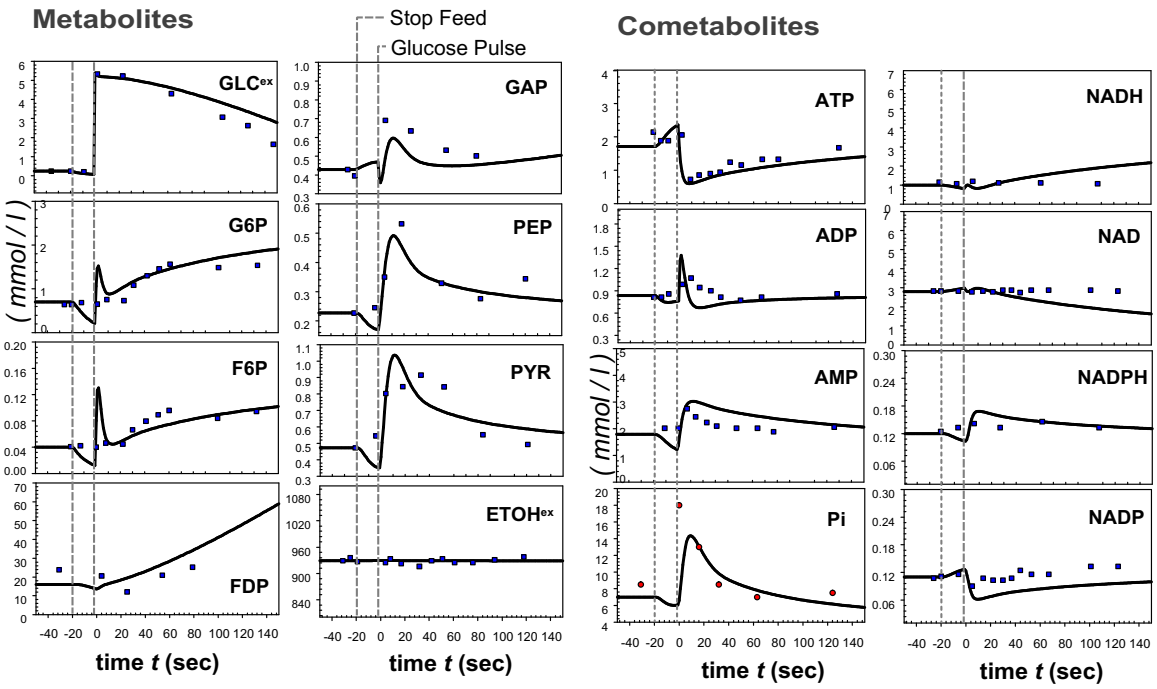

Figure 6: Comparison between model simulation and measured concentrations of glycolytic metabolites and cometabolites after dynamic system excitation. Data shown left from the broken lines represent steady state values at a growth rate of $\mu = 0.1\,h^{-1}$.

Sensitivity Analysis

The next step towards solving the envisaged design problem is to calculate the so called flux control coefficients, or—in the terminology of engineering—sensitivity coefficients. The flux control coefficient $C^J_{E_i}$ has been defined as the fractional change of the network flux J caused by a fractional change in the level of enzyme activity E_i. Thus,

$$Flux\ Control\ Coefficient = \frac{dJ\,(E_i)}{dE_i}. \qquad (23)$$

or normalized

$$C^J_{E_i} = \frac{dJ/J}{dE_i/E_i} = \frac{d\,lnJ\,(E_i)}{d\,lnE_i}. \qquad (24)$$

Figure 8 depicts the results of these calculations for those enzymes involved on the path from glucose to ethanol.

Since only a subset of the flux control coefficients with respect to ethanol formation are shown in Figure 8 (excluding, for example, the coefficients of enzymes involved in the PPP shunt), flux coefficients of this subset do not necessarily sum up to one. From the hierarchy of sensitivities it can be concluded that the enzyme responsible for the transport of glucose via the cell membrane (permease) shows the overwhelming control strength upon the ethanol production rate. Consequently, amplification of this enzyme should lead to an increased flux from glucose to ethanol.

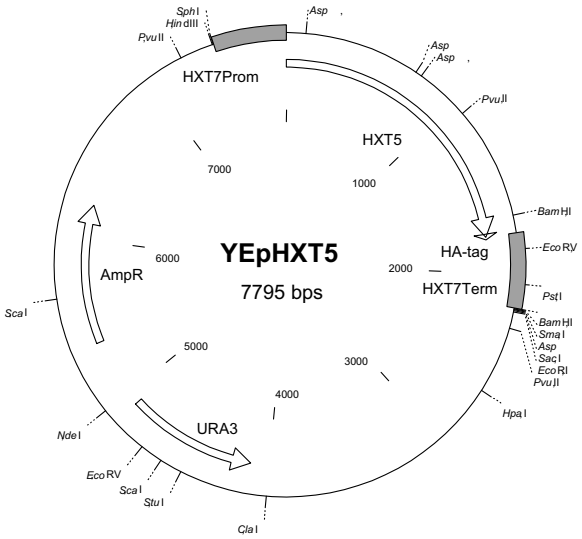

Figure 9: HXT5 multi copy plasmid.

Synthesis—Amplification of Hexose Transporters

Experimental verification of the above-named design proposal has been performed in a collaboration project with Institute of Microbiology from the University of

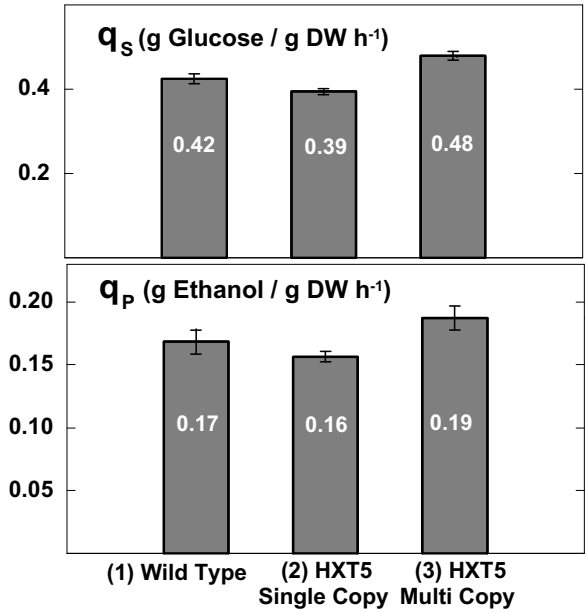

Figure 10: Specific substrate uptake rate q_s and specific product formation rate q_p during anaerobic chemostat cultivation. (1) Wild strain. (2) HXT5 with a single copy of the gene integrated the chromosome and constitutively expressed. (3) Multi copy plasmid with HXT5.

Düsseldorf (Dr. Boles). Within this project, the kinetics of the three most important hexose transporters out of the 17 transporters identified from the yeast genome project are investigated (Boles et al., 1997; Buziol et al., to be published). One of these s—HXT5—has been also expressed with a multi copy plasmid, shown in Figure 9 (Boles, to be published).

The transporter gene in HXT5 is flanked by an enhanced HXT7 promoter and HXT7 terminator, respectively. HXT7 is a high affinity transporter expressed at low glucose concentrations. The construct shown in Figure 9 ensures expression of HXT5 at target growth conditions. The same transporter has been integrated as single copy in the chromosome of a yeast strain in which all the genes of the other transporters has been knocked out (Wieczorke et al., 1999). As a result of these genetic constructions, it is possible to compare the flux of glucose through the glycolysis between two strains differing only in the amount of the hexose transporter. According to the hierarchy of flux control coefficients discussed in context with Figure 8, one would expect a noticeable increase of ethanol flux.

Experiments with three different strains were performed in continuous cultures at a dilution rate of $D = 0.07\,h^{-1}$ (Buziol et al., to be published). The results of these experiments are summarized in Figure 10.

Compared to the wild type, ethanol excretion rate and substrate uptake rate of the strain with higher transporter activity have found to be 10% and 25% higher, respectively. The discussion of the relevance of these results from industrial point of view is beyond the horizon of this paper. However, the rather modest effect prompts the question: Are there any other alternative design possibilities resulting in a substantial increase of the ethanol production rate?

Objective Function

The apparent failure to produce significant increase of the glycolytic flux points to the fundamental question if the underlying assumptions leading to the design suggestion are adequate. Keeping in mind that the pathway of interest is part of a whole—the living cell—, the idea that one need to amplify a single or multiple enzymes according to the hierarchy of flux control coefficients may not correspond to physiological reality. There are two aspects that should attract attention. First, an increased expression of enzymes is linked with energetically expensive protein synthesis. Glycolytic enzymes in yeast are known to contribute in the order of 30% to the total amount of cellular proteins. Thus, it seems to be likely that overexpression of these enzymes result in a stress situation with an unforeseeable impact on cell physiology. Instead of a single or simultaneous elevation of enzyme activities, a more robust strategy should try to keep the total concentration of proteins at a constant value and redistribute the activities according to the required objective. The optimization problem may then be stated as

$$\text{Maximize } J\left(\mathbf{r}_{max}^{Path}\right) \tag{25}$$

subject to

$$\frac{1}{w}\sum_{i=1}^{w}\frac{r_i^{max}}{r_{i,reference}^{max}} \leq \Omega. \tag{26}$$

In writing Equation 26 we assume that the maximal rate r_i^{max} is proportional to the enzyme amount. Therefore, Equation 26 specifies a fixed level for the total enzyme activity Ω.

Another relevant issue concerns the pool concentration of the metabolites within the cell. Attempts to increase metabolic fluxes by changes in individual enzyme concentrations may lead to substantial changes in metabolite concentrations. Again, a substantial change in metabolite concentrations either proves to be cytotoxic or at least leads to an undesired flux diversion (Kell and Mendes, 2000). Therefore, preservation of the metabolite concentrations close to the steady-state values of the wild strain is at any rate desirable to meet the basic property of well established metabolic systems for which the metaphor *homeostasis* has been coined (Reich and Selkov, 1981). It is also known that part of an optimal performance of cells with respect to the use of, for example, energy and carbon sources is a process of adaptation leading to a change of structure to control homeostasis. Interestingly, such structural changes may

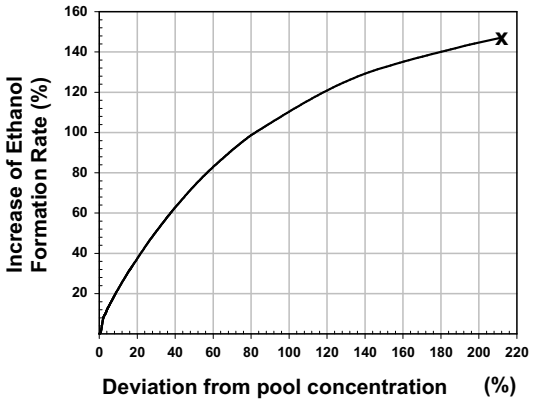

Figure 11: Increase of the Ethanol formation rate as a function of the maximal deviation from initial pool concentrations.

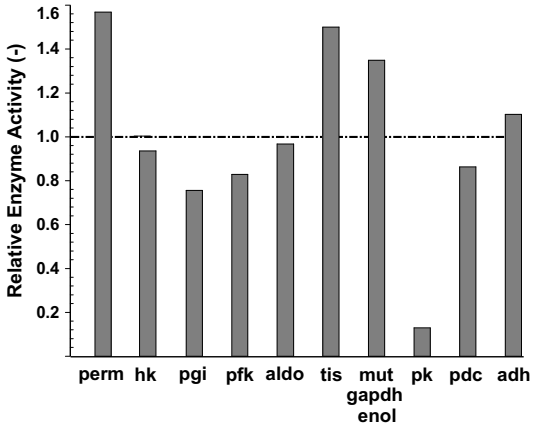

Figure 12: Optimal distribution of enzyme activities. All activities are related to their initial values.

include a change of enzyme concentrations via regulation of enzyme synthesis. Hence, homeostasis puts a further constraint on the metabolic redesign. Mathematically, this could take the form

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\left|c_{i,optimum}^{steady\ state}-c_{i,reference}^{steady\ state}\right|}{c_{i,reference}^{steady\ state}}\leq\Theta. \qquad (27)$$

Finally, the optimum must be constraint to stable steady states, leading to condition (28)

$$\frac{d}{dt}\mathbf{c}=\mathbf{0}\ \text{ and }\ Re\left(\lambda_{i,optimum}\right)\leq 0, \qquad (28)$$

with $Re\left(\lambda_i\right)$ denoting the real part of the system's eigenvalues.

Optimal Solutions

Retaining the total activity at the system's initial state at $\Omega = 1$, non-linear optimization of enzyme activities yielded significantly higher ethanol formation rates (see Figure 11.

As shown in Figure 11, amplification of ethanol formation largely depends on the allowed deviation from the initial pool concentrations Θ. Since the ratios of enzyme levels on the elementary flux mode glucose—ethanol are subject to modifications, no amplification can be expected for $\Theta = 0\%$ while the maximal possible amplification of ethanol formation is found to be as much as 144% at $\Theta = 210\%$. No stable steady states have been detected above this value.

Figure 12 shows the optimal distribution of enzyme activities for the selected example, 40% increase of pool concentrations resulting in an amplification of ethanol formation of 63%.

The optimized modulation of the enzyme activities results in an unexpected and interesting redistribution of enzyme activities. Due to its large share in the control of ethanol formation rate, the activity of the hexose transporter (perm) is found to be significantly higher in the optimized metabolic system. In contrast, even though the control coefficients of tis, gapdh, mut, enol and adh are relatively small, activities of those enzymes are also found to be larger compared to the non-optimized system. This outcome can be explained by the enzymes ability of counteracting an increase of pool concentrations provoked by a risen glucose influx effectively. Obviously, this is the result of the superposition of the three objectives: maximization of flux at more or less homeostatic conditions and unchanged total amount of enzymes.

Concluding Remarks

This paper has presented typical examples of computer aided design problems in Metabolic Engineering. The examples shown refer to the two different characteristics of biological systems: (1) Topological properties and (2) Kinetic properties of the individual reactions.

Topological analysis of metabolic networks turns out to be of immense value for relating genotypes and phenotypes. The further application of the illustrated concept of elementary flux modes critically depends on the development of new and more effective algorithms to treat larger networks. In case of biotechnological production processes, the most important application concerns the prediction of optimal topological properties for maximizing the product yield.

The second example illustrates model based design of an enhanced product formation rate. In addition to information on the network topology, the suggested solution of this important task in Metabolic Engineering requires detailed knowledge on kinetic rate expressions. Armed with a dynamic model for the most important part of the system, it is then possible to apply non-linear

optimization methods. The advantages of this approach are twofold: (1) Effects on large changes in enzyme concentrations can be easily studied and (2) Constraints such as limits on the total amount of enzymes and deviations from steady state metabolite pool concentrations can be taken into account in parallel.

References

Bailey, J. E., "Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities," *Biotechnol. Prog.*, **14**, 8–28 (1998).

Bakker, B. M., F. R. Michels, F. R. Opperdoes, and H. V. Westerhoff, "What controls glycolysis in bloodstream from *Trypanosoma brucei*?," *J. Biol. Chem.*, **274**, 14551–14559 (1999).

Christensen, B. and J. Nielsen, "Metabolic network analysis of *Penicillium chrysogenum* using 13C-labeled glucose," *Biotech. Bioeng.*, **68**, 652–659 (2000).

Dandekar, T., S. Schuster, B. Snel, M. Huynen, and P. Bork, "Pathway alignment: application to the comparative analysis of glycolytic enzymes," *Biochem. J.*, **343**, 115–124 (1999).

Dauner, M. and U. Sauer, "GC-MS analysis of amino acids rapidly provides rich information for isotopomer balancing," *Biotechnol. Prog.*, **16**, 642–649 (2000).

Edwards, J. S. and B. O. Palsson, "Systems properties of the *Haemophilus influenzae* Rd metabolic genotype," *J. Biol. Chem.*, **274**(25), 17410–17416 (1999).

Fell, D. A. and J. R. Small, "Fat synthesis in adipose tissue," *Biochem. J.*, **238**, 781–786 (1986).

Happel, J. and P. H. Sellers, "The characterization of complex systems of chemical reactions," *Chem. Eng. Commun.*, **83**, 221–240 (1989).

Heinrich, R. and S. Schuster, *The regulation of cellular systems*. Chapman & Hall, New York (1996).

Kell, D. B. and P. Mendes, Snaphots of systems: Metabolic Control Analysis and biotechnology in the post-genomic era, In *Technological and medical implications of Metabolic Control Analysis*, pages 3–25. Kluwer, Dordrecht (2000).

Liao, J. C., S. Y. Hou, and Y. P. Chao, "Pathway analysis, engineering and physiological considerations for redirecting central metabolism," *Biotech. Bioeng.*, **52**, 129–140 (1996).

Marx, A., A. A. de Graaf, W. Wiechert, L. Eggeling, and H. Sahm, "Determination of the fluxes in the central metabolism of *Corynebacterium glutamicum* by Nuclear Magnetic Resonance Spectroscopy combined with metabolite balancing," *Biotech. Bioeng.*, **49**, 111–129 (1996).

Mauch, K., S. Vaseghi, and M. Reuss, Quantitative analysis of metabolic and signalling pathways in *Saccharomyces cerevisiae.*, In Schügerl, K. and B. Bellgardt, editors, *Bioreaction Engineering*, pages 435–477. Springer, Berlin (2000).

Palsson, B., "The challenges of in silico biology. Nat. Biotechnol," *Nat. Biotechnol.*, **18**, 1147–1150 (2000).

Reich, J. G. and E. E. Selkov, *Energy metabolism of the cel. A theoretical treatise*. Academic Press, New York (1981).

Rizzi, M., M. Baltes, U. Theobald, and M. Reuss, "*In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II Mathematical model," *Biotech. Bioeng.*, **55**, 592–608 (1997).

Savinell, J. M. and B. O. Palsson, "Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism," *J. Theor. Biol.*, **154**, 421–454 (1992).

Schilling, C. H. and B. O. Palsson, "Assesment of the metabolic capablitites of *Haemophilus influenza* Rd through genome-scale pathway analysis.," *J. Theor. Biol.*, **203**, 249–283 (2000).

Schilling, C. H., S. Schuster, B. O. Palsson, and R. Heinrich, "Metabolic pathway analysis: basic concepts and scientific applications in the post genome era," *Biotechnol. Prog.*, **15**, 296–303 (1999).

Schilling, C. H., D. Letschert, and B. O. Palsson, "Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from pathway-oriented perspective," *J. Theor. Biol.*, **203**, 229–248 (2000).

Schmidt, K., J. Carlsen, J. Nielsen, and J. Villadsen, "Modeling isotopomer distribution in biochemical networks using isotopomer mapping matrices," *Biotech. Bioeng.*, **55**, 831–840 (1997).

Schmidt, K., J. Nielsen, and J. Villadsen, "Quantitative analysis of metabolic fluxes in *Escherichia coli*, using two-dimensional NMR spectroscopy and complete isotopomer models," *Biotech. Bioeng.*, **71**, 175–190 (1999).

Schuster, S. and C. Hilgetag, "On elementary flux modes in biochemical reaction systems at steady state," *J. Biol. Sys.*, **2**, 165–182 (1994).

Schuster, R. and S. Schuster, "Refined algorithm and computer program for calculating all non-negative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed," *CABIOS*, **9**, 79–85 (1993).

Schuster, S., T. Dandekar, and D. A. Fell, "Detection of elemntary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering," *Trends Biotechn.*, **17**, 53–60 (1999).

Schuster, S., D. A. Fell, and T. Dandekar, "A general definition of metabolic patheways useful for systematic organization and analysis of complex metabolic networks," *Nat. Biotechnol.*, **18**, 326–332 (2000b).

Schuster, S., T. Dandekar, K. Mauch, M. Reuss, and D. A. Fell, Recent developments in metabolic pathway analysis and their potential implications for biotechnology and medicine, In *Technological and medical implications of Metabolic Control Analysis*, pages 57–66. Kluwer, Dordrecht (2000a).

Seressiotis, A. and J. E. Bailey, "MPS: an artificially intelligent software system for the analysis and synthesis of metabolic pathways," *Biotech. Bioeng.*, **31**, 587–602 (1988).

Stephanopoulos, G. N., A. A. Aristidou, and J. Nielsen, *Metabolic engineering*. Academic Press, New York (1998).

Szyperski, T., R. W. Glaser, M. Hochuli, and J. Fiaux, "Bioreaction network topology and metabolic flux ratio analysis by biosynthetic fractional 13C labeling and two-dimensional NMR spectroscopy," *Metabolic Engineering*, pages 189–197 (1999).

Theobald, U., W. Mailinger, M. Baltes, M. Rizzi, and M. Reuss, "*In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*: I. Experimental observations," *Biotech. Bioeng.*, **55**, 305–316 (1997).

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

Vanderbei, R. J., *Linear programming: foundations and extensions*. Kluwer, Boston (1998).

Vaseghi, S., A. Baumeister, M. Rizzi, and M. Reuss, "*In vivo* dynamics of the pentose phosphate pathway in *Saccharomyces cerevisiae*," *Metabolic Engineering*, **1**, 128–140 (1999).

Wiechert, W. and A. A. de Graaf, "Bidirectional reaction steps in metablic networks: I. Modeling and simulation of carbon isotope labeling experiments," *Biotech. Bioeng.*, **55**, 101–117 (1997).

Wiechert, W., C. Siefke, A. A. de Graaf, and A. Marx, "Bidirectional reaction steps in metablic networks: II. Flux estimation and statistical analysis," *Biotech. Bioeng.*, **55**, 118–134 (1997).

Wieczorke, R., S. Krampe, W. T., K. Freidel, C. P. Hollenberg, and E. Boles, "Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexose in *Saccharomyces cerevisiae*," *FEBS Letters*, **464**, 123–128 (1999).