Calculability Analysis in Underdetermined Metabolic Networks Illustrated by a Model of the Central Metabolism in Purple Nonsulfur Bacteria

Steffen Klamt,[1] Stefan Schuster,[2] Ernst Dieter Gilles[1]

[1]Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstrassse 1, D-39106 Magdeburg, Germany, e-mail: klamt@mpi-magdeburg.mpg.de
[2]Department of Bioinformatics, Max Delbrück Center for Molecular Medicine, D-13092 Berlin-Buch, Germany; e-mail: stschust@mdc-berlin.de

Received 27 December 2000; accepted 27 September 2001

Abstract: Metabolite balancing has turned out to be a powerful computational tool in metabolic engineering. However, the linear equation systems occurring in this analysis are often underdetermined. If it is difficult or impossible to find the missing constraints, it is nevertheless feasible in some cases to determine the values of a subset of the unknown rates. Here, a procedure for finding out which reaction rates can be uniquely calculated in underdetermined metabolic networks and computing these rates is given. The method is based on the null space to the stoichiometry matrix corresponding to the reactions with unknown rates. It is shown that this method is considerably easier to handle than an algorithm given previously (Van der Heijden et al., 1994a). Furthermore, a useful elementary representation of the null space is presented which is closely related with the elementary flux modes. This unique representation is central to a more general approach to observability/calculability analysis. In particular, it allows one to find, in an easy way, those sets of measurable rates that enable a calculation of a certain unknown rate. Besides, rates which are never calculable by metabolite balancing may be easily detected by this method.

The applicability of these methods is illustrated by a model of the central metabolism in purple nonsulfur bacteria. The photoheterotrophic growth of these representatives of anoxygenic photosynthetic bacteria is stoichiometrically analyzed. Interesting metabolic constraints caused by the necessary balancing of NAD(P)H can be detected in a highly underdetermined system. This is, to our knowledge, the first application of stoichiometric analysis to the metabolic network in this bacteria group using metabolite balancing techniques. A new software tool, the *FluxAnalyzer*, is introduced. It allows quantitative and structural analysis of metabolic networks in a graphical user interface. *Biotechnol Bioeng* **77**: 734–751, 2002; DOI 10.1002/bit.10153
Keywords: metabolic flux analysis; metabolite balancing; calculability; observability; underdetermined systems; stoichiometric analysis; purple nonsulfur bacteria; photosynthetic bacteria

INTRODUCTION

Stoichiometric analysis of metabolic networks has become an important tool for studying the capabilities of biochemical reaction systems in microorganisms, particularly in light of biotechnological applications (Edwards et al., 2001; Heinrich and Schuster, 1996; Mavrovouniotis et al., 1990; Schuster et al., 2000; Varma and Palsson, 1993). One method in this framework is the metabolic flux analysis (MFA) or, more specifically, metabolite balancing (Bonarius et al., 1997; Holms, 1986; Noorman et al., 1996; Rizzi et al, 1997; Savinell and Palsson, 1992; Stephanopoulos et al., 1998; Van der Heijden et al., 1994a,b). It uses mass balancing techniques and the assumption of metabolic steady state for the formulation of linear constraints. These constraints, together with measurements of extracellular excretion and uptake rates, allow the calculation of intracellular net reaction rates and, thus, the determination of metabolic flux distributions in a defined stoichiometric network.

Nevertheless, several complications may occur upon application of mass balancing techniques. The most restrictive fact is that, in underdetermined networks, complete flux distributions may not be computed because not enough measurements or constraints are available (e.g., Vallino and Stephanopoulos, 1993). Several suggestions have been made to overcome these problems and to find missing constraints (cf. Bonarius et al., 1997):

- Introduction of additional metabolic constraints, e.g., additional balancing of co-metabolites (NAD(P)H, ATP) or using knowledge of flux ratios such as P/O
- Usage of linear optimization techniques with appropriate objective functions
- Isotopic tracer experiments (e.g., by NMR technology or GC-MS)

Although the auxiliary constraints in the first suggestion are often used, they imply the problem that for full ATP balancing the requirement of ATP for maintenance has to be known. However, this can only be

roughly estimated because not all of the processes involved in ATP maintenance are known in detail. Separate balancing of NADH and NADPH is often complicated due to the presence of transhydrogenases. Moreover, NADH balancing in respiring organisms is sensitive to the estimated value of ATP maintenance.

Usage of optimization techniques raises the question of what the true biological objective function is. Frequently used optimality criteria are, for example, maximization of the NADH, NADPH, or ATP production (Meléndez-Hevia et al., 1997; Pfeiffer et al., 2001; Varma and Palsson, 1993) or the maximization of growth rate (Edwards and Palsson, 1999; Edwards et al., 2001). Maximization of ATP production is certainly not always relevant because otherwise, the Entner-Doudoroff pathway, which transduces less energy than glycolysis, would never be used (cf. Meléndez-Hevia et al., 1997). Moreover, in the maximization of ATP production, a trade-off between production rate and yield needs to be found (Pfeiffer et al., 2001). To some extent, objective functions such as ''maximize product yield,'' are also applied in metabolic engineering where the estimated theoretically optimal flux distribution is then the basis of genetic manipulations in the considered organism.

Isotopic tracer experiments carried out by using NMR technology or GC-MS have become a very important method for experimental quantification of the metabolism in underdetermined networks (cf. Wiechert et al., 1999). The additional information generated by isotopic tracer experiments is often sufficient for the problem to be no longer underdetermined. Unfortunately, however, only metabolites with sufficiently large concentrations can be measured, and this method is expensive and laborious and therefore not suitable for all problems, especially in industrial applications.

In this article it will be shown that even if the missing constraints cannot be found to obtain a determined system, underdetermined networks can contain very important information. The constraints can be analyzed to find pathways which are nevertheless quantitatively determined. These fluxes have the special property that their values are fixed, despite linear dependencies within the stoichiometry matrix or the lack of enough measurements.

We will first summarize calculation routines for underdetermined and/or redundant systems, based on the work by Van der Heijden et al. (1994a) and Stephanopoulos et al. (1998). Then, the suggested procedure for treating underdetermined systems is described and compared with a routine proposed by Van der Heijden et al. (1994a). In particular, we will show that our routine considerably simplifies the classification of unknown rates as being calculable or noncalculable. A more general approach for analyzing calculability in metabolic networks is presented based on a suitable elementary representation of the null space to the stoichiometry matrix.

An interactive graphical program, called *Flux Analyzer*, is also presented, which allows the flexible and efficient usage of routines for all network types via a comfortable menu-controlled interface. Calculated flux distributions can be visualized in graphical flux maps.

Later we present an illustrative example of an underdetermined system, notably the metabolic network in purple nonsulfur bacteria, which are representatives of anoxygenic photosynthetic bacteria. These versatile organisms are capable to grow aerobically in the dark or photosynthetically under anaerobic conditions in the light (both autotrophically or heterotrophically). It is shown that particularly under photosynthetic growth conditions their special stoichiometric network is worth being analyzed even if no measurements of extracellular uptake or excretion rates are available. Interesting conclusions concerning the balancing of reducing equivalents in these bacteria can be drawn.

METABOLITE BALANCING AND CALCULABILITY ANALYSIS

Principles of Metabolite Balancing

Metabolite balancing (MB) is part of metabolic flux analysis, which is, in addition, comprised of the above-mentioned optimization approaches (cf. Edwards and Palsson, 1999; Edwards et al., 2001) as well as isotopomer balancing (Wiechert et al., 1999). Metabolic balancing is widely used for quantification of metabolic fluxes in cell cultures, whose metabolism is in steady state. The great advantage of MB is that no kinetic enzyme parameters are needed. Its fundamentals and the derivation of the linear constraints are briefly described below and can be studied in detail elsewhere (Savinell and Palsson, 1992; Stephanopoulos et al., 1998). The first step in MB is the construction of the metabolic network, which contains the most important metabolites and reactions between them:

- m metabolites (nodes in the network) with concentrations $\boldsymbol{c} = (c_1,c_2,\ldots,c_m)$.
- q reactions with net reaction rates $\boldsymbol{r} = (r_1,r_2,\ldots,r_q)$ in [mmol/(g dry weight · h)]. The growth rate μ is to be included in $\boldsymbol{r}$ as well, if biomass synthesis is considered.
- **N:** stoichiometry matrix with dimension $[m \times q]$, where the element n_{ij} is the stoichiometric coefficient of metabolite i in reaction j.

The central assumption of steady state [$\mu(t)$, $\boldsymbol{r}(t)$, $\boldsymbol{c}(t)$ are constant] and neglecting the dilution of metabolites due to growth leads to a homogeneous system of m linear equations

$$0 = \frac{\mathrm{d}c_i(t)}{\mathrm{d}t} \cong \sum_{j=1}^{q} n_{ij} r_j \qquad (1)$$

This can be written in matrix notation as

$$\boldsymbol{0} = \mathbf{N}\boldsymbol{r} \tag{2}$$

As the number of reactions q is usually greater than the rank of the stoichiometry matrix, the system is underdetermined if no reaction rates have been measured. A complete set of linearly independent solutions of equation (2) form the null space matrix **K**, whose columns span the null space (kernel) *ker*(**N**) of **N** (cf. Reder, 1988)

$$\boldsymbol{0} = \mathbf{N}\mathbf{K} \tag{3}$$

Note that, if **K** exists, there are infinitely many representations of **K**, because its columns, $\boldsymbol{k}_i$, can be linearly combined with each other. Also, all possible nontrivial fluxes $\boldsymbol{J}$ that fulfill the steady-state condition (2) are linear combinations of the columns $\boldsymbol{k}_i$

$$\boldsymbol{J} = \mathbf{K}\mathbf{a} = \sum_i a_i \boldsymbol{k}_i \tag{4}$$

The number of columns (called the degrees of freedom in **K** is q–*rank*(**N**). The stoichiometry matrix **N** can be considered to have full row rank: *rank*(**N**) = m. Possible dependencies between the rows can easily be removed by deleting those rows that cause the linear dependencies; they are of no importance for equation (3). It then follows that the dimension of **K** is $[q \times (q-m)]$.

Several interesting results in stoichiometric analysis can be obtained by analysis of the null space without information about kinetic parameters. For example, the null space is a useful concept in metabolic control analysis (Reder, 1988; cf. Heinrich and Schuster, 1996). Moreover, elementary flux modes can be computed, that is, minimal sets of enzymes that fulfill the steady-state condition (2) and irreversibility restrictions (Schuster and Hilgetag 1994; Schuster et al., 2000). Elementary flux modes are useful to detect essential structural features of the biochemical network.

Equation (2) is usually highly underdetermined due to a considerable number of degrees of freedom. Only in a few cases, some reaction rates can be exactly determined by the analysis of **K** alone. Any rate for which the corresponding row of **K** contains zeros only, must be zero, because of equation (4). These reactions, which are called strictly detailed balanced reactions (Schuster and Schuster, 1991) can be eliminated from the network in the further analysis, because their rate is a priori known and is not needed to compute any other rates.

Figure 1 shows a simple example. The hypothetical network consists of five metabolites and seven reactions. The incoming and outgoing fluxes of the network are marked as dashed arrows. The stoichiometry matrix **N** for this network as well as the resulting null space matrix **K** can easily be formulated:

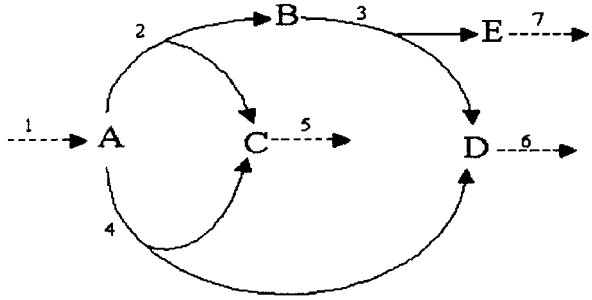

Figure 1. Network example 1.

$$\mathbf{N} = \begin{pmatrix} 1 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}, \mathbf{K} = \begin{pmatrix} 4 & 0 \\ 1 & 1 \\ 1 & 1 \\ 3 & -1 \\ 4 & 0 \\ 4 & 0 \\ 1 & 1 \end{pmatrix}$$

In this example no reaction rate is determined a priori, because no row of **K** is completely zero. If, for example, a metabolite F and a reaction B $\leftrightarrow$ F were included, the null space matrix would be augmented by a null row corresponding to this reaction. Accordingly, this reaction would be strictly detailed balanced.

Metabolic balancing is based on the knowledge of some fluxes, usually the fluxes of the reactions going into, and out of the cell. Well-defined assumptions (for instance, respiration under anaerobic conditions is zero) further decrease the degrees of freedom. The vector of reaction rates can be partitioned into the known part, $\boldsymbol{r}_b = (r_{b_1} \ldots r_{b_k})$, and unknown part, $\boldsymbol{r}_n = (r_{n_1} \ldots r_{n_u})$, possibly after rearranging the components. In the same way the matrix **N** is partitioned into $\mathbf{N}_b$ and $\mathbf{N}_n$:

$$\boldsymbol{r} = \begin{pmatrix} \boldsymbol{r}_b \\ \boldsymbol{r}_n \end{pmatrix}, \quad \mathbf{N} = (\mathbf{N}_b, \mathbf{N}_n)$$

If no rates are known (unknown), then $\mathbf{N}_b$ ($\mathbf{N}_n$) is set to the null vector of dimension m. The steady-state equation (2) can now be written in the following form:

$$\boldsymbol{0} = \mathbf{N}\boldsymbol{r} = \mathbf{N}_b\boldsymbol{r}_b + \mathbf{N}_n\boldsymbol{r}_n \Leftrightarrow \mathbf{N}_n\boldsymbol{r}_n = -\mathbf{N}_b\boldsymbol{r}_b \tag{5}$$

Equation (5) is the fundamental relation in metabolite balancing. Note that a similar equation with constant terms on the right-hand side is obtained when one or more time derivatives $\mathrm{d}c_i/\mathrm{d}t$ in equations (1) are set equal to a non-zero constant.

The aim is the determination of preferably all unknown rates of $\boldsymbol{r}_n$ and the utilization of possibly occurring redundancies for detection of measurement or modeling errors. Equation (5) is the starting point for classifying the defined system and the elements of $\boldsymbol{r}_b$ and $\boldsymbol{r}_n$. The results of this classification affect the decision as to which routines have to be used for further calculations.

Balanceability and Calculability: Classification of Systems and Rates

The following section summarizes the classification of systems and rates and is based on the work of Van der Heijden et al. (1994a). Some modifications have been made to simplify the classification procedure.

Relation (5) allows the assignment of two classification properties to the complete network, both of which depend only on the rank of $\mathbf{N}_n$ (u denotes the number of unknown rates):

Determinacy:

- *Underdetermined:* $rank(\mathbf{N}_n) < u$; there are not enough linearly independent constraints for computing all rates of $\boldsymbol{r}_n$ uniquely.
- *Determined:* $rank(\mathbf{N}_n) = u$; there are enough linearly independent constraints for computing all rates of $\boldsymbol{r}_n$ uniquely.

Redundancy:

- *Redundant:* $rank(\mathbf{N}_n) < m$; that means that some rows in $\mathbf{N}_n$ can be expressed as linear combinations of other rows. This can lead to an inconsistent system if the vector $\boldsymbol{r}_b$ contains such values that no $\boldsymbol{r}_n$ exists that exactly solves equation (5). Below it is shown how, for any given $\boldsymbol{r}_b$, consistency can be checked in a redundant system.
- *Not redundant:* $rank(\mathbf{N}_n) = m$; no linear dependencies between the rows occur; therefore, the system is automatically consistent for any $\boldsymbol{r}_b$.

The components in $\boldsymbol{r}_n$ and $\boldsymbol{r}_b$ can be classified by the following properties:

Calculability (Observability): A rate of $\boldsymbol{r}_n$ is called

- *calculable* if it can be *uniquely* computed using equation (5).
- *noncalculable* if it cannot *uniquely* be determined using equation (5).

Balanceability: A rate of $\boldsymbol{r}_b$ is called

- *Balanceable* if the consistency of the system (5) depends on the value of this rate. Because nonredundant systems are consistent for every $\boldsymbol{r}_b$, balanceable rates occur only in redundant systems.
- *Nonbalanceable* if the consistency of system (5) is independent of this rate. (Hence, all rates are nonbalanceable in a nonredundant system).

A special case arises if the system is *exactly* (*fully*) *determined*, that is, it is determined but not redundant ($rank(\mathbf{N}_n) = u = m$). This is the easiest case because all unknown rates can easily be determined via the inverse of the then square nonsingular matrix $\mathbf{N}_n$:

$$\boldsymbol{r}_n = -\mathbf{N}_n^{-1}\mathbf{N}_b\boldsymbol{r}_b \tag{6}$$

Unfortunately, this is a very rare case. To treat all classes of systems, where balanceable and/or noncalculable rates may also occur, the more general least-squares solution of equation (5) is used:

$$\boldsymbol{r}_n = -\mathbf{N}_n^{\#}\mathbf{N}_b\boldsymbol{r}_b \tag{7}$$

where $\mathbf{N}_n^{\#}$ is the $u \times k$ – Penrose pseudo-inverse (Ben-Israel and Greville, 1974; Strang, 1980) which can be computed, for example, by MATLAB (command *pinv*). Note that its definition is more general than the definition via the product of $\mathbf{N}_n$ and its transpose (as used, for example, in Stephanopoulos et al., 1998, Ch. 8). This general pseudo-inverse exists for any matrix and is essential for all further algorithms. It allows the computation of a least-squares solution of equation (5), where the unique parts of the solution (the calculable rates of $\boldsymbol{r}_n$) can be determined even if the system involves noncalculable rates.

Checking Balanceability and Consistency

Balanceable rates occur in equation (5) only if the system is redundant. In a redundant system, consistency is not given a priori. A consistent system fulfills the consistency conditions

$$\boldsymbol{0} = \mathbf{R}\boldsymbol{r}_b \tag{8}$$

where $\mathbf{R}$ is the redundancy matrix, which is easily constructed by equations (5) and (7) (cf. Van der Heijden et al., 1994a)

$$\mathbf{R} = \mathbf{N}_b - \mathbf{N}_n\mathbf{N}_n^{\#}\mathbf{N}_b \tag{9}$$

Only if the system is not redundant, the matrix $\mathbf{R}$ equals the null matrix and therefore equation (8) is automatically fulfilled. In redundant cases, the originally given or measured set of $\boldsymbol{r}_b$ does not normally fulfill equation (8) due to unavoidable measurement errors or modeling inaccuracy and, hence, the system is inconsistent. In such a case the least-squares solution (7) cannot exactly solve equation (5). Balanceable rates are "responsible" for potential inconsistencies. They can be detected by an inspection of matrix $\mathbf{R}$: Exactly those $\boldsymbol{r}_{b_j}$ are balanceable for which the corresponding j-th column of $\mathbf{R}$ contains at least one non-zero value. In contrast to nonbalanceable rates, which have no impact on the consistency of equation (5), balanceable rates should be adjusted (balanced) to obtain a consistent system in which the steady-state condition (2) is fulfilled. This procedure uses the variance–covariance matrix of the measurements and the reduced redundancy matrix as described in detail in Stephanopoulos et al. (1998, Ch. 8) and also allows the prediction of measurements or modeling errors by a consistency check.

Checking Calculability

The aim of MB is the calculation of as many components of $\boldsymbol{r}_n$ as possible. It is therefore very important to detect the calculable rates. It is noteworthy that an un-

derdetermined network can, in some cases, contain all types of rates: balanceable and nonbalanceable, calculable and noncalculable rates. Thus, a system can simultaneously be redundant and underdetermined. Therefore, the term "overdetermined" is ambiguous and misleading. (We propose to use this term only for a system which is redundant *and* determined.) These facts are often not taken in account.

In Van der Heijden et al. (1994a) the set of calculable rates was determined by using singular value decomposition (SVD). This method of calculability analysis was applied, for example, by Noorman et al. (1996). Singular value decomposition requires the calculation of three matrices, and the determination of calculable rates using this method seems not to be so straightforward. Here an approach that is much easier to justify and employ is introduced as follows. As pointed out above, the system (5) is underdetermined if

$$rank(\mathbf{N}_n) < u \tag{10}$$

(u: number of unknown rates = number of columns in $\mathbf{N}_n$). This is equivalent to the statement that the dimension of the kernel $\mathbf{K}_n = ker(\mathbf{N}_n)$, which equals u–$rank(\mathbf{N}_n)$, is greater than zero. If inequality (10) does not hold, then all rates are calculable. To detect the calculable rates in the underdetermined case, we consider the general solution (7) of equation (5) which is, in underdetermined systems, rather a solution space having the same dimension as $\mathbf{K}_n$:

$$\boldsymbol{r}_{n,s} = -\mathbf{N}_n^{\#}\mathbf{N}_b\boldsymbol{r}_b + \mathbf{K}_n\mathbf{a} = \boldsymbol{r}_{n,part} + \mathbf{K}_n\boldsymbol{a} \tag{11}$$

where $\boldsymbol{r}_{n,part}$ is a particular least-squares solution determined by equation (7); $\boldsymbol{a}$ is an arbitrary vector with u–$rank(\mathbf{N}_n)$ elements and reflects the indeterminacy of $\boldsymbol{r}_{n,s}$. It can easily be verified that for an arbitrary vector a the resulting vector $\boldsymbol{r}_{n,s}$ is a least-squares solution of equation (5) as is $\boldsymbol{r}_{n,part}$, because

$$\mathbf{N}_n\boldsymbol{r}_{n,s} = \mathbf{N}_n\boldsymbol{r}_{n,part} + \mathbf{N}_n\mathbf{K}_n\boldsymbol{a} = \mathbf{N}_n\boldsymbol{r}_{n,part} \tag{12}$$

Thus, $\mathbf{K}_n$ represents the degrees of freedom of the solution space. Those elements of $\boldsymbol{r}_{n,s}$ for which the value of $\mathbf{K}_n\boldsymbol{a}$ does not change for any $\boldsymbol{a}$, have fixed values. This is exactly the case for such elements in $\boldsymbol{r}_{n,s}$, for which the corresponding row in $\mathbf{K}_n$ is a null row. That is all one has to do for checking calculability:

Such elements $\boldsymbol{r}_{n_j}$ of $\boldsymbol{r}_n$ are determined and therefore calculable, whose corresponding (j-th row) in $\boldsymbol{K}_n$ is a null row.

The values of the detected calculable rates can be taken from $\boldsymbol{r}_{n,part}$, a special least-squares solution of equation (5). Noncalculable rates cannot be considered as determined by equation (7).

The detected set of calculable rates is the same as one would get by using the SVD as described in Van der Heijden et al. (1994a). However, the analysis of the null space matrix of $\mathbf{N}_n$ is more straightforward and the null space $\mathbf{K}_n$ can be determined more easily than the complete SVD. Computation of the null space is a standard routine implemented in many biochemical simulation programs such as GEPASI (Mendes, 1997), JARNAC (http://www.sys-bio.org), and METATOOL (Pfeiffer et al., 1999). Besides, in the next section, we will show that the calculability condition based on an inspection of the null space allows also a more general (global) calculability analysis.

In summary, the following steps are proposed to treat the network characterized by equation (5):

1. Balancing
   a. Determine whether the system is redundant by the condition rank ($\mathbf{N}_n$) $< m$
   b. If the system is not redundant, go on straight to Step 2.
      If the system is redundant: Determine the balanceable rates by analyzing the redundancy matrix **R**. Thereafter, preferably balance the rates to obtain a consistent system and make a consistency check of measurements and/or model.
2. Calculating rates
   a. Determine which rates of $\boldsymbol{r}_n$ are calculable by searching for null rows in the null space matrix $\mathbf{K}_n$.
   b. Take the values of the calculable rates from the least-squares solution (7).

Please note:

1. It is not mandatory to balance the balanceable rates and, thus, to construct a consistent system. Equation (8) does then use the original values of $\boldsymbol{r}_b$ for calculating the least-squares solution, which do generally not fulfill equation (5), due to inconsistencies in $\boldsymbol{r}_b$.
2. As mentioned above, in an exactly determined system, equation (6) can be used for calculating the complete set $\boldsymbol{r}_n$, because in this case, the pseudo-inverse is exactly the inverse of $\mathbf{N}_n$.
3. Underdetermined systems can also be treated by optimization routines to calculate the undetermined components of $\boldsymbol{r}_n$ (Edwards et al., 2001; Stephanopoulos et al., 1998; Varma and Palsson, 1993). For this purpose a linear objective function $\boldsymbol{f}$ is constructed which is used to find such $\boldsymbol{r}_n$ that minimize this objective function and fulfil the equation constraints (5) and perhaps additional inequality constraints. However, it should be checked whether the optimal solution is unique.
4. All described routines (including optimization) are implemented in the graphical application *Flux Analyzer*, which is introduced in a later section.

Examples of Calculability and Balanceability

For illustration, particularly of the new calculability check, we use the example in Figure 1, where we consider four possible scenarios:

Case 1: $\boldsymbol{r}_b$ is empty

No rates are determined a priori, because the matrix $\mathbf{K}_n$ then coincides with the null space matrix, $\mathbf{K}$, of the entire system, which has no null rows. The system is not redundant, because $\mathbf{N}_n = \mathbf{N}$ has full rank.

Case 2: $\boldsymbol{r}_b = (\boldsymbol{r}_1)$

If only $\boldsymbol{r}_1$ is measured or known, one gets the following matrices:

$$\mathbf{N}_b = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \boldsymbol{r}_n = \begin{pmatrix} \boldsymbol{r}_2 \\ \boldsymbol{r}_3 \\ \boldsymbol{r}_4 \\ \boldsymbol{r}_5 \\ \boldsymbol{r}_6 \\ \boldsymbol{r}_7 \end{pmatrix}$$

$$\mathbf{N}_n = \begin{pmatrix} -1 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix} \quad \mathbf{K}_n = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$$

Because $\mathbf{N}_n$ has full row rank ($= 5$) the system is not redundant ($\mathbf{R}$ is the null matrix). The system is underdetermined because $rank(\mathbf{N}_n) = 5 < 6 = u$. The analysis of $\mathbf{K}_n$ shows that $\boldsymbol{r}_5$ and $\boldsymbol{r}_6$ are calculable, because their corresponding rows in $\mathbf{K}_n$ are zero. Both have the same value as $\boldsymbol{r}_1$ as one could easily verify without using the pseudo-inverse.

Case 3: $\boldsymbol{r}_b = (\boldsymbol{r}_1,\boldsymbol{r}_7)^T$

The matrices now have the following structure:

$$\mathbf{N}_b = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & -1 \end{pmatrix}, \quad \boldsymbol{r}_n = \begin{pmatrix} \boldsymbol{r}_2 \\ \boldsymbol{r}_3 \\ \boldsymbol{r}_4 \\ \boldsymbol{r}_5 \\ \boldsymbol{r}_6 \end{pmatrix}$$

$$\mathbf{N}_n = \begin{pmatrix} -1 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{K}_n = ()$$

Because $rank(\mathbf{N}_n) = 5 = u = m$ this system is exactly determined and all unknown rates can be determined as in equation (6) by the inverse of $\mathbf{N}_n$, which also represents the pseudo-inverse. $\mathbf{K}_n$ is empty, because no degrees of freedom exist. If, for example, $\boldsymbol{r}_b = (\boldsymbol{r}_1,\boldsymbol{r}_7)^T = (1, 0.3)^T$ then the unique solution would be $\boldsymbol{r}_n = (\boldsymbol{r}_2, \boldsymbol{r}_3, \boldsymbol{r}_4, \boldsymbol{r}_5, \boldsymbol{r}_6)^T = (0.3, 0.3, 0.7, 1, 1)^T$.

Case 4: $\boldsymbol{r}_b = (\boldsymbol{r}_1, \boldsymbol{r}_5)^T$

Now, we obtain the following system:

$$\mathbf{N}_b = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \boldsymbol{r}_n = \begin{pmatrix} \boldsymbol{r}_2 \\ \boldsymbol{r}_3 \\ \boldsymbol{r}_4 \\ \boldsymbol{r}_6 \\ \boldsymbol{r}_7 \end{pmatrix}$$

$$\mathbf{N}_n = \begin{pmatrix} -1 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{pmatrix}, \quad \mathbf{K}_n = \begin{pmatrix} 1 \\ 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}$$

$$\mathbf{R} = \begin{pmatrix} 0.5 & -0.5 \\ 0 & 0 \\ 0.5 & -0.5 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Because $rank(\mathbf{N}_n) = 4$ is less than m, the system is redundant. Furthermore, the system is also underdetermined because the rank of $\mathbf{N}_n$ is smaller than the number of unknown rates u. The balanceable rates can be detected by analysis of the redundancy matrix $\mathbf{R}$, which is constructed via equation (9) using the pseudo-inverse. The rank of $\mathbf{R}$ is 1, therefore the degree of redundancy is 1 as well. Because the columns corresponding to $\boldsymbol{r}_1$ and $\boldsymbol{r}_5$ both have elements unequal zero, both rates are balanceable. The analysis of $\mathbf{K}_n$ shows that $\boldsymbol{r}_6$ is calculable. If we assume measured values $\boldsymbol{r}_b = (\boldsymbol{r}_1, \boldsymbol{r}_5)^T = (2,3)^T$ and that $\boldsymbol{r}_1$ and $\boldsymbol{r}_5$ have the same variance of measurement errors, then their balancing results in the estimated values $\tilde{\boldsymbol{r}}_b = (\tilde{\boldsymbol{r}}_1, \tilde{\boldsymbol{r}}_5) = (2.5,2.5)^T$ (representing now a consistent system). The value of $\boldsymbol{r}_6$ is then computed to be 2.5, which follows directly or can be calculated via the least-squares solution using the pseudo-inverse (7) and the estimated rates $\tilde{\boldsymbol{r}}_b$.

Elementary Representation of the Null Space and Global Calculability Analysis

In addition to the questions of calculability and balanceability considered in the previous section, it is worth dealing with the following questions. Which combinations of given rates allow one to calculate a certain unknown rate in the complete system defined by equation (2)? Which rates must we additionally provide to be able to calculate a certain unknown rate, which is not calculable by a given set of known rates? And are there rates that can never be observed, because the set of measurable rates does not allow their calculation?

To answer these questions, a more detailed investigation of the null space of $\mathbf{N}_n$ is helpful. We recall that this subspace contains all possible vectors $\boldsymbol{k}$ with

$$\boldsymbol{0} = \mathbf{N}_n \boldsymbol{k} \tag{13}$$

A (non-unique) representation of the null space can be given by the null space (or kernel) matrix $\mathbf{K}_n$ of $\mathbf{N}_n$, which contains a set of linearly independent vectors fulfilling equation (13). First, the impact of changing a rate from "unknown" to "known" on the null space of matrix $\mathbf{N}_n$ in relation (5) is studied. Are we able to derive the new null space matrix from the old one directly by inspection? For such considerations, the following fact will be helpful:

Property P1: Consider a network with given sets of known rates ($\boldsymbol{r}_b$) and unknown rates ($\boldsymbol{r}_n$). Let a vector $\boldsymbol{k}$ be contained in the null space of $\mathbf{N}_n$. If a rate of $\boldsymbol{r}_n$, say the z-th component, is now considered to be measured and therefore shifted to the set $\boldsymbol{r}_b$, then the vector $\hat{\boldsymbol{k}}$ constructed from $\boldsymbol{k}$ by deleting the z-th component, is contained in the new null space, if this component is zero.

This follows immediately from equation (13), because the new matrix $\hat{\mathbf{N}}_n$ is obtained from the original matrix $\mathbf{N}_n$ by canceling the z-th column. Therefore, all linear combinations of columns in $\mathbf{N}_n$ without the z-th one, which yield the null vector, are valid for $\mathbf{N}_n$ as well. For example, consider the simple network in Figure 2. The stoichiometry matrix $\mathbf{N}$ and a possible representation of the kernel matrix $\mathbf{K}$ of $\mathbf{N}$ is given by the following:

$$\mathbf{N} = \begin{pmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & 1 & -1 \end{pmatrix}, \quad \mathbf{K} = \begin{pmatrix} 1 & 0 \\ 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{14}$$

These matrices represent also the matrices $\mathbf{N}_n$ and $\mathbf{K}_n$ of equation (5), if no rates are known ($\mathbf{N}_b = \boldsymbol{0}$). Inspection of the kernel matrix indicates that in this case no rates are calculable. Let us now consider the case where the rate $\boldsymbol{r}_1$ could be measured. The matrices and vectors have now the following structure (the circumflex indicates the new matrices/vectors):

$$\hat{\mathbf{N}}_n = \begin{pmatrix} -1 & -1 & 0 \\ 1 & 1 & -1 \end{pmatrix}, \qquad \hat{\mathbf{N}}_b = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$$\hat{\boldsymbol{r}}_n = \begin{pmatrix} \boldsymbol{r}_2 \\ \boldsymbol{r}_3 \\ \boldsymbol{r}_4 \end{pmatrix}, \quad \hat{\mathbf{K}}_n = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

The representation of $\hat{\mathbf{K}}_n$ reflects the proposition of Property P1. Because the second column of $\mathbf{K}$ contains a zero at the first position, this vector is contained in the new null space after removal of this component. In our example, this vector represents the only remaining dimension of the null space and is used for the construction of $\hat{\mathbf{K}}_n$.

One could assume that the "surviving" columns from $\mathbf{K}_n$ could be used for the construction of the new matrix, $\hat{\mathbf{K}}_n$. In general, however, this fails: Use the same system as in equation (14) but now consider the rate $\boldsymbol{r}_2$ to be measured instead of $\boldsymbol{r}_1$. In this case, no column of $\mathbf{K}_n$ would "survive," because both columns contain, in the second row, a value unequal to zero. However, one can easily verify that the remaining null space is of dimension one!

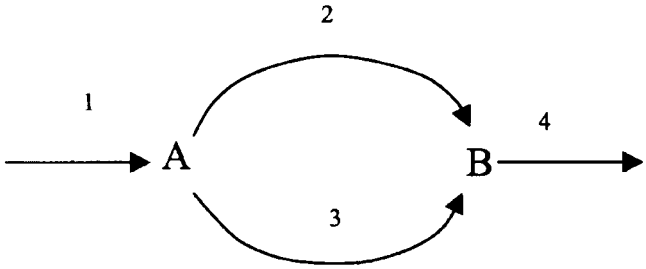

Figure 2. Network example 2.

The representation of the null space as a kernel matrix is not sufficient for our purpose to derive the new null space from the old one. The null space has to be represented as a matrix, $\boldsymbol{\Delta}_n$, in such a way that the column vectors in $\boldsymbol{\Delta}_n$ are maximal with respect to the zeroes involved and span the null space. The idea is that $\boldsymbol{\Delta}_n$ (here called *elementary representation of the null space*) could be used in such a way that if the z-th rate is newly considered to be known, we can derive the (complete) new $\hat{\Delta}_n$ by retaining those column vectors of $\boldsymbol{\Delta}_n$ that have a zero at the z-th position. For this purpose a definition of the matrix $\boldsymbol{\Delta}_n$ and its elements with regard to the problem as formulated in equation (5) is given:

C1. Every column vector $\boldsymbol{v}$ of $\boldsymbol{\Delta}_n$ is a vector contained in the null space: $\boldsymbol{0} = \mathbf{N}_n \boldsymbol{v}$.

C2. $\boldsymbol{v}$ is nondecomposable in the sense that there is no nontrivial vector $\boldsymbol{k}$ in the null space, which contains zero elements wherever $\boldsymbol{v}$ does and has at least one additional zero component.

C3. All possible $\boldsymbol{v}$ fulfilling conditions C1 and C2 (apart from vectors obtained by rescaling) are contained in $\boldsymbol{\Delta}_n$.

With conditions C1–C3 the columns in $\boldsymbol{\Delta}_n$ correspond to all possible sets of linearly dependent columns of $\mathbf{N}_n$ that cannot be further decomposed into smaller sets of linearly dependent vectors. The column vectors $\boldsymbol{v}$ of $\boldsymbol{\Delta}_n$ are elementary flux modes, if all reactions are considered to be reversible (cf. Heinrich and Schuster, 1996; Schuster et al., 2000). (The assumption of reversibility is necessary here so that the *complete* null space is covered). Therefore, the algorithm given by Schuster et al. (2000) can be used for the calculation of $\boldsymbol{\Delta}_n$. However, each pair of modes differing only in the directionality must be counted only as one mode. As the elementary modes are, in the completely reversible case, related to the direct mechanisms in chemistry, the algorithm by Happel and Sellers (1989) for computing such mechanisms can also be used. Note that, again in the reversible case, the number of elementary modes (and therefore the number of columns in $\boldsymbol{\Delta}_n$) is in general greater (but never smaller) than the dimension of the null space (see

Pfeiffer et al., 1999, and the example below). So what can $\mathbf{\Delta}_n$ be used for? First, $\mathbf{\Delta}_n$ has the following useful properties:

a. $\mathbf{\Delta}_n$ is a unique and complete representation of the null space of $\mathbf{N}_n$ (up to scaling).
b. Assume that the z-th element of $\boldsymbol{r}_n$ is newly introduced to have a measured value. The elementary representation, $\hat{\mathbf{\Delta}}_n$, of the new null space can easily be obtained by copying all column vectors of $\mathbf{\Delta}_n$ containing a zero at the z-th position, which is then removed.
c. The calculability condition described in the previous section on the classification of systems and rates is valid if applied to $\mathbf{\Delta}_n$ rather than to $\mathbf{K}_n$: A rate $\boldsymbol{r}_i$ contained in $\boldsymbol{r}_n$ is calculable if (and only if) all vectors in $\mathbf{\Delta}_n$ have a zero at the i-th position.

Property (a) follows from condition C3. A consequence of (a) is that every vector in the null space can (not necessarily uniquely) be represented as a linear combination of columns of $\mathbf{\Delta}_n$. Property (b) follows by this argumentation: Let $\boldsymbol{k}$ be a column vector of $\mathbf{\Delta}_n$, whose z-th element is zero. Because of Property P1, $\boldsymbol{k}$ is also contained in the new null space (after deletion of the z-th element). It must also be contained in the new $\hat{\mathbf{\Delta}}_n$, because otherwise it would be decomposable (cf. Condition C2) and could, hence, not be contained in $\mathbf{\Delta}_n$. This is also valid for the opposite direction: every $\hat{\boldsymbol{k}}$ in $\hat{\mathbf{\Delta}}_n$ must be contained in $\mathbf{\Delta}_n$ (if a zero element is inserted at the z-th position in $\hat{\boldsymbol{k}}$). Property (c) follows from the calculability condition and Property (a).

With these useful properties we are now able to answer the questions posed at the beginning of this section, by the following three tests. (The elementary representation of the null space of the complete stoichiometry matrix $\mathbf{N}$ is now called $\mathbf{\Delta}$.)

Test T1. If we want to calculate a certain unknown rate $\boldsymbol{r}_i$, we can inspect $\mathbf{\Delta}$ to find an appropriate subset, $\boldsymbol{r}_m$, of the set of measurable rates which has to satisfy the following: Every vector in $\mathbf{\Delta}$, which has a non-zero value at the i-th position must contain at least one further non-zero value at a position corresponding to a rate in $\boldsymbol{r}_m$. This can be explained in more detail as follows. Using an appropriate $\boldsymbol{r}_m$, matrix $\mathbf{\Delta}$ can, possibly after rearrangement of rows and columns, be partitioned as

$$\mathbf{\Delta} = \begin{pmatrix} \mathbf{\Delta}_{\mathrm{mx}} & \mathbf{\Delta}_{\mathrm{m1}} & \mathbf{0}_{\mathrm{m0}} \\ \mathbf{\Delta}_{\mathrm{ix}} & \mathbf{0}_{\mathrm{i1}} & \mathbf{0}_{\mathrm{i0}} \\ \mathbf{\Delta}_{\mathrm{px}} & \mathbf{\Delta}_{\mathrm{p1}} & \mathbf{\Delta}_{\mathrm{p0}} \end{pmatrix} \tag{15}$$

where $\mathbf{\Delta}_{\mathrm{ix}}$, $\mathbf{0}_{\mathrm{i1}}$, and $\mathbf{0}_{\mathrm{i0}}$ are row vectors corresponding to the rate $\boldsymbol{r}_{\mathrm{i}}$. They differ in that $\mathbf{\Delta}_{\mathrm{ix}}$ contains only non-zero elements and $\mathbf{0}_{\mathrm{i1}}$ and $\mathbf{0}_{\mathrm{i0}}$ are null vectors. $\mathbf{\Delta}_{\mathrm{mx}}$, $\mathbf{\Delta}_{\mathrm{m1}}$ and $\mathbf{\Delta}_{\mathrm{m0}}$ are submatrices corresponding to the rates in $\boldsymbol{r}_m$ and, accordingly, $\mathbf{\Delta}_{\mathrm{px}}$, $\mathbf{\Delta}_{\mathrm{p1}}$ and $\mathbf{\Delta}_{\mathrm{p0}}$ are submatrices corresponding to all rates considered not to be measured (except $\boldsymbol{r}_i$). The vector $\boldsymbol{r}_m$ must be chosen in such a way that each column of $\mathbf{\Delta}_{\mathrm{mx}}$ contains at least one non-zero element. It is possible that some columns of $\mathbf{\Delta}$ contain a zero at the i-th position and at least one non-zero value at a position of a rate contained in the chosen $\boldsymbol{r}_m$. These columns are arranged in the columns represented by $\mathbf{\Delta}_{\mathrm{m1}}$, $\mathbf{0}_{\mathrm{i1}}$, and $\mathbf{\Delta}_{\mathrm{p1}}$.

Let a set $\boldsymbol{r}_m$ be found so that equation (15) is fulfilled for a certain rate $\boldsymbol{r}_i$. (Note that this choice may not be unique.) Due to Property P1, the matrix $\mathbf{\Delta}_n$ is obtained from $\mathbf{\Delta}$ by deleting all submatrices except $\mathbf{0}_{\mathrm{i0}}$ and $\mathbf{\Delta}_{\mathrm{p0}}$.

$$\mathbf{\Delta}_n = \begin{pmatrix} \mathbf{0}_{\mathrm{i0}} \\ \mathbf{\Delta}_{\mathrm{p0}} \end{pmatrix} \tag{16}$$

As the first row (which originates from the i-th row in $\mathbf{\Delta}$) is a null vector, the calculability condition for $\boldsymbol{r}_i$ is fulfilled. (This is also valid for a scenario as in equation (5), where $\mathbf{\Delta}_n$ rather than $\mathbf{\Delta}$ has to be analyzed, see comment below).

Test T2. If the z-th component of a column vector in $\mathbf{\Delta}$ is zero whenever the i-th component of this vector is non-zero then rate $\boldsymbol{r}_z$ cannot provide a contribution to the calculation of $\boldsymbol{r}_i$, even if the former were a known rate. This follows because measuring the rate $\boldsymbol{r}_z$ would not result in the removal of any column of $\mathbf{\Delta}$ in which the i-th component is non-zero.

Test T3. Assume there is, in $\mathbf{\Delta}$, a vector that has non-zero values only at those positions that correspond to nonmeasurable rates. Then all these rates with a non-zero value will never be calculable with metabolite balancing! This is true because this vector of $\mathbf{\Delta}$ is also contained in $\mathbf{\Delta}_n$ for any choice of known rates from among the set of measurable rates.

Note that the tests T1–T3 can be applied not only to the complete matrix $\mathbf{\Delta}$, which is then reduced to $\mathbf{\Delta}_n$, but also to any matrix $\mathbf{\Delta}_n$ which is then reduced to $\hat{\mathbf{\Delta}}_n$ by considering additional rates to be known. If the reduced matrix is empty (determined system), it should be replaced by the null vector (representing a 0-dimensional kernel) to guarantee applicability of test T1. Furthermore, test T2 implies, in particular, that if the null space matrix and, hence, the matrix $\mathbf{\Delta}$ are block-diagonalizable (Schuster and Schuster, 1991), the rates belonging to one diagonal block do not contribute to the calculation of the rates in any other block.

For illustration we return to our example in Figure 2, which is formulated stoichiometrically in equation (14). If no rates are given, then $\mathbf{\Delta}_n$ is equal to the matrix $\mathbf{\Delta}$ of the complete stoichiometry matrix $\mathbf{N}$, given by:

$$\mathbf{\Delta}_n = \mathbf{\Delta} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \tag{17}$$

The first and second columns describe the elementary flux modes in which the entire flux goes through reac-

tions [1, 2, and 4] and [1, 3, and 4], respectively. The third column describes a cyclic flux, if the reactions are reversible. Note that any two vectors of $\boldsymbol{\Delta}$ could be used for representation of a kernel matrix [see $\mathbf{K}$ in equation (14)]. For the calculation of rate $\boldsymbol{r}_4$ we need only rate $\boldsymbol{r}_1$, because those vectors of $\boldsymbol{\Delta}$ that contain a non-zero element at the fourth position, contain also a non-zero element at the first position. On the other hand, if $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$ could be measured, the new $\hat{\boldsymbol{\Delta}}_n$ would be empty (because every vector in $\boldsymbol{\Delta}$ contains a non-zero value at the position of rate $\boldsymbol{r}_2$ and/or $\boldsymbol{r}_3$) and therefore the remaining rates ($\boldsymbol{r}_1$ and $\boldsymbol{r}_4$) would be calculable. If rate $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$ would not be measurable, then they could never be calculated, because the third vector in $\boldsymbol{\Delta}$ contains non-zero values only for these non-measurable rates.

Now, reaction rate $\boldsymbol{r}_1$ is measured. The first two vectors in $\boldsymbol{\Delta}$ must be removed, because they contain a non-zero value at the first position corresponding to $\boldsymbol{r}_1$. The third vector "survives" by deleting the zero-component at the first position:

$$\hat{\mathbf{N}}_n = \begin{pmatrix} -1 & -1 & 0 \\ 1 & 1 & -1 \end{pmatrix}, \qquad \hat{\mathbf{N}}_b = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$$\hat{\boldsymbol{r}}_n = \begin{pmatrix} \boldsymbol{r}_2 \\ \boldsymbol{r}_3 \\ \boldsymbol{r}_4 \end{pmatrix}, \qquad \hat{\boldsymbol{\Delta}}_n = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

Inspection of $\hat{\boldsymbol{\Delta}}_n$ shows that $\boldsymbol{r}_4$ is calculable as expected, because the only remaining vector in $\hat{\boldsymbol{\Delta}}_n$ has a zero value at the position of $\boldsymbol{r}_4$.

Now, we again consider the case, where system (14) is given with a measured rate $\boldsymbol{r}_2$ instead of $\boldsymbol{r}_1$. The new elementary representation, $\hat{\boldsymbol{\Delta}}_n$, of the null space reads:

$$\hat{\mathbf{N}}_n = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}, \qquad \hat{\mathbf{N}}_b = \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

$$\hat{\boldsymbol{r}}_n = \begin{pmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \boldsymbol{r}_4 \end{pmatrix}, \quad \hat{\boldsymbol{\Delta}}_n = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Because $\hat{\boldsymbol{\Delta}}_n$ contains no null rows, no rate is calculable. However, by inspection of $\hat{\boldsymbol{\Delta}}_n$, we see that an arbitrary additionally measured rate would be sufficient to calculate the others because the new $\hat{\boldsymbol{\Delta}}_n$ would then be empty. Note that, because $\hat{\boldsymbol{\Delta}}_n$ comprises only one vector, it is a possible kernel matrix (while this is not valid for the matrix $\boldsymbol{\Delta}$ calculated above).

FluxAnalyzer

The software tool, *FluxAnalyzer* was developed for a convenient performance of metabolic flux analysis using the described methods and for a visualization of the calculated flux distributions. This package was developed with the programming language of MATLAB, in which several required routines of linear algebra are available. The *FluxAnalyzer* represents an interactive graphical interface for metabolic networks which can easily be defined by the user. The most important features are:

- Arbitrary metabolic networks can be symbolically composed. For this purpose, masks are provided where each network element and its properties can be defined. Each element belongs to one of the types: *metabolite*, *reaction*, or *biomass constituent*.
- Graphical network representations, called flux maps and designed by the user, can be linked to the symbolic network representation by user interfaces (text boxes) predefined in MATLAB. Each text box refers to one network element and can be arranged at a proper place in the flux map and allows the input of values (e.g., reaction rates or biomass composition) or the output of calculated fluxes (cf. Fig. 3: text boxes refer to reaction rates).
- Actions can be easily performed via menu items (the user does not need to be completely aware of the mathematical details of the used algorithms).
- Classification of rates for an entered set of known rates (distinction by color).
- Routines for balancing/calculating all balanceable/calculable reaction rates.
- Sensitivity analysis of calculated rates with respect to measured ones.
- Routine for optimization by defining an objective function and constraints.
- Topology analysis including calculation and display of elementary flux modes (on the basis of a defined set of irreversible reactions) and of the elementary representations of the null space $\boldsymbol{\Delta}$ or $\boldsymbol{\Delta}_n$.
- Documentation: saving of data sets or flux maps.
- Several network projects of one or more organisms can be managed.

The catabolic flux map for purple nonsulfur bacteria used for investigations described in the following section is presented in Figure 3. The routines and algorithms can be started under the menu item "MFA."

STOICHIOMETRIC ANALYSIS OF PHOTOHETEROTROPHIC GROWTH OF PURPLE NONSULFUR BACTERIA

An illustrative example of a stoichiometric analysis in underdetermined networks can be given by the study of the metabolism in purple nonsulfur bacteria (PNB) growing under photoheterotrophic conditions. These bacteria (e.g., *Rhodobacter capsulatus*, *Rhodobacter sphaeroides*, *Rhodospirillum rubrum*) are representatives

[1] Further information on the *FluxAnalyzer* can be obtained on the website http://www.mpi-magdeburg.mpg.de/research/project_a/pro_a5a/mfaeng/fluxanaly.html A free version for academic users is available from the corresponding author.

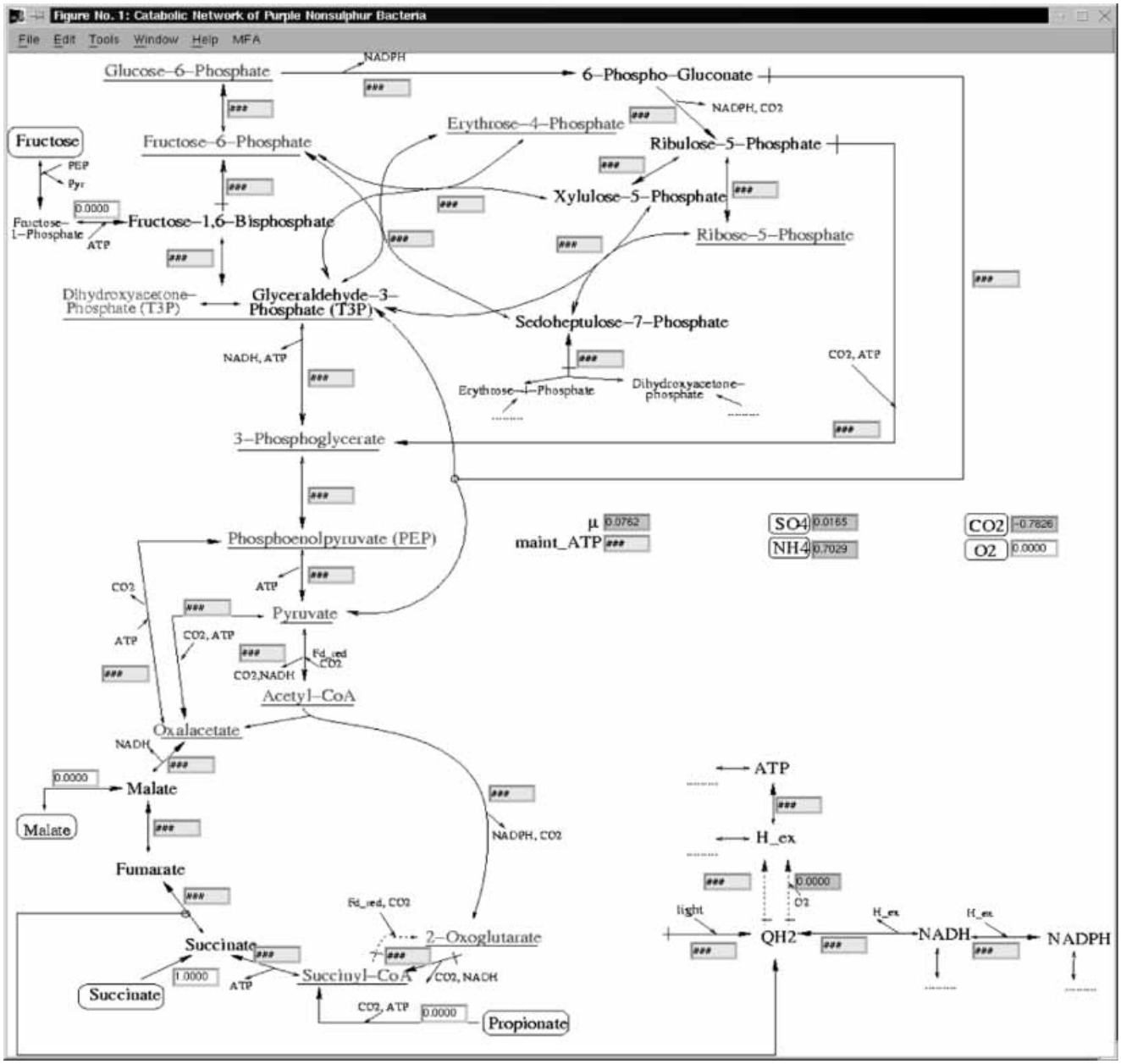

Figure 3. Catabolic flux map of PNB used by the *FluxAnalyzer* (screen shot). The calculated rates of scenario S1 are shown. Unknown rates (that is, rates that are neither given nor calculable) are marked by ###. Bright background: entered (known) rates; dark background: calculated values. The signs of the reaction rates are defined according to the directions with the large arrow heads. Double-headed arrows indicate reversible reactions. Only the considered true branchpoint metabolites are displayed. Underlined metabolites are precursors, framed symbols are external substances.

of anoxygenic photosynthetic bacteria for which recently some biotechnological applications have been published (Cornet and Albiol, 2000; Cornet et al., 1998; Hai et al., 2000; Matsunaga et al., 2000).

Here, a first attempt to analyze the photoheterotrophic growth of PNB stoichiometrically is presented. We investigate a proposed stoichiometric network of the central carbon and energy metabolism without any measurements, and only with some well-known assumptions about the stationary metabolism. The algorithms described in the previous section are useful for stoichiometric analysis of such underdetermined linear systems. It is shown that important metabolic constraints may be detected. To our knowledge, no attempt has been made so far to analyze the central metabolic network in anoxygenic photosynthetic bacteria with MFA. Firstly, a short overview of the metabolism in PNB is given.

Purple Nonsulfur Bacteria (*Rhodospirillaceae*)

Purple nonsulfur bacteria (PNB) are anoxygenic, facultatively photosynthetic bacteria and are considered to be the most versatile organisms with respect to metabolism (Tabita, 1995). Aerobically in the dark, the cells grow chemoheterotrophically by respiration, while anaerobically in the light, a switch to photoheterotrophic or photoautotrophic growth, respectively, takes place. Purple nonsulfur bacteria grow as chemotrophs when exposed to air, regardless of whether light is available or not. The components of the photosynthetic apparatus are synthesized and arranged in the membrane only under anaerobic conditions. Here, only the photoheterotrophic growth is considered, which is reviewed in detail in McEwan (1994) and Tabita (1995).

Photoheterotrophic growth

The PNB have only one photosystem transforming light energy into a cyclic electron transport (in contrast to noncyclic electron transport in plants or cyanobacteria; McEwan, 1994). The result of these membrane processes is a proton-motive force gradient which can be used for ATP formation by ATP synthase. In the net balance no reducing equivalents are produced during this cycle. Furthermore, in contrast to cyanobacteria and plants, PNB are unable to use water as an electron donor and therefore no oxygen is produced in photosynthesis ("anoxygenic"). Instead, they require other electron donors for production of the necessary reduction equivalents:

- Photoautotrophic growth: Molecular hydrogen can be used lithotrophically as electron donor; CO_2 is then used autotrophically as sole carbon source.
- Photoheterotrophic growth: Electrons are supplied by organic substrates (such as succinate); the substrates are then also the major carbon source.

Thus, NAD(P)H is either generated directly in NAD(P)-depending reactions (e.g., malate dehydrogenase) or by reverse transport of electrons (e.g., coming from hydrogen or succinate dehydrogenase), which requires energy produced by photosynthesis. The central carbon metabolism during photoheterotrophic growth is composed of some well-known pathways and some specialized routes (cf. Fig. 3):

- Embden-Meyerhoff pathway (or the oppositely directed gluconeogenesis) (Conrad and Schlegel, 1977)
- Entner-Doudoroff pathway (Conrad and Schlegel, 1977)
- Pentose phosphate pathway (oxidative and nonoxidative branches)
- TCA (tricarboxylic acid cycle): It has been shown that this pathway (or parts of it) might also be driven in the reductive direction under anaerobic conditions. Two necessary ferredoxin-dependent enzymes catalysing α-ketoglutarate synthesis and pyruvate synthesis are inducible (Buchanan et al., 1967), while the α-ketoglutarate-dehydrogenase is repressed during anaerobic growth.
- Calvin cycle: This pathway (with the two key enzymes ribulose-1,5-bisphosphate carboxylase and phosphoribulokinase) is especially involved under photolithoautotrophic conditions, where CO_2 must be fixed to provide organic carbon; under photoheterotrophic conditions this cycle seems to play an important role for balancing of reduction equivalents because it functions as a redox sink (McEwan, 1994)
- Transport systems for uptake of organic substrates (e.g., fructose-PTS; Saier et al. 1971)
- The most important anaplerotic reactions are catalyzed by PEP carboxykinase, PEP carboxylase, and pyruvate carboxylase.

In contrast to respiration in the dark, the photosynthetic cyclic electron transport under anaerobic conditions in the light does not consume reducing power such as NADH. Under aerobic conditions, oxidases are used for respiratory processes, where oxygen is the electron acceptor. This also allows an aerobic growth of these bacteria in the dark. Further special pathways induced only under certain conditions are not considered here, e.g., the fixation of nitrogen via a nitrogenase.

Proposed Stoichiometric Network

The proposed stoichiometric network for the central carbon metabolism of PNB involves the 30 most important catabolic branchpoint-metabolites (Table II) for photoheterotrophic growth on several substrates (malate, fructose, succinate, propionate). Branchpoint means that the balanced metabolites participate in at least three reactions. Catabolites participating in only two reactions are not considered because these two reactions can be combined into one. Forty-one catabolic reactions are contained in the network (several of them being combined ones and representing *reaction sequences*, Table III). Oxygen and an overall reaction of the oxidases are also incorporated to compare photoheterotrophic and respiratory growth (scenario 5).

The anabolic network consists of 46 building blocks whose synthesis is described in 46 anabolic reaction sequences using the stoichiometries given in Neidhardt et al. (1990). The fluxes through these pathways depend on the biomass composition and growth rate. As complete data on the biomass composition of purple nonsulfur bacteria are not available, we assume that, under photosynthetic conditions, these data can be approximated by those of *Escherichia coli*, except that a higher lipid percentage (photosynthetic membranes) is assumed and that poly-hydroxybutyrate (PHB) is the major storage component in PNB (Table I). For the estimated content of lipids (16% of dry weight) we used the fact that steady-state phototrophic cells have a 50–70%

Table I. Estimated biomass composition of photosynthetically grown purple nonsulfur bacteria. Except for the lipids (which were increased to take into account the photosynthetic membranes) and poly-hydroxybutyrate and bacteriochlorophyll (which were included in addition) data from *E. coli* (Neidhardt et al., 1990) were taken and reduced proportionately. The values were chosen to be in agreement with data reported by Kaplan (1978) for the lipid content, by Göbel (1978) for bacteriochlorophyll and by Fuller (1995) for PHB.

Biomass components	Estimated biomass proportion
Protein	0.51 g/g dry weight
Lipids	0.16 g/g dry weight
RNA	0.16 g/g dry weight
DNA	0.03 g/g dry weight
Poly-hydroxybutyrate (PHB)	0.05 g/g dry weight
Bacteriochlorophyll	0.01 g/g dry weight
Lipopolysaccharide	0.03 g/g dry weight
Peptidoglycan	0.03 g/g dry weight
Polyphosphate	0.02 g/g dry weight

higher phospholipid content relative to chemotrophically grown cells (Kaplan, 1978). For an approximation of the PHB content (5%) we took values given by Fuller (1995). Besides, the considerable synthesis of bacteriochlorophyll is taken into account and estimated (1%) by using values reported in (Göbel, 1978). Small deviations from the estimated biomass composition are not expected to have a great impact and are of minor importance for qualitative conclusions. The resulting fluxes from the catabolic network into the anaerobic network were calculated by the *FluxAnalyzer*, [Table II, in (mmol of metabolite/g DW)].

A network scenario was created by *FluxAnalyzer* allowing the investigations presented in the following. The catabolic flux map of this network scenario (comprising the catabolism of PNB as proposed) is depicted in Figure 3.

Table II. Catabolic (branchpoint) metabolites considered in the catabolic network of purple nonsulfur bacteria.

Catabolite (intracellular)	Symbol	Fluxes into biomass [mmol/(g DW)]	Notes
Oxygen	O_2	0	Only available in aerobic medium
Carbon dioxide	CO_2	−1.3665 (negative value: produced during biomass synthesis)	Considered as one pool for bicarbonate and molecular CO_2
Sulfate	SO_4	0.2165	For synthesis of some amino acids
Ammonia	NH_4	9.2244	
Glucose-6-phosphate	G6P	0.0449	
Fructose-6-phosphate	F6P	0.0801	
Fructose-l,6-bisphosphate	F16P	0	
Triose-phosphate	T3P	0.2225	One pool involving glyceraldehyde-3-phosphate and dihydroxyacetone phosphate
3-Phosphoglycerate	PG	1.6707	
Phosphoenolpyruvate	PEP	0.7233	
Pyruvate	Pyr	2.5098	
Acetyl-CoA	AcCoA	5.7626	
Alpha-Ketoglutarate	AlK	0.9520	
Succinyl-CoA	SuccCoA	0.5624	
Succinate	Succ	−0.4708	
Fumarate	Fum	−0.9318	
Malate	Mal	0	
Oxaloacetate	OxA	2.3706	
6-Phospho-gluconate	PGluc	0	
Ribulose-5-phosphate	R15P	0	
Ribose-5-phosphate	R5P	0.7427	
Xylulose-5-phosphate	X5P	0	
Sedoheptulose-7-phosphate	S7P	0	
Erythrose-4-phosphate	E4P	0.3347	
ATP	ATP	34.7753	
NADH	NADH	−2.2166	
NADPH	NADPH	19.0476	
Reduced Ubiquinone	QH2	0	
Electrons in electron transport phosphorylation chains	ElecPhos	0	Represents the quantity of electrons in respiratory and photosynthetic electron transport phosphorylation chains (also considered as constant in steady state)
Periplasmic protons	Hp	0	Protons in periplasm pumped outside by respiration or cyclic photosynthesis (representing proton motive force)

Table III. Reaction sequences considered in the catabolic network of purple nonsulfur bacteria (cf. Fig. 3 and Tabita, 1995).

Name of reaction (sequence)	(Cumulative) equation	Notes/participating enzymes
μ	cf. Table II	Biomass synthesis
O_2_up	= O_2	Uptake of oxygen
N_up	=N	Uptake of ammonia
CO_2_up	= CO_2	Uptake/release of carbon dioxide/bicarbonate
Succ_up	=Succ	Uptake of succinate
F_up	ATP + PEP = F16P + Pyr	Uptake of fructose by fructose-PTS
Mal_up	= Mal	Uptake of malate
S_up	=S	Uptake of sulfate
Prop_up	CO_2 + ATP = SuccCoA	Uptake of propionate; reaction cumulates propinate uptake and conversion to SuccCoA via methylmalonyl-CoA (propionyl-CoA carboxylase, methylmalonyl-CoA racemase, methylmalonyl-CoA mutase)
F6P::G6P	F6P = G6P	Glucose-phosphate-isomerase
F16P::F6P	F16P = F6P	Fructose-1,6-bisphosphatase (irreversible, induced with Calvin cycle) or/and pyrophosphate-dependent 6-phosphofructo-1-kinase (reversible)
F16P::T3P	F16P = 2 T3P	Fructose-bisphosphate-aldolase
T3P::PG	T3P = PG + NADH + ATP	Glyceraldehyde dehydrogenase (the enzyme induced during the Calvin cycle also uses NADH for the reverse reaction) and 3-phosphoglycerate kinase
PG::PEP	PG = PEP	Phosphoglycerate mutase and enolase
PEP::Pyr	PEP = Pyr + ATP	Pyruvate kinase; the opposite direction is possible through PEP synthase
Pyr::AcCoA	Pyr = AcCoA + NADH + CO_2	Pyruvate dehydrogenase (irreversible); the reverse reaction could be performed due to the inducible ferredoxin dependent pyruvate synthase; the reduced ferredoxin is here considered as produced from NADH via a NADH:ferredoxin-oxidoreductase
AcCoA::AlK	AcCoA + OxA = AlK + CO_2 + NADPH	Citrate-synthase, aconitase, isocitratc dehydrogenase (NADP); considered as reversible under phototrophic conditions (reductive TCA)
AlK::SuccCoA	AlK = SuccCoA + NADH + CO_2	Alpha ketoglutarate dehydrogenase (irreversible and only induced under aerobic conditions); the opposite direction (reductive TCA) is possible due to alpha-ketoglutarate-synthase (uses reduced ferredoxin – here considered as produced from NADH)
SuccCoA::Succ	SuccCoA = Succ + ATP	Succinyl-CoA hydrolase
Succ::Fum	Succ = Fum + QH2 + 2 ElecPhos	Succinate dehydrogenase: releases electrons to redox carrier in the membrane; the opposite direction can be considered as an inducible fumarate reductase (feeding of electrons in the membrane is then done via proton-pumping NADH-dehydrogenase)
Fum::Mal	Fum = Mal	Fumarase
Mal::OxA	Mal = OxA + NADH	Malate dehydrogenase
G6P::PGluc	G6P = PGluc + NADPH	Glucose-6-phosphate dehydrogenase (irreversible), 6-phosphogluconolactonase (irreversible)
EntDoud	PGluc = T3P + Pyr	Entner-Doudoroff pathway
PGluc::R15P	PGluc = R15P + NADPH + CO_2	Phosphogluconate dehydrogenase (irreversible)
R15P::X5P	R15P = X5P	Ribulose-5-phosphate epimerase
R15P::R5P	R15P = R5P	Ribose-5-phosphate isomerase
X5P|R5P::S7P|T3P	X5P + R5P = S7P + T3P	Transketolase
S7P|T3P::F6P|E4P	S7P + T3P = F6P + E4P	Transaldolase
X5P|E4P::F6P|T3P	X5P + E4P = F6P + T3P	Transketolase
R15P::PG	R15P + ATP + CO_2 = 2 PG	The two reactions of the Calvin-Bassham-cycle (only anaerobically inducible); phosphoribulokinase, ribulose-1,5-bisphosphate-carboxylase (RubisCO); irreversible
E4P|T3P::S7P	E4P + T3P = S7P	Sedoheptulose-1,7-bisphosphate aldolase and fructose-1,6-/sedoheptulose-1,7-bisphosphatase (irreversible, part of the Calvin cycle)
OxA::PEP	OxA + ATP = PEP + CO_2	PEP-carboxykinase (considered as reversible; opposite direction also possible by PEP-carboxylase)
Pyr::OxA	Pyr + CO_2 + ATP = OxA	Pyruvate carboxylase
Photo	= QH2	Photosynthesis: light induced reduction of ubiquinone
CyclicET	QH2 = 3 Hp	Cyclic electron transport during photosynthesis with generation of proton motive force; no electrons are removed during this cycle (expect reverse electron transport via NADH dehydrogenase)
NADHDehydro	NADH = QH2 + 2 ElecPhos + 2 Hp	NADH dehydrogenase (used for respiration and reverse electron transport)

Table III. Continued

Name of reaction (sequence)	(Cumulative) equation	Notes/participating enzymes
Oxidase	QH2 + 0.5 O_2 + 2 ElecPhos = 4 Hp	Respiration with removal of electrons; (used P/O-ratio of 2—but all scenarios performed here are independent of this value)
Transhydro	NADH + Hp = NADPH	Transhydrogenase, uses proton motive force, the NADPH-producing direction is likely
ATPSynth	3 Hp = ATP	ATP-synthase; uses the proton motive force provided by respiration or photosynthesis
MaintATP	ATP =	Maintenance demand: all energy equivalents not contained in the energy requirement for biomass synthesis (needed for non-growth associated energy consumption, such as transport, futile cycles, maintaining membrane potential); difficult to estimate

Note: In most cases the symbol of the reaction sequence is constructed as "starting substance::product." In a few cases an abbreviation of the key enzyme or key reaction is used. Substrate uptake/excretion is abbreviated as 'Substrate_up'. Only the considered metabolites (Table II) are contained in the cumulative equations.

Photoheterotropic Growth on Several Substrates

In this section, the photoheterotrophic growth of PNB on several substrates is studied using the proposed reaction system. We attempt to derive important metabolic constraints for this versatile network without measurements or special assumptions. The results of the studied scenarios are summarized in Table IV, where the given, calculated, and noncalculable rates can be distinguished.

First, the phototrophic growth under anaerobic conditions on succinate is analyzed. In scenario 0 (S0, Table IV) we use only the fact that the oxygen uptake is clearly zero. Besides, we assume that no intermediates are excreted.

Using this strongly underdetermined system (degrees of freedom: 7) we can hardly determine any further reaction rates, except that the flux through the oxidase must be zero. This indeterminacy is not surprising because we have no information about the growth rate nor even about succinate uptake.

In scenario S1 (shown in Table IV and in Fig. 3) we want to investigate whether it is possible to determine any fluxes for a given uptake of substrate. We assume at first that 1 $mmol \cdot g^{-1}DW \cdot h^{-1}$of succinate is taken up (Succ_up = 1; for the abbreviations of reaction sequences, see Table III). The determination of calculable rates in this still strongly undetermined system (degrees of freedom: 6) brings about the surprising result that the net CO_2 fixation/evolution rate as well as the growth rate are determined, although there are many possible pathways where CO_2 is evolved or fixed. (Remember: The cell is free in this scenario to use the Calvin cycle, reductive/oxidative TCA, pentose phosphate pathway, Entner-Doudoroff pathway, anaplerotic sequences, glycolysis/gluconeogenesis, transhydrogenase, and an arbitrary maintenance demand). That this fixed constraint for the steady-state physiology of the cell can be obtained with such a minimum of information and a large degree of freedom is an inherent property of the network and is closely connected with the carbon and redox balance. The calculated rates of μ and CO_2_up are the only ones for which all intermediates, precursors, and the reduction equivalents NAD(P)H may be properly balanced simultaneously. (Note: If we fictively changed the stoichiometric coefficient of NAD(P)H in any of the catabolic reactions the calculability of μ and CO_2_up would be lost.)

The rates of nitrogen and sulfur uptake are directly coupled to the growth rate and therefore also calculable.

The calculated fluxes in S1 correspond to a normalized uptake of succinate (see above). The calculable rates for another value of succinate influx, J_{Succ_up}, could be easily calculated by multiplication with the normalized rate, i.e., for instance (cf. Table IV):

$$J_{CO_2\_up} = -0.783 \cdot J_{Succ\_up} \tag{18}$$

The actual influx of substrate depends mainly on the rate of photosynthesis, because the necessary energy equivalents for biomass composition and maintenance processes must be produced there. Besides, a limiting reaction rate in the metabolic network could also be the limiting factor for growth and substrate uptake.

Because there occurs a net release of CO_2, succinate is less reduced than biomass [taken as the average redox status of the carbons involved, cf. Ormerod and Sirevag (1983)]. Some further possible substrates were investigated with the same metabolic network as in S1 (S2: malate, S3: propionate, S4: fructose). The above results are also valid for the other substrates and, therefore, the same rates are calculable. As expected, the more reduced a substrate is, the less CO_2 is released. Growth on propionate is only possible if CO_2 or bicarbonate is provided, as experimentally verified in previous studies (Ormerod, 1956; Richardson et al., 1988). A comparison between the calculated CO_2 net fixation (per mmol substrate used) and measured rates found in the literature (Muller, 1933; Ormerod, 1956) yields a very good agreement (Table V) and confirms the detected metabolic constraint as well as the proposed network structure.

Table IV. Assumed (bold, italic) and calculated rates (in [mmol/(g DW *h)]) in the scenarios S0–S6. Empty fields represent noncalculable rates.

Reaction	S0	S1	S2	S3	S4	S5	S6
μ		0.076	0.065	0.076	0.131	0.076	
O_2_up	***0***	***0***	***0***	***0***	***0***	***0***	
N_up		0.703	0.603	0.703	1.205	0.703	
CO_2_up		−0.783	−1.242	0.217	−0.485	−0.783	
Succ_up		***0***	***0***	***0***	***0***	***1***	***1***
Mal_up	***0***	***0***	***1***	***0***	***0***	***0***	***0***
S_up		0.017	0.014	0.017	0.028	0.017	
F_up	***0***	***0***	***0***	***0***	***1***	***0***	***0***
Prop_up	***0***	***0***	***0***	***1***	***0***	***0***	***0***
F6P::G6P						0.690	
F16P::F6P							
F16P::T3P							
T3P::PG						−2.746	
PG::PEP						−0.039	
PEP::Pyr						0.720	
Pyr::AcCoA						0.530	
AcCoA::AlK						0.073	
AlK::SuccCoA						***0***	
SuccCoA::Succ						−0.039	
Succ::Fum						0.997	
Fum::Mal						1.065	
Mal::OxA						1.065	
EntDoud						***0***	
G6P::PGluc						0.686	
PGluc::R15P						0.686	
R15P::X5P						−0.530	
R15P::R5P						−0.198	
X5P[R5P::S7P|T3P						−0.252	
S7P|T3P::F6P|E4P							
X5P|E4P::F6P|T3P							
RI5P::PG						1.415	***0***
E4P|T3P::S7P							***0***
OxA::PEP						0.814	
Pyr::OxA						***0***	
Photo							***0***
CyclicET							0
NADHDehydro						−0.997	
Oxidase	0	0	0	0	0	0	
Transhydro						***0***	
ATPSynth							
MaintATP							

The fact that the growth rate is calculable enables one to determine the maximal biomass yield by dividing the growth rate by the substrate uptake (leading to the unit of gram biomass dry weight per mmol (or gram) substrate used, Table VI). This maximal biomass yield can only be produced by the cell if no excretion of metabolites takes place (as assumed in the scenarios). If propionate is used as a substrate, the cells are able to produce more biomass than the mass of the consumed substrate, due to the net fixation of CO_2. Again, these calculated biomass yields (Table VI) are independent of the maintenance demand and the many degrees of freedom in the central carbon metabolism in scenarios S1–S4.

Table V. Calculated and measured net carbon dioxide fixation (positive) or production (negative) in photosynthetically grown purple nonsulfur bacteria (in mmol per mmol substrates used).

	Succinate (S1)	Malate (S2)	Propionate (S3)	Fructose (S4)
Calculated (MFA)	−0.783	−1.242	0.217	−0.485
Source: Muller, 1933	−0.70	−1.22	—	—
Source: Ormerod, 1956	−0.74	−1.14	0.23	—

It is a special property of the considered network that catabolic reaction rates are calculable during photosynthetic growth, because energy production is decoupled from NADH (not so during respiration, where the calculability of many rates is lost if no measurements are provided; see scenario S6). The necessary balancing of NAD(P)H in the central carbon metabolism during anaerobic photosynthetic growth is a stoichiometric constraint which implies the calculability of some rates.

An example of a more determined flux distribution during anaerobic photosynthetic growth on succinate is

Table VI. Calculated (maximal) biomass yield per used substrate in photosynthetically grown purple nonsulfur bacteria.

	Succinate (S1)	Malate (S2)	Propionate (S3)	Fructose (S4)
g DW/ mmol substrate used (μ/*substrate_up*)	0.076	0.065	0.076	0.131
g DW/ g substrate used (μ/*substrate_up*)	0.644	0.485	1.026	0.728

given in scenario S5, where, in comparison to S1, the following rates are additionally set to zero to further reduce the solution space (these assumptions seem to be reasonable by the given references): the flux through the Entner-Doudoroff pathway (cf. Conrad and Schlegel, 1977), the net flux from AlK to SucCoA (cf. Joshi and Tabita, 2000), the net flux of pyruvate carboxylase (cf. Joshi and Tabita, 2000) and the net flux through the transhydrogenase. This is a hypothetical scenario to show the power of our method in the case where some additional rates are known. As shown in Table IV, considerably more rates than in S1 are calculable (the dimension of the null space decreased from six to two). Of course, the rates calculable in the more general scenario S1 have still the same value. Interestingly, a relatively high flux is necessary through the Calvin cycle (*R15P*::*PG*) catalyzed by RubisCO and phosphoribulokinase, although there is a net release of CO_2! This has been confirmed by the investigation of RubisCO deletion mutants (cf. Wang et al., 1993). The phenomenon can now quantitatively be explained by metabolite balancing: the Calvin cycle (or another adequate pathway serving as redop sink) is absolutely necessary for redox balancing, because reducing equivalents not needed for biomass synthesis are used for refixation of some of the carbon lost in other reactions (*OxA*::*PEP*, *G6P*::*PGluc*):

Analysis of the Elementary Representation of the Null Space

Applying the algorithm for computing the elementary modes to scenarios S1 and S5 (considering all reactions as reversible as is necessary for the calculation of $\mathbf{\Delta}_n$, as shown in a previous section, we obtain 452 and 2 vectors in $\mathbf{\Delta}_n$, respectively. The considerable former number is caused by the high degree of indeterminacy of the systems. One of those vectors occurring in $\mathbf{\Delta}_n$ of S1 but not in the $\mathbf{\Delta}_n$ of S5 corresponds to the following reactions (only the reactions of the vector with a non-zero value are shown; these values are all equal to unity here):

$$(Pyr::OxA, OxA::PEP, PEP::Pyr, Photo, CyclicET, ATPSynth) \quad (19)$$

This is a well-known cycle caused by anaplerotic sequences. It fulfills the steady-state condition. One can assume that no rate in equation (19) is measurable. Therefore, all six rates will never be calculable with metabolite balancing (even if all other rates would be known!) unless one of them can be measured, for example, by isotopic tracer experiments. (19) was not a member of $\mathbf{\Delta}_n$ in scenario S5 anymore, because *Pyr*::*OxA* was set to zero.

The following vector is a member of $\mathbf{\Delta}_n$ in S1 as well as in S5:

$$(-F16P::F6P, F16P::T3P, T3P|S7P::F6P|E4P, E4P|T3P::S7P), \quad (20)$$

(where the minus sign means that the reaction proceeds hypothetically in the backward direction). Because no rate in (20) is measurable, none of these rates will be calculable under phototrophic conditions [cf. S5, where no rate of (20) is calculable]. On the other hand, the reaction sequence *E4P*|*T3P*::*S7P* requiring a fructose 1,6-/sedoheptulose-1,7-bisphosphatase is only induced with the Calvin cycle which is repressed under aerobic conditions. Therefore, under aerobic conditions *E4P*|*T3P*::*S7P* can be set to zero and (20) is not a member of $\mathbf{\Delta}_n$ anymore (and the other rates are possibly calculable provided that an appropriate set of known rates can be given).

The second and last member of $\mathbf{\Delta}_n$ in S5 (also contained in $\mathbf{\Delta}_n$ of S1) reads:

$$(Photo, CyclicET, ATPSynth, MaintATP) \quad (21)$$

Perhaps, it is possible to estimate the rate of photosynthesis or the maintenance demand of ATP. Otherwise, no rate of (21) can be determined under phototrophic conditions. But again, under aerobic conditions in the dark, the rate of photosynthesis is known to be zero and thus, (21) disappears from $\mathbf{\Delta}_n$ giving the chance to calculate the rate of *MaintATP* or *ATPSynth*.

Chemotrophic Growth Under Aerobic Conditions (Scenario S6)

Purple nonsulfur bacteria grow well on several substrates under aerobic conditions in the dark. In scenario S6, we adapt the network to aerobic growth on succinate. In contrast to S1, the rates of *Photo*, *R15P*::*PG* and *E4P*|*T3P*::*S7P* (not induced under aerobic conditions) are set to zero, while oxygen uptake is possible: the rate of *O*$_2$*_up* is set to be unknown. Although seven rates are known (in S1 only five!) no rates are calculable except that of *CyclicET*, the calculation of which is straightforward because the rate of photosynthesis (*Photo*) is zero. The reason was explained above: NADH is now coupled via *Oxidase* to the energy-producing respiratory electron transport phosphorylation, the flux of which is dependent on the unknown maintenance demand of

ATP (only estimations have been proposed; Stephanopoulos et al., 1998). Therefore, further investigations on aerobic growth require measurements of oxygen and substrate uptake, CO_2 and product excretion to obtain a determined and possibly redundant network, where also a consistency check would be possible. Measurements of these rates would then also allow a calculation of the maintenance demand of ATP (*MaintATP*).

DISCUSSION, CONCLUSIONS, AND OUTLOOK

Underdetermined networks are a general problem for metabolic flux analysis, which has become a very helpful and suitable tool for studying and engineering the metabolism of micro-organisms. In this article, we have shown that stoichiometric analysis of underdetermined networks may in some cases lead to new insight and conclusions or at least help to improve the understanding of metabolic constraints. The investigation of the metabolic network of purple nonsulfur bacteria has clearly illustrated that even underdetermined systems can contain very useful information about the metabolic capabilities of an organism. For a highly simplified metabolic network of the aerobic bacterium *Acinetobacter calcoaceticus*, this was shown earlier by Noorman et al. (1996), based on the method of Van der Heijden et al. (1994a).

In the first part of this contribution, the methodical basis of metabolite balancing is reviewed and illustrated to an extent necessary for further presentation herewith. A new approach for the detection of all calculable rates in a network was then introduced. It is simpler than the method proposed by Van der Heijden et al. (1994a) because it does not require singular value decomposition. The present approach is based on the analysis of the null space matrix to that part of the stoichiometry matrix that corresponds to the unknown rates. While null space matrices, which can easily be computed by the Gaussian elimination method, have been extensively used in metabolic control analysis (Heinrich and Schuster, 1996; Reder, 1988) and structural analysis of metabolic networks (Pfeiffer et al., 1999; Schuster and Hilgetag, 1994; Stephanopoulos et al., 1998), its importance for the MFA has, in our opinion, been underestimated so far. Anaerobic networks especially seem to contain metabolic constraints that could be derived even in underdetermined systems using calculability analysis. While in several earlier studies on underdetermined networks, a lumping of sets of reaction rates (e.g., in internal cycles) is used (Vallino and Stephanopoulos, 1993), our approach does not require this. On the other hand, lumping can allow one to compute certain overall fluxes.

We also introduced a more general approach for calculability analysis based on an elementary representation of the null space. This is useful to answer questions such as "Which rates can contribute to the calculation of a certain rate?", "Which combinations of measurements or known rates allow one to calculate a certain rate?" or "Are some of the rates never calculable?"

The *FluxAnalyzer*, an easy-to-use and interactive program written in the Matlab environment, allows a flexible symbolic formulation of complex metabolic networks. Thus, an efficient analysis of stoichiometric networks with the complete set of mathematical tools for doing metabolite balancing can be performed in a graphical interface.

The investigation of the metabolic network in purple nonsulfur bacteria under photoheterotrophic conditions, performed here without any measurements of extracellular fluxes, reveals interesting metabolic constraints with respect to balancing of reduction equivalents (NAD(P)H). During photoheterotrophic growth, the overall growth rate and CO_2-net fixation/release are determined as a function of the substrate used, independently of the real fluxes through the TCA, Calvin cycle, oxidative pentose-phosphate pathway, anaplerotic sequences, and transhydrogenase as well as of the ATP maintenance demand. Aerobic growth is coupled to consumption of NADH and therefore balancing for these conditions is only possible with measurements of uptake and excretion rates. On the other hand, an analysis of the elementary representation of the null space shows that (beside anaplerotic reaction rates) some additional intracellular reaction rates are not calculable by metabolite balancing under photoheterotrophic conditions, unless additional data become available (e.g., by isotropic tracer experiments). Nevertheless, metabolite balancing has turned out to be a valuable tool for analyzing the metabolism in the versatile purple nonsulfur bacteria.

We would like to thank Dietrich Flockerzi, MPI Magdeburg, for critical reading of the mathematical part of the manuscript.

References

Ben-Israel A, Greville TNE. 1974. Generalized inverses: Theory and applications. New York: John Wiley & Sons.

Bonarius HPJ, Schmid G, Tramper J. 1997. Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. Trends Biotechnol 15:308–314.

Buchanan BB, Evans MCW, Arnon DI. 1967. Ferredoxin-dependent carbon assimilation in *Rhodospirillum rubrum*. Arch Mikrobiol 59:32–40.

Conrad R, Schlegel HG. 1977. Influence of aerobic and phototrophic growth conditions on the distribution of glucose and fructose carbon into the Entner-Doudoroff and Embden-Meyerhof pathways in *Rhodopseudomonas sphaeroides*. J Gen Microbiol 101:277–290.

Cornet JF, Dussap CG, Gro JB. 1998. Kinetics and energetics of photosynthetic microorganisms in photobioreactors. Adv Biochem Eng/Biotech 59:153–224.

Cornet JF, Albiol, J. 2000. Modeling photoheterotrophic growth kinetics of *Rhodospirillum rubrum* in rectangular photobioreactors. Biotechnol Progr 16:199–207.

Edwards JS, Palsson BO. 1999. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. J Biol Chem 274:17410–17416.

Edwards JS, Ibarra RU, Palsson BO. 2001. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat Biotechnol 19:125–130.

Fuller RC. 1995. Polyesters and photosynthetic bacteria. In: Blankenship RE, Madigan MT, Bauer CE, editors. Anoxygenic photosynthetic bacteria. Dordrecht: Kluwer Academic Publishers, p 1245–1256.

Göbel F. 1978. Quantum efficiencies of growth. In: Clayton RK, Sistrom WR, editors. The photosynthetic bacteria. New York: Plenum Press. p 908–925.

Hai T, Ahlers H, Gorenflo V, Steinbüchel A. 2000. Axenic cultivation of anoxygenic phototrophic bacteria, cyanobacteria, and microalgae in a new closed tubular glass photobioreactor. Appl Microbiol Biotechnol 53:383–389.

Happel J, Sellers PH. 1989. The characterization of complex systems of chemical reactions. Chem Eng Comm 83:221–240.

Heinrich R, Schuster S. 1996. The regulation of cellular systems. New York: Chapman & Hall.

Holms WH. 1986. 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. Curr Top Cell Regul 28:69–105.

Joshi HM, Tabita FR. 2000. Induction of carbon monoxide dehydrogenase to facilitate redox balancing in a ribulose bisphosphate carboxylase/oxygenase-deficient mutant strain of *Rhodospirillum rubrum*. Arch Microbiol 173:193–199.

Kaplan S. 1978. Control and kinetics of photosynthetic membrane development. In: Clayton RK, Sistrom WR, editors. The photosynthetic bacteria. New York: Plenum Press. p 808–839.

Matsunaga T, Hatano T, Yamada A, Matsumoto M. 2000. Microaerobic hydrogen production by photosynthetic bacteria in a double-phase photobioreactor. Biotechnol Bioeng 68:647–651.

Mavrovouniotis ML, Stephanopoulos G, Stephanopoulos G. 1990. Computer-aided synthesis of biochemical pathways. Biotechnol Bioeng 36:1119–1132.

McEwan AG. 1994. Photosynthetic electron transport and anaerobic metabolism in purple non-sulfur phototrophic bacteria. Antonie van Leeuwenhoek 66:151–164.

Meléndez-Hevia E, Waddell TG, Heinrich R, Montero F. 1997. Theoretical approaches to the evolutionary optimization of glycolysis - chemical analysis. Eur J Biochem 244:527–543.

Mendes P. 1997. Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3. Trends Biochem Sci 22:361–363.

Muller FM. 1933. On the metabolism of purple sulfur bacteria in organic media. Arch Mikrobiol 4:131–166.

Neidhardt FC, Ingraham JL, Schaechter M. 1990. Physiology of the bacterial cell: A molecular approach. Sunderland: Sinauer Associates.

Noorman HJ, Romein, B, Luyben KCAM, Heijnen JJ. 1996. Classification, error detection, and reconciliation of process information in complex biochemical systems. Biotechnol Bioeng 49:364–376.

Ormerod JG. 1956. The use of radioactive carbon dioxide in the measurement of carbon dioxide fixation in *Rhodospirillum rubrum*. Biochem J 64:373–380.

Ormerod JG, Sirevag R. 1983. Essential aspects of carbon metabolism. In: Ormerod JG, editor. The phototrophic bacteria: Anaerobic life in the light. Oxford: Blackwell Scientific Publications. p 100–119.

Pfeiffer T, Sánchez-Valdenebro I, Nuño JC, Montero F, Schuster S. 1999. METATOOL: For studying metabolic networks. Bioinformatics 15:251-257.

Pfeiffer T, Schuster S, Bonhoeffer S. 2001. Cooperation and competition in the evolution of ATP producing pathways. Science 292:504–507.

Reder C. 1988. Metabolic control theory: A structural approach. J theor Biol 135:175–201.

Richardson DJ, King GF, Kelly DJ, McEwan AG, Ferguson SJ, Jackson JB. 1988. The role of auxiliary oxidants in maintaining redox balance during photosynthetic growth of *Rhodobacter capsulatus* on propionate and butyrate. Arch Microbiol 150:131–137.

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

Saier MH, Feucht BU, Roseman S. 1971. Phosphoenolpyruvate-dependent fructose phosphorylation in photosynthetic bacteria. J Biol Chem 246:7819–7821.

Savinell JM, Palsson BO. 1992. Optimal selection of metabolic fluxes for *in vivo* measurement. I. Development of mathematical methods. J Theor Biol 155:201–214.

Schuster S, Hilgetag C. 1994. On elementary flux modes in biochemical reaction systems at steady state. J Biol Syst 2:165–182.

Schuster S, Schuster R. 1991. Detecting strictly detailed balanced subnetworks in open chemical reaction networks. J Math Chem 6:17–40.

Schuster S, Fell D, Dandekar T. 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature Biotechnol 18:326–332.

Stephanopoulos GN, Aristidou AA, Nielsen J. 1998. Metabolic engineering. San Diego: Academic Press.

Strang G. 1980. Linear algebra and its applications. New York: Academic Press.

Tabita FR. 1995. The biochemistry and metabolic regulation of carbon metabolism and CO_2-fixation in purple bacteria. In: Blankenship RE, Madigan MT, Bauer CE, editors. Anoxygenic photosynthetic bacteria, Dordrecht: Kluwer Academic Publishers. p 885–914.

Vallino JJ, Stephanopoulos G. 1993. Metabolic flux distributions in *Corynebacterium glutamaticum* during growth and lysine overproduction. Biotechnol Bioeng 41:633–646.

Van der Heijden RTJM, Heijnen JJ, Hellinga C, Romein B, Luyben KChAM. 1994a. Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balancebility of conversion rates. Biotechnol Bioeng 43:3–10.

Van der Heijden RTJM, Romein B, Heijnen JJ, Hellinga C, Luyben KChAM. 1994b. Linear constraint relations in biochemical reaction systems: II. Diagnosis and estimation of gross errors. Biotechnol Bioeng 43:11–20.

Varma A, Palsson BO. 1993. Metabolic capabilities of *Escherichia coli*: I. Synthesis of biosynthetic precursors and cofactors. J Theor Biol 165:477–502.

Wang X, Falcone, DL, Tabita FR. 1993. Reductive pentose phosphate-independent CO_2 fixation in *Rhodobacter sphaeroides* and evidence that ribulose bisphosphate carboxylase/oxygenase activity serves to maintain the redox balance of the cell. J Bac 175(11):3372–3379.

Wiechert W, Möllney M, Isermann N, Wurzel M, De Graaf AA. 1999. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. Biotechnol Bioeng 66:69–85.