Decomposition into Minimal Flux Modes: A Novel Flux-Balance Approach to Predict Flux Changes in Metabolic Networks from Changes of Enzyme Concentrations

Sabrina Hoffmann, Andreas Hoppe and Hermann-Georg Holzhutter

Institut für Biochemie, Universitätsmedizin Berlin (Charité), Humboldt-Universität zu Berlin, Monbijoustr. 2, 10117 Berlin, Germany

E-Mail: hergo@charite.de

Received: 25th September 2006 / Published: 31st August 2007

Abstract

The dynamic behaviour of metabolic networks is determined by the kinetic properties and the cellular levels of the enzymes and transporters involved. Changes in the concentrations of enzymes can be assessed by proteomics measurements or – more indirectly – by gene expression analyses. However, a straightforward interpretation of such data with respect to metabolic functions of the cell is difficult as a simple correlation between changes of enzyme levels and changes of fluxes in a metabolic network does not exist. Here we outline a theoretical concept to exploit information on changes of enzyme concentrations for predicting changes of stationary fluxes and this way to characterize changes in the functional status of cells or tissues. The basis of our concept is a novel variant of flux-balance analysis which we call *MinMode-decomposition*. The basic idea of this concept is to approximate flux distributions in metabolic networks as linear combinations of functionally motivated minimal flux modes (*MinModes*). They are defined as minimal flux modes supporting a unit flux through only one of the target reactions of the network. This theoretical concept will be applied to metabolic networks of bacteria (*Methylobacterium extorquens*) and human red blood. Based on simu-

lated data we demonstrate that a good prediction of observed flux changes can be achieved if the decomposition of flux changes into MinModes is performed such that a maximal correlation with observed changes in enzyme activities is accomplished.

Introduction

All cellular functions are ultimately linked to the presence of metabolites (such as proteins, nucleotides, fatty acids, phospholipids etc.) produced by the so-called metabolic network comprising thousands of enzyme-catalysed chemical reactions and carrier-mediated transport processes. The rate (herein called flux) through a given process, i.e. the amount of material chemically converted or transported per time unit, is controlled by various regulatory mechanisms. The set of all fluxes in a metabolic network is called flux distribution. The flux distribution may dramatically change with changing functional status of the cell (e.g. turning on the glycolytic flux when switching from the resting to the working muscle). It is an important goal of quantitative biochemistry to determine the flux distribution that determines the functionality of the cell. Such studies may help to reveal the relative importance of a specific enzyme and to predict the impact on the flux distribution if the enzyme is not active, e.g. due to a mutation or due to the administration of an enzyme inhibitor. The latter aspect is of central importance for the development of novel drugs interfering with the cellular metabolism.

Experimental determination of metabolic flux rates by means of tracer studies is time-consuming and tedious. Therefore, various mathematical concepts have been developed to analyse the full spectrum of flux modes possible in a metabolic network (structural analysis) or to predict flux distributions (semi-quantitative analysis). The common basis for all these concepts is the stoichiometric matrix $S=(S_{ij})$ representing the number of molecules of metabolite (i) formed or utilized in reaction (j). The stoichiometric matrix $\mathbf{S}$ is a $m \times n$ matrix where m corresponds to the number of metabolites and n is the number of reactions for which at least one catalysing enzyme is available in a given cell type [1]. The presence of a particular reaction can be evidenced by biochemical studies or – with some precaution – deduced from proteomic or genomic data [2 – 5].

Most modelling approaches assume the spatial distribution of metabolites to be homogeneous so that the kinetic behaviour of the network can be described by a system of ordinary differential equation systems,

$$\frac{d[X_i]}{dt} = \sum_{j=1}^{m} S_{ij}\, v_j \qquad (1)$$

$[X_i]$ is the concentration of the i-th metabolite ($i = 1,2,...,m$) and v_j denotes the flux through the j-th reaction ($j = 1,2,...,n$). The fluxes v_j constitute the so-called flux vector $\mathbf{v} = (v_j)$, in this paper referred to as a flux distribution. For quasi-stationary metabolic states where changes of the external conditions of the cell (e.g. changes of substrate concentrations or

hormonal effectors) are slow compared with the characteristic time-dependent response of the intra-cellular metabolism, a further simplification in the mathematical description of the network can be achieved by calculating the stationary solution of equation system (1),

$$\sum_{j=1}^{m} S_{ij}\, v_j = 0 \qquad (2)$$

The most advanced and satisfactory modelling approach is to solve the equation systems (1) or (2) with explicit flux vector **v** composed of rate equations relating the fluxes to the concentrations of the metabolites and external signals. However, such a straightforward modelling approach requires detailed knowledge of the kinetic properties of each participating enzyme. Even for the relatively simple metabolic network of bacteria as, for example, *Escherichia coli* this information is currently only available for the minority of enzymes involved. Therefore, computational studies of whole-cell metabolic networks demand alternative mathematical concepts. Existing non-kinetic concepts can be subdivided into two categories: Structural methods of network analysis and flux-balance analysis (FBA). Structural network analysis aims at exploring the full set of flux modes that may exist in a network with known stoichiometry. Simply speaking, these methods provide an overview of the many routes along which a given metabolite can be converted into another metabolite. Various algebraic concepts have been developed to define a basic set of "fundamental" flux modes which linearly combine to all possible flux modes in the network [1, 6, 7]. The definitions of such basic flux modes differ in the way that the reversible reactions are partitioned in forward and backward rates [8]. The two most prominent "fundamental" sets of flux modes are the so-called elementary modes [7] and the extremal pathways [6]. They have been used to re-define metabolic pathways [9], to check the robustness of the metabolic network against enzyme knock-outs [10, 11], the identification of thermodynamically infeasible cycles [12, 13] and the determination of so-called minimal cut sets, i.e. minimal sets of enzymes that have to be knocked out in order to completely abolish the flux through a given set of reactions [14]. The main obstacle in the application of structural analyses to large networks is the enormous number of possible flux modes arising from the combinatorial multiplicity with which single reactions can be composed to a longer route. For example, for a simplified metabolic network of *Escherichia coli* consisting of 106 reactions and producing five different end products from one initial substrate, 27100 elementary modes exist. Addition of three further initial substrates increases the number of elementary modes to 507632 [15]. This combinatorial explosion [16] implies that basic mode sets for networks with several hundreds of reactions cannot be calculated on commonly available computers. A lot of effort has been put into the development of faster algorithms to reduce computation time [15]. However, even if better algorithms allow the computation of fundamental modes for larger networks, inspection of all these modes and evaluation of their physiological significance remains extremely difficult. Attempts have also been undertaken to decrease the size of fundamental mode sets by incorporating additional constraints, for example, transcriptional regulation and environmental conditions [17] or kinetic and physiological feasibility [18, 19] into the computing algorithm. However, the formulation of such constraints requires profound *a priori* knowledge of physiological and regulatory details.

Besides structural network analyses, the concept of flux-balance analysis (FBA) has become a widely used method to estimate unknown fluxes in metabolic networks [20]. FBA postulates an objective function relating the flux distribution to a specific physiological function of the cell and to determine a flux distribution that optimizes this objective function. The idea behind this approach is that cells are capable of setting up an optimal flux distribution to produce a functionally relevant metabolic output. Most applications of FBA have used an objective function that considers only a single cardinal function of the cell as, for example, the accumulation of cell material (biomass) during the S-phase of the cell cycle. However, even primitive cells have to generate a metabolic output that simultaneously meets several functional demands. To overcome the restriction of FBA to monofunctional objective functions we have recently proposed the principle of flux minimization [21 – 23]. According to this principle, functionally relevant target fluxes, i.e. fluxes generating metabolites that are either used as building blocks for the synthesis of complex biomolecules or exported, should be accomplished with a minimal sum of internal network fluxes.

This work presents some ideas how the concept of structural network analyses can be unified with the concept of flux-minimization. As exemplary metabolic networks we will consider the central metabolism of *Methylobacterium extorquens* and the redox- and energy metabolism of human erythrocytes: Both networks have already been studied in previous work of our group [24, 25]. In the first part of this paper we show, only very few elementary modes are actually needed to decompose flux distributions calculated by flux-balance methods. However, the decomposition of the FBA solution into elementary modes is not unique and not all of the fundamental modes used in this decomposition allow for a clear physiological interpretation. Therefore, in the second part of the paper we propose a new type of fundamental modes which we call minimal flux modes (or short: *MinModes*). They are defined as minimal flux modes required to maintain a unit flux through a single target reaction of the network. According to this definition, there are only as many *MinModes* as there are functionally relevant target fluxes (17 in the network of Methylobacterium and 4 in the network of erythrocytes). Although MinModes do not form a basis in strict mathematical sense the examples considered in this article suggest that they can be linearly combined to provide a good approximation of a given flux distribution. The striking advantage of such a representation in terms of MinModes is that the coefficients used for the linear combination have a clear physiological meaning: they represent the metabolic output of the network in a given steady state and thus can be used as a measure for the functionality of the cell. Parts of these results have already been published [26].

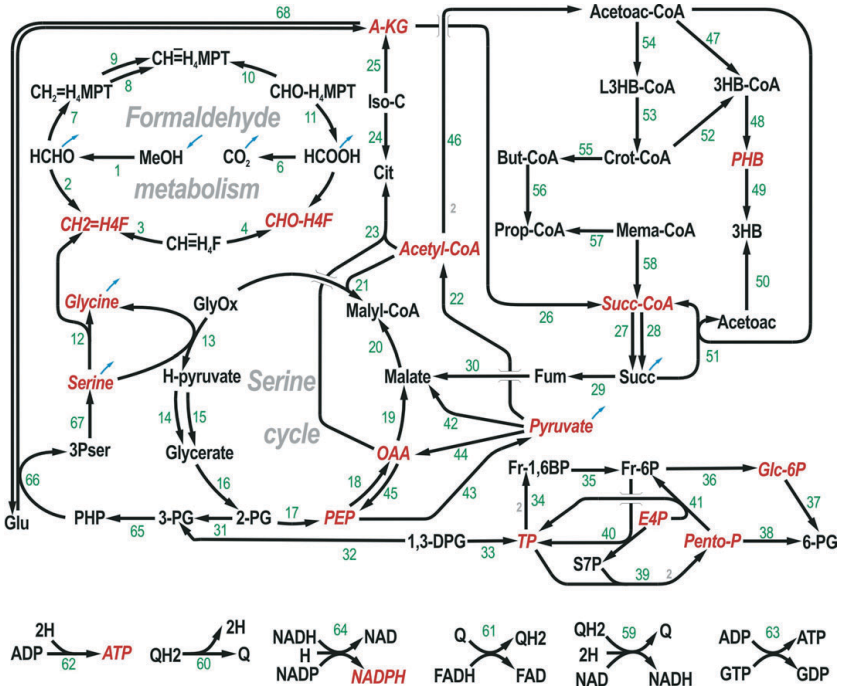
In the third part of this article we study the relationship between changes of enzyme levels and changes in the flux distribution. As consistent data sets encompassing changes in the expression levels of metabolic enzymes and measured flux changes in the network are not available yet, we settle this study on simulated data. We take the change of the v_{max} value of an enzyme as a measure for the change of its concentration. Subsequently we use the validated kinetic model for the erythrocyte network to simulate changes of the steady state fluxes elicited by changes in the v_{max} values of the participating enzymes. As expected,

there is no simple correlation between these two changes. However, using the decomposition of flux changes into MinModes as a side constraint we show that the prediction of flux changes from changes of enzyme levels can be greatly improved.

Exemplary Networks

The theoretical concepts considered in this paper will be applied to metabolic networks of two different cell types: The bacterium *Methylobacterium extorquens* AM1 (in the following referred to a B-network) and the human erythrocyte (in the following referred to as E-network). The metabolic scheme for the B-network was originally published by Van Dien [24, 25]. The scheme used in this paper contains some corrections which we have made in the light of recent findings [27, 28]. The B-network comprises 68 internal chemical reactions (43 of which are considered reversible), 8 exchange processes with the extracellular medium (for methanol, succinate, carbon dioxide, formate, formaldehyde, pyruvate, glycine, and serine) and 17 target reactions producing those metabolites required for biomass synthesis (see Table 1 for more details). While the model includes methanol, succinate, and pyruvate as alternative substrates, in all calculations methanol was considered the only available carbon source.

The reaction scheme for the E-model was originally published by Heinrich, Schuster and Holzhütter [29, 30]. It comprises 22 internal reactions, 4 exchange processes (for glucose, phosphate, pyruvate and lactate) and 4 target reactions delivering those metabolites which are essential for the integrity and functionality of the erythrocyte (2,3-bisphosphoglycerate, ATP, glutathione and phosphoribosyl pyrophosphate). For the E-network a detailed kinetic model is available [30] which takes into account all known kinetic properties of the participating enzymes. This kinetic model allows the computation of reliable stationary and time-dependent metabolic states which can be compared with results obtained by the FBA outlined in this paper. The reaction schemes for both networks are shown in Fig. 1. The involved reactions are explained in the legend of this figure.



$$\mathrm{si}\left(\mathrm{MM},\mathrm{MM}^{'}\right)=\frac{1}{\mathrm{N_r}}\sum_{\mathrm{i=1}}^{\mathrm{N_r}}\rho\left(\mathrm{MM_i},\mathrm{MM_i^{'}}\right),\quad \rho\left(\mathrm{MM_i},\mathrm{MM_i^{'}}\right)=\begin{cases}1 \text{ if } \mathrm{sgn}(\mathrm{MM_i})=\mathrm{sgn}(\mathrm{MM_i^{'}})\\ 0 \text{ else}\end{cases}$$

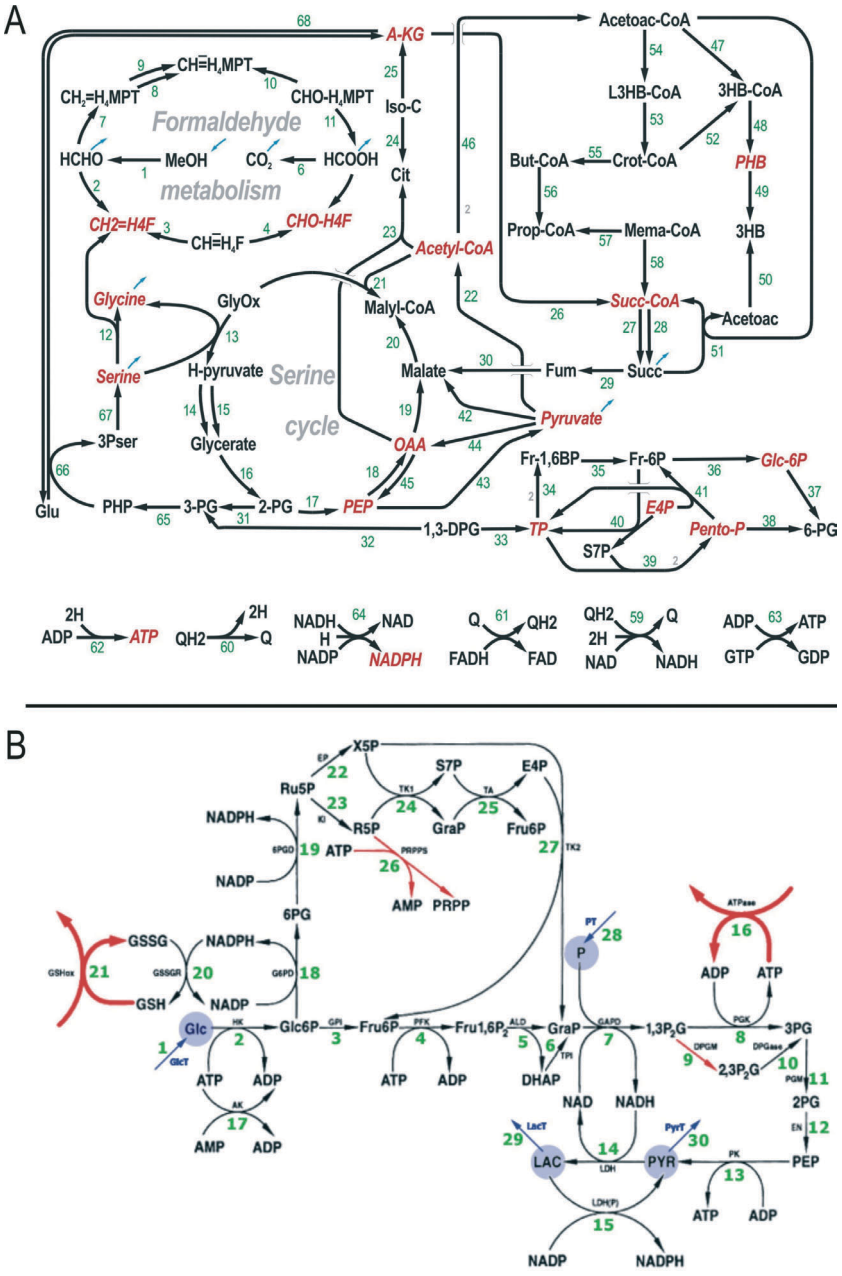
Figure 1. Metabolic networks considered in this article.

A. B-network: Reaction scheme of the central metabolism of *Methylobacterium extorquens*. The reaction arrows point in the direction of the net reaction under standard conditions. Compound names in red italic indicate utilization or generation of the corresponding metabolite during biomass production, blue arrows indicate exchange fluxes with the external environment. Cofactors have been dropped for better readability. The complete reaction scheme is shown in Table1A. The scheme is based on information outlined in [21] and derived from the KEGG data base (http://www.genome.ad.jp/kegg/).

Reactions/Enzymes:

1-methanol dehydrogenase (1.1.1.244), 2-not catalysed, 3-methylene H_4F dehydrogenase (MtdA)(1.5.1.5), 4-methenyl H_4F cyclohydrolase (3.5.4.9), 5-formyl H_4F synthetase (6.3.4.3), 6-formate dehydrogenase (1.2.1.2), 7-formaldehyde-activating enzyme, 8-methylene H_4MPT dehydrogenase (MtdB), 9-methylene H_4MPT dehydrogenase (MtdA) n/a, 10-methenyl H_4MPT cyclohydrolase (3.5.4.27), 11-formyl MFR:H_4MPT formyltransferase (1.2.99.5), 12-formyl MFR dehydrogenase (1.2.99.5), 13-serine hydroxymethyltransferase (2.1.2.1), 14- serine-glyoxylate aminotransferase (2.6.1.45), 15-hydroxypyruvate reductase (1.1.1.81), 16-glycerate kinase (2.7.1.31), 17-enolase (4.2.1.11), 18-PEP carboxylase (4.1.1.31), 19-malate dehydro-

genase (1.1.1.37), 20-malate thiokinase (6.2.1.9), 21-malyl-CoA lyase (4.1.3.24), 22-pyruvate dehydrogenase (1.2.4.1), 23-citrate synthase (2.3.3.1), 24-aconitase (4.2.1.3), 25-isocitrate dehydrogenase (1.1.1.42), 26-a-KG dehydrogenase (1.2.1.52), 27-succinyl-CoA synthetase (6.2.1.4), 28-succinyl-CoA hydrolase (3.1.2.3), 29-succinate dehydrogenase (1.3.5.1), 30-fumarase (4.2.1.2), 31-phosphoglycerate mutase (5.4.2.1) ,32-phosphoglycerate kinase (2.7.2.3), 33-glyceraldehyde-3-P dehydrogenase (1.2.1.12), 34-aldolase (4.1.2.13), 35-fructose-1,6-bisphosphatase (3.1.3.11), 36-phosphoglucose isomerase (5.3.1.9), 37-glucose-6-phosphate dehydrogenase (1.1.1.49), 38 – 6-phosphogluconate dehydrogenase (1.1.1.44), 39 transketolase (2.2.1.1), 40-transaldolase (2.2.1.2), 41- transketolase (2.2.1.1), 42-malic enzyme (1.1.1.38), 43-pyruvate kinase (2.7.1.40), 44-pyruvate carboxylase (6.4.1.1), 45-PEP carboxykinase (4.1.1.32), 46- b-ketothiolase (2.3.1.16), 47- acetoacetyl-CoA reductase (NADPH) (1.1.1.36), 48- PHB synthase (2.3.1._), 49- PHB depolymerase (3.1.1.75), 50 b-hydroxybutyrate dehydrogenase (1.1.1.30), 51-acetoacetate-succinyl-CoA transferase (2.8.3.5), 52-d-crotonase (4.2.1.17), 53 l-crotonase (4.2.1.17), 54-acetoacetyl-CoA reductase (NADH) (1.1.1.35), 55-crotonyl-CoA reductase (1.3.1.8), 56-unknown pathway, 57-propionyl-CoA carboxylase (6.4.1.3), 58-methylmalonyl-CoA mutase (5.4.99.2), 59-NADH-quinone oxidoreductase (1.6.99.5), 60-cytochrome oxidase (1.10.2.2), 61- ubiquinone oxidoreductase (1.5.5.1), 62-ATPase, 63-NDP kinase (2.7.4.6), 64-transhydrogenase (1.6.1.2), 65 – 3-phosphoglycerate dehydrogenase (1.1.1.95), 66-phosphoserine transaminase (2.6.1.52), 67-phosphoserine phosphatase (3.1.3.3), 68-glutamate dehydrogenase (1.4.1.4)



B. E-network: Reaction scheme of the energy- and redox metabolism of human erythrocytes.

Reaction/Enzymes:

1-glucose transporter GlcT, 2-hexokinase HK (2.7.1.1), 3-phosphohexose isomerase GPI (5.3.1.9), 4-phosphofructokinase PFK (2.7.1.11), 5-aldolase ALD (4.1.2.13), 6-triosephosphate isomerase TPI (5.3.1.1), 7-triosephosphate dehydrogenase (NAD) GAPDH (1.2.1.12), 8-phosphoglycerate kinase PGK (2.7.2.3), 9-bisphosphoglycerate mutase DPGM (5.4.2.4), 10-bisphosphoglycerate phosphatase DPGase (3.1.3.13), 11-phosphoglycerate mutase PGM (5.4.2.1), 12-enolase EN (4.2.1.11), 13-pyruvate kinase PK (2.7.1.40), 14-lactate dehydrogenase LDH (NADH) (1.1.1.28), 15-lactate dehydrogenase LDH (NADPH) (1.1.1.28), 16-ATPase (total) ATPase, 17-myokinase (adenylate kinase) AK (2.7.4.3), 18-glucose-6-phosphate dehydrogenase G6PD (1.1.1.49), 19-phosphogluconate dehydrogenase 6PGD (1.1.1.44), 20-glutathione reductase GSSGR (1.8.1.7), 21-glutathione oxidation (total) GSHox, 22-phosphoribulose epimerase EP (5.1.3.1), 23-ribose phosphate isomerase KI (5.3.1.6), 24-transketolase (1) TK1 (2.2.1.1), 25-transaldolase TA (2.2.1.2), 26-phosphoribosylpyrophosphate synthetase PRPPS (2.7.6.1), 27-transketolase (2) TK2 (2.2.1.1), 28-phosphate transporter PT, 29- lactate exchange LacT, 30- pyruvate exchange PyrT

Computational Methods

Calculation of flux-minimized steady-state flux distributions

The computation of flux modes is based on flux balance analysis (FBA). The core of this method is the optimization of an objective function which relates the flux distribution to cellular functions. According to the principle of flux minimization [31] the objective function to be minimized is chosen as:

$$\Phi = \sum_{j} \mathrm{pos}\left(v_j\right) + K_j \ \mathrm{neg}\left(v_j\right) \tag{3}$$

where the sum runs over all fluxes in the network and K_j denotes the equilibrium constant for the j-th reaction. The real functions pos(x) and neg(x) return the absolute value of the argument x if $x \geq 0$ and $x \leq 0$, respectively, and otherwise 0. The functional state of the cell is defined by fixing non-zero fluxes through so-called 'target reactions' which together with the steady-state represent constraints of the minimization problem. The constrained flux minimization problem is solved using the software package CPLEX [32]. Details of the computational protocol have been described elsewhere [31].

The *in vivo* state of the B-network is determined by the following values of the fluxes through the 17 target reactions involved in the production of biomass:
V_{77}= 13.4 (glycine), v_{82}= 1.96 (CH_2=H_4F), v_{86}= 11.1 (pep), v_{92}= 5.09 (ery4P), v_{89}= 1.92 (tp), v_{85}= 7.24 (serine), v_{83}= 11.0 (CHO-H_4F), v_{93}= 92.9 (phb), v_{90}= 16.4 (glc6P), v_{88}= 17.1 (akg), v_{84}= 41.8 (pyruvate), v_{81}= 53.5 (acetylCoA), v_{87}= 41.1 (oaa), v_{80}= 6.8 (succCoA), v_{78}= 585.3 (atp), v_{91}= 10.4 (pentoP), v_{79}= 235.13 (nadph)
The unit of fluxes is moles precursor metabolite per 1000 g atoms C in biomass. Release of NADH during biomass production was included in the target flux for NADPH. The relations between the target flux, i. e. the stoichiometry with which the 17 precursor metabolites enter the biomass, have been determined by Van Dien [24].

The *in vivo* state of the E-network is determined by the following values of the fluxes through the 4 target reactions:
$v_9 = 0.49$ (2,3DPG), $v_{16} = 2.38$ (ATP), $v_{21} = 0.093$ (GSH), $v_{26} = 0.026$ (PRPP)
The unit of fluxes is mM/h.
In the following we will refer to the solution of the minimization problem (3) fulfilling the steady-state conditions (2) and providing the above values of target fluxes as the *global flux minimum.*

Decomposition of the global flux minimum into elementary modes
For the two exemplary networks, elementary modes and the convex basis of elementary modes were computed using a recent version of the software tool FluxAnalyzer [33]. Decomposition of *global flux minimum* into the convex basis was performed by solving the linear program

$$\mathbf{e}\,\mathbf{C} = \mathbf{v}_g \;,\; \|\mathbf{e}\| \text{ minimal} \tag{4}$$

where $\mathbf{C}$ is the convex basis of elementary modes written as a matrix, $\mathbf{v}_g$ is the global flux minimum and $\mathbf{e}$ is a vector of non-negative real numbers.

Definition of minimal flux modes (MinModes)
Besides the *global flux minimum* defined as the optimal flux distribution in the network with all target fluxes kept at pre-defined non-zero values, we computed special flux-minimized steady-states by putting the value of only one of the target fluxes to either unity (in the used flux units). The resulting minimized flux modes we call minimal flux modes (short: MinModes). They are defined as follows:

> A MinMode is a minimal (according to the flux minimization principle) steady state flux distribution that accomplishes a unit flux through one of the (independent) target reactions whilst the fluxes through the other target reactions are zero.

This definition is more rigorous than that given in [26], because it presumes the target fluxes to be independent from each other. Independency of target fluxes means the existence of a flux-minimized solution that accomplishes a non-zero flux through the chosen target flux without the necessity to have non-zero fluxes through other target reactions. This condition may be not always fulfilled. For example, if two metabolites (say A and B) are produced in one and the same reaction, then under steady state conditions utilization of A necessarily entails utilization of B. Hence, the first step towards the computation of MinModes requires to identify clusters of intrinsically coupled target reactions. In the following we will assume that the definition of 'target fluxes' in the network eventually includes clusters of coupled output reactions. For the two networks studied in this paper, the output fluxes are independent, i.e. there are 17 of such MinModes for the B-network and 4 MinModes for the E-network.

RESULTS

Elementary modes of the network of Methylobacterium

The model of central metabolism of *Methylobacterium extorquens* shown in Fig. 1 was subjected to elementary mode analysis. The network model consists of 93 reactions whereby 43 reactions are considered irreversible. The complete set of elementary modes for this model is too large to be computable by means of the program *FluxAnalyzer* [33] using a PC equipped with a memory capacity of 768MB. The enormous amount of elementary modes network can be envisaged by noting that there are 450251 elementary modes if the computation is simplified by setting 5 of the 17 target fluxes to zero. However, for the representation of an arbitrary flux distribution knowledge of the convex basis of elementary modes is sufficient [34]. A convex basis of elementary modes was computable for the complete B-network. It consists of 7033 elementary modes. Intriguingly, only 21 elementary flux modes (out of 7033) of the convex basis were actually needed (i.e. had non-zero coefficients in the linear representation) for the decomposition of the *global flux minimum*. Regarding the physiological interpretation of these 21 elementary flux modes further analysis showed that most of them contain more than one non-zero target flux (see Fig. 2). Moreover, the choice of basic elementary modes is not unambiguous. This makes it difficult to assign a specific output of the network to these elementary flux modes.

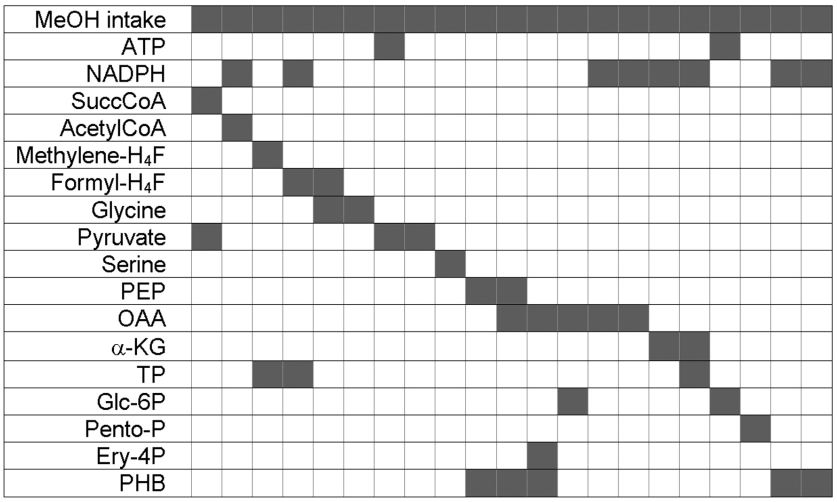


Figure 2. Occurrence of non-zero input and output fluxes in the 21 elementary modes required for the decomposition of the global flux minimum for the B-network. Each column represents an elementary mode, each row represents an input/output flux. Table fields shaded in grey indicate a non-zero value of the respective flux.

Elementary modes of the network of the Erythrocyte

The model of central metabolism of the human Erythrocyte has 20 elementary modes. Only 4 elementary modes were actually needed (i.e. had non-zero coefficients in the linear representation) for the decomposition of the global flux minimum where each of these modes exactly corresponds to a singular target flux. Thus, for this simple model the difficulties mentioned for the B-network do not occur.

Calculation of MinModes

According to the definition given above, MinModes are flux-minimized states of the network where only one of the (independent) target reactions carries a unit flux whereas the flux through all other target reactions is put to zero. As an important property, each of these modes is associated with a specific output of the network. For the B-network (*Methylobacterium extorquens)* each MinMode represents the minimal flux distribution required for the synthesis of a single biomass precursor metabolite. For the E-network (erythrocyte) the target metabolites ATP and GSH are essential for cellular integrity (maintenance of the intracellular ionic milieu by the ATP-driven Na-K-pump, protection against secondary reaction products of radicals via GSH), the two other target metabolites 2,3DPG and PRPP are indispensable for oxygen binding to haemoglobin and the salvage of adenine nucleotides. The MinModes for the two networks are listed in Table 1.

Inspecting the MinModes of the B-network, we note that the amount of methanol consumed in each of the MinModes can be split into a carbon providing and an energy (ATP and redox equivalents) providing part. As an example, the MinMode for the production of the biomass precursor glycine is plotted in Fig. 3.

Table 1. Model scheme and the corresponding minimal flux modes (MinModes) of the B-network. *cmm* and *gfm* denote the *MinMode composition* and the *global flux minimum*, respectively.

#	reaction	k_1	acetyl/CoA 1	akG 2	ATP 3	CH_2=H4F 4	Ery4P 5	CHO-H4F 6	Glc6P 7	glycine 8	NADPH 9	OAA 10	Pentop 11	PEP 12	PHB 13	pyruvate 14	serine 15	Succ 16	TP 17	*cmm*	*gfm*
	Formaldehyde metabolism																				
1	MeOH + Q → HCHO + QH_2	100000	2	5	1	1	5	1	6.6	2.8	0.4	4	5.8	3.55	4	3.36	3.6	1	4.09	32.63	23.59
2	HCHO + H_4F → CH_2=H_4F	100000	1	3	0	1	2	1	2	1	0	2	2	2	2	0	2	0	2	10.23	10.23
3	$CH\equiv H_4F$ + NADPH ↔ CH_2=H_4F + NADP	7.14	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	-0.19	-0.19
4	$CH\equiv H_4F$ ↔ CHO-H_4F	2.3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0.19	0.19
5	HCOOH + H_4F + ATP ↔ CHO-H_4F + ADP	41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	HCOOH + NAD → CO_2 + NADH	420	1	2	0	0	3	0	4.6	1.8	0.4	2	3.8	1.55	2	1.36	1.6	0	2.09	12.09	13.37
7	HCHO + H_4MPT → CH_2=H_4MPT	100000	1	2	0	0	3	0	4.6	1.8	0.4	2	3.8	1.55	2	1.36	1.6	0	2.09	12.09	13.37
8	CH_2=H_4MPT + NAD → $CH\equiv H_4MPT$ + NADH	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1.62	0
9	CH_2=H_4MPT + NADP → $CH\equiv H_4MPT$ + NADPH	1	1	2	0	0	3	0	4.6	1.8	0.4	2	3.8	1.55	1	1.36	1.6	0	2.09	10.47	13.37
10	$CH\equiv H_4MPT$ ↔ CHO-H_4MPT	1	-1	-2	0	0	-3	0	-4.6	-1.8	-0.4	-2	-3.8	-1.55	-2	-1.36	-1.6	0	-2.09	-12.09	-13.37
11	HCOOH + H_4MPT ↔ CHO-H_4MPT	100000	-1	-2	0	0	-3	0	-4.6	-1.8	-0.4	-2	-3.8	-1.55	-2	-1.36	-1.6	0	-2.09	-12.09	-13.37
	Serine cycle																				
12	Serine + H_4F ↔ Glycine + CH_2=H_4F	1022	-1	-3	0	0	-2	0	-2	-1	0	-2	-2	-2	-2	-2	-2	0	-2	-10	-10
13	Serine + GlyOx ↔ Glycine + H-Pyruvate	1	1	3	0	0	2	0	2	2	0	2	2	2	2	2	2	0	2	10.23	10.23
14	H-Pyruvate + NADH → Glycerate + NAD	1	0	1	0	0	2	-1	4.4	2.2	0.6	1	3.2	2	2	2	2.4	0	2	11.26	10.23
15	H-Pyruvate + NADPH → Glycerate + NADP	1	1	2	0	0	0	1	-2.4	-0.2	-0.6	1	-1.2	0	0	0	-0.4	0	0	-1.02	0
16	Glycerate + ATP → 2-PG + ADP	100000	1	3	0	0	2	0	2	2	0	2	2	2	2	2	2	0	2	10.23	10.23
17	PEP ↔ 2-PG	3	-1	-3	0	0	-1	0	-1	-1	0	-2	-1	-2	-2	-2	-1	0	-1	-9.28	-9.28
18	PEP + CO_2 → OAA	1	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	3.27	1.01
19	OAA + NADH → Malate + NAD	6260	1	0	0	0	0	3	0	0	0	1	0	0	0	0	0	0	0	1.65	0
20	Malate + CoA + ATP → Malyl-CoA + ADP	100000	1	3	0	0	2	0	2	2	0	2	2	2	2	2	2	0	2	10.23	10.23
21	Acetyl-CoA + GlyOx ↔ Malyl-CoA	345	-1	-3	0	0	-2	0	-2	-2	0	-2	-2	-2	-2	-2	-2	0	-2	-10.23	-10.23
	Citrate cycle																				
22	Pyruvate + CoA + NAD → Acetyl-CoA + NADH + CO_2	100000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	Acetyl-CoA + OAA → Cit + CoA	100000	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.3	0.3
24	Iso-C ↔ Cit	14.4	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.3	-0.3
25	α-KG + CO_2 + NADPH ↔ Iso-C + NADP	1.3	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.3	-0.3
26	α-KG + CoA + NAD → Succ-CoA + CO_2 + NADH	100000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	Succ + CoA + GTP ↔ Succ-CoA + GDP	1.68	0	-1	0	0	0	0	-1	-1	0	0	-1	-1	0	-1	-1	1	-1	-2.05	-2.77
28	Succ-CoA → Succ + CoA	100000	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0.72	0
29	Succ + FAD ↔ Fum + FADH	1	0	1	0	0	1	0	1	1	0	1	1	1	0	1	1	0	1	2.88	2.88
30	Fum ↔ Malate	4.86	0	1	0	0	1	0	1	1	0	1	1	1	0	1	1	0	1	2.88	2.88
	Gluconeogenesis and pentose phosphate pathway																				
31	2-PG ↔ 3-PG	7.3	0	0	0	0	1	0	1	1	0	0	1	0	0	0	1	0	1	0.95	0.95
32	1,3-DPG + ADP ↔ 3-PG + ATP	3226	0	0	0	0	-1	0	-1	0	0	0	-1	0	0	0	0	0	-1	-0.59	-0.59
33	1,3-DPG + NADH ↔ TP + NAD	2.77	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0.59	0.59
34	2 TP → Fru-1,6-BP	5555	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	Fru-1,6-BP ↔ Fru-6-P	174	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	Fru-6-P ↔ Glc-6-P	2.22	0	0	0	0	-1	0	-2	0	0	0	-2	0	0	0	0	0	0	-1.02	-1.02
37	Glc-6-P + NADP ↔ 6-PG + NADPH	2.08	0	0	0	0	-1	0	-3	0	0	0	-2	0	0	0	0	0	0	-1.31	-1.31
38	Pento-P + CO_2 + NADPH ↔ 6-PG + NADPH	13.5	0	0	0	0	1	0	3	0	0	0	2	0	0	0	0	0	0	1.31	1.31
39	TP + S7P ↔ 2 Pento-P	1.5	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0.47	0.47
40	TP + S7P ↔ Ery-4-P + Fru-6-P	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0.47	0.47
41	Ery-4-P + Pento-P ↔ TP + Fru-6-P	10	0	0	0	0	-1	0	-1	0	0	0	-1	0	0	0	0	0	0	-0.56	-0.56
42	Malate + NAD ↔ Pyruvate + CO_2 + NADH	1	0	-2	0	0	-1	0	-1	-1	0	0	-1	-1	-2	-1	-1	0	-1	-5.7	-7.35
43	PEP + ADP → Pyruvate + ATP	18000	0	2	0	0	1	0	1	1	0	0	1	1	2	2	1	0	1	6.43	8.08
44	Pyruvate + CO_2 + ATP ↔ OAA + ADP	6.55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45	OAA + GTP ↔ PEP + GDP + CO_2	12	0	0	0	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0.61	0
	PHB synthesis and glyoxylate regeneration cycle																				
46	2 Acetyl-CoA ↔ Acetoac-CoA + CoA	1	0	1	0	0	1	0	1	1	0	1	1	1	1	1	1	0	1	4.5	4.5
47	Acetoac-CoA + NADPH → 3HB-CoA + NADP	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1.62	1.62
48	3HB-CoA → PHB + CoA	100000	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1.62	1.62
49	PHB → 3HB	100000	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
50	Acetoac + NADH ↔ 3HB + NAD	526	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
51	Acetoac-CoA + Succ ↔ Acetoac + Succ-CoA	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(continuation on the next page)

Table 1: (Continued).

#	reaction	k_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	*cmm*	*gfm*
52	Crot-CoA ↔ 3HB-CoA	5.55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
53	L3HB-CoA ↔ Crot-CoA	5.55	0	1	0	0	1	0	1	1	0	1	1	1	0	1	1	0	1	2.88	2.88
54	Acetoac-CoA + NADH → L3HB-CoA + NAD	1587	0	1	0	0	1	0	1	1	0	1	1	1	0	1	1	0	1	2.88	2.88
55	Crot-CoA + NADPH ↔ But-CoA + NADP	1	0	1	0	0	1	0	1	1	0	1	1	1	0	1	1	0	1	2.88	2.88
56	But-CoA + NAD → Prop-CoA + CO_2 + NADH	100000	0	1	0	0	1	0	1	1	0	1	1	1	0	1	1	0	1	2.88	2.88
57	Mema-CoA + ADP ↔ Prop-CoA + ATP + CO_2	123	0	-1	0	0	-1	0	-1	-1	0	-1	-1	-1	0	-1	-1	0	-1	-2.88	-2.88
58	Mema-CoA ↔ Succ-CoA	18.6	0	1	0	0	1	0	1	1	0	1	1	1	0	1	1	0	1	2.88	2.88
	Energy metabolism																				
59	NADH + Q ↔ NAD + QH_2 + 2 H	1	0	-1	0	0	-1	1	-1.8	-0.4	-0.2	0	-1.4	-0.91	-1	-1.27	-0.8	0	-0.82	-4.72	-2.16
60	QH_2 ↔ Q + 2 H	1	2	5	1	1	5	2	5.8	3.4	0.2	5	5.4	3.64	3	3.09	3.8	1	4.27	30.78	24.31
61	FADH-S + Q ↔ FAD-S + QH_2	1	0	1	0	0	1	0	1	1	0	1	1	1	0	1	1	0	1	2.88	2.88
62	ADP + 2 H ↔ ATP	1	2	4	1	1	4	3	4	3	0	5	4	3	2	2	3	1	4	26.26	23.29
63	GDP + ATP ↔ GTP + ADP	1.04	0	-1	0	1	-1	3	-1	-1	0	0	-1	-1	0	-1	-1	1	-1	-1.44	-2.77
64	NADH + NADP + H → NADPH + NAD	1	0	0	0	0	0	0	0	0	0	0	0	-0.55	0	-0.36	0	0	-1.09	-0.41	-2.28
	Serine biosynthesis																				
65	PHP + NADH ↔ 3-PG + NAD	10000	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	-1	0	0	-0.36	-0.36
66	α-KG + 3-Pser ↔ PHP + Glu	6.66	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	-1	0	0	-0.36	-0.36
67	3-Pser ↔ Serine	673	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0.36	0.36
68	α-KG + NADPH → Glu + NADP	100000	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0.36	0.36
	Exchange reactions																				
69	CO_2 → CO_2(out)	1	0	0	0	0	1	0	0.6	0.8	0.4	0	0.8	0.55	0	0.36	0.6	0	1.09	4.77	6.05
70	CHOOH → CHOOH(out)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
71	Glycine → Glycine(out)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
72	HCHO → HCHO(out)	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	10.31	0
73	Pyruvate → Pyruvate(out)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
74	Serine → Serine(out)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
75	MeOH(out) → MeOH	1	2	5	1	1	5	1	6.6	2.8	0.4	4	5.8	3.55	4	3.36	3.6	1	4.09	32.63	23.59
76	Succ → Succ(out)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Additional reactions for biomass synthesis																				
77	glycine → glycine_biomass	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0.23	0.23
78	ATP → ATP_biomass + ADP	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10.2	10.2
79	NADPH → NADPH_biomass + NAD	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	4.1	4.1
80	Succ-CoA → Succ-CoA_biomass + Succ + CoA	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0.12	0.12
81	Acetyl-CoA → Acetyl-CoA_biomass + CoA	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.93	0.93
82	CH_2=H_4F → CH_2=H_4F_biomass + H_4F	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0.03
83	CHO-H_4F → CHO-H_4F_biomass + H_4F	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0.19	0.19
84	pyruvate → pyruvate_biomass	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0.73	0.73
85	serine → serine_biomass	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0.13	0.13
86	PEP → PEP_biomass	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0.19	0.19
87	OAA → OAA_biomass	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0.72	0.72
88	α-KG → α-KG_biomass	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.3	0.3
89	TP → TP_biomass	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.03	0.03
90	Glc-6-P → Glc-6-P_biomass	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0.29	0.29
91	Pento-P → Pento-P_biomass	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0.18	0.18
92	Ery-4-P → Ery-4-P_biomass	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0.09	0.09
93	PHB → PHB_biomass	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1.62	1.62

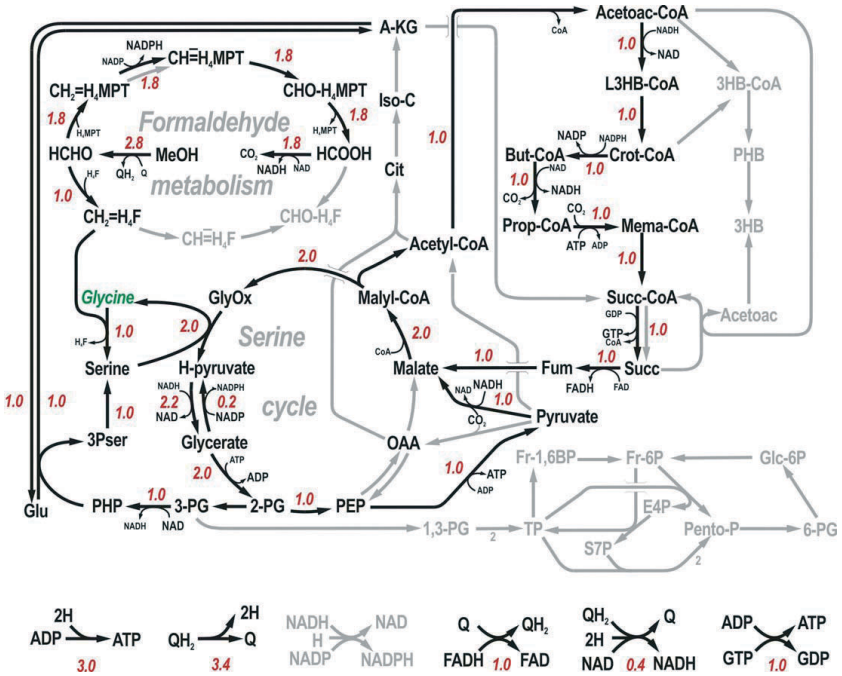


Figure 3. Minimal flux mode (*MinMode* $MM_{glycine}$) for glycine synthesis in the B-network. The MinMode was calculated by minimizing the objective function (3) where the flux v_{70} for the release of glycine to the biomass was put to unity and the other 17 target flux were put to zero. Flux values are depicted next to the reaction arrow, grey arrows indicate zero fluxes.

Here, the H_4MPT-cycle is used for the production of NADPH and NADH (utilized in other parts of the network) as well as for the complete oxidation of methanol to CO_2. The released carbon is fixed in another part of the network to provide the carbon needed for the actual synthesis of glycine. The second carbon atom is incorporated via reactions of the serine cycle. Only two of the reactions that traditionally form the citrate acid cycle, are used in this MinMode. They are catalysed by the succinate dehydrogenase (29) and the fumarase (30).

To illustrate the relative importance of the individual reactions for the functionality of the network we counted how often a non-zero flux through each of the reactions occurs in the 17 MinModes of the B-network. The corresponding statistics (Fig. 4) reveals the ubiquitous usage of reactions involved in methanol uptake and the energy metabolism. Based on the frequency with which a reaction has a non-zero flux in the various MinModes one may depict the 'backbone' of the network. For the B-network this backbone is constituted by the reactions of serine cycle and formaldehyde metabolism (H_4F and H_4MPT cycle). On the other hand, there are 15 apparently redundant reactions that have a zero-flux in all Min-Modes. This relatively high number of apparently abdicable reactions is certainly due to the

fact that only those reactions are considered as targets of the network which deliver metabolites for the synthesis of cellular biomass. Because metabolites other than the biomass precursors may be relevant for cellular functionality, we have calculated an extended set of MinModes considering each of the 63 metabolites as a possible and relevant output of the network. In this set of 58 MinModes only 6 reactions turn out to be apparently dispensable (reactions: 5, 22, 26, 34, 49 and 52).

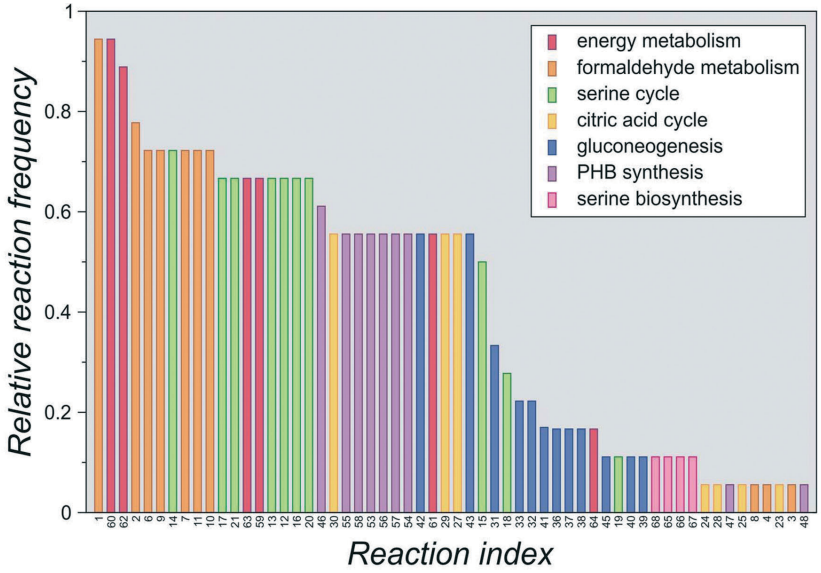


Figure 4. Frequency of non-zero fluxes in the 17 MinModes of the in the B-network

Inspecting the 4 MinModes of the E-network, which are equivalent to the 4 elementary modes required for the decomposition of the *global flux minimum*, the only difference between the two MinModes associated with the production of ATP and the production of 2,3DPG is the respective component referring to the particular target reaction.

Composition of flux distributions into MinModes

The MinModes represent canonical flux modes of the network supporting a single metabolic output (or a group of coupled outputs, see above). In real situations the flux distribution of a cell has to assure simultaneously a multitude of metabolic functions as, for example, the production of ATP, repair of DNA, elimination of reactive oxygen species or the synthesis of proteins. Because these metabolic functions must be controlled independently in the cell, it is straightforward to postulate that the total flux distribution is a linear combination of independent component fluxes rather than a globally optimized flux distribution. The concept presented here of minimal flux modes is an attempt to define those component fluxes by the following equation:

$$\mathbf{v_d} \approx \sum_i \alpha_i \, \mathbf{MM_i} \qquad (5)$$

Here $\mathbf{MM_i}$ denotes the MinMode supporting the flux 1 (flux unit) through the i-th target reaction and the numerical value of the (dimensionless) coefficient α_i corresponds to the actual flux. For example, applying the linear combination (4) of MinModes to the B-network, the coefficient α_{gly} that is multiplied with $MM_{gly,}$ the MinMode for the production of the biomass precursor glycine (Gly), is put to 13.4 as the measured flux of glycine into biomass amounts to 13.4 flux units.

To check the feasibility of the MinMode decomposition we compared the resulting flux distribution $\mathbf{v_d}$ (=*MinMode decomposition*) with the *global flux minimum* and with observed flux values. For the B-network, Fig. 5 depicts the values of the individual fluxes predicted by the *MinMode decomposition* and those of the *global flux minimum*.

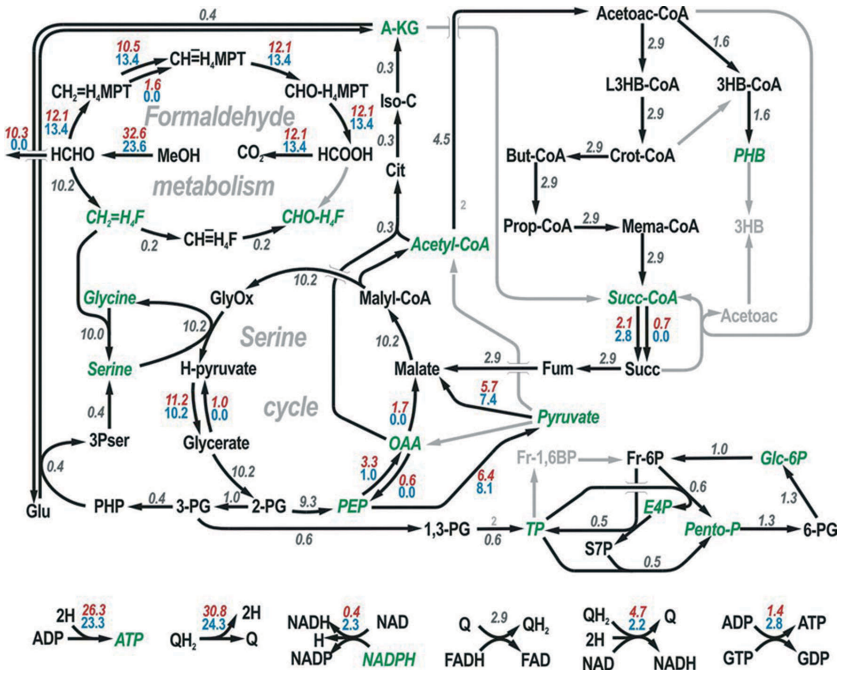
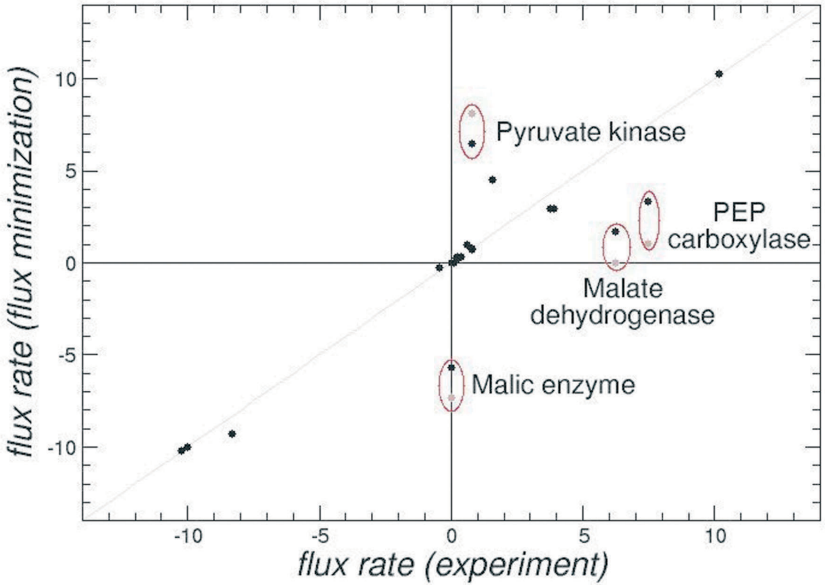


Figure 5. Comparison of flux values of the *global flux minimum* of the B-network (regular numbers) with flux values obtained by *MinMode composition* (italic numbers). Equal flux values in both approaches are displayed only once (grey/italic). All metabolites fed into biomass synthesis by target reactions are depicted with green/italic letters.

Obviously, the larger the target flux with which a precursor metabolite enters the biomass the higher the influence of the corresponding MinMode in the combination. In the B-network, the ATP consumption for biomass formation has a large impact. In contrast to an "along the way" synthesis of ATP in the *global flux minimum*, the MinMode MM_{78} exclusively provides ATP. This makes it plausible, why the *MinMode composition* predicts somewhat higher fluxes for ATP synthesis. Minor differences in the flux pattern occurred for reactions of the H_4MPT-cycle, transhydrogenase, NDP kinase, and reactions of alternative pathways that work with different cofactors, as for example the conversion of PEP into malate (reactions 18, 19, 45 and 42, 43) or from methylene-H_4MPT into methenyl-H_4MPT (reactions 8 and 9). All flux differences can be accounted for by differences in the production, conversion or dissipation of energy. To further check the feasibility of the *MinMode composition* we compared it with measured fluxes [24]. For 18 out of 21 reactions the fluxes predicted by the *MinMode composition* are in good accordance with the experimental data (Fig. 6A). In the cases of malic enzyme, malate dehydrogenase, PEP carboxylase, and pyruvate kinase, significant differences between predictions and observations occurred. Noteworthy, the remaining flux discrepancies are even smaller than those with respect to the *global flux minimum.*



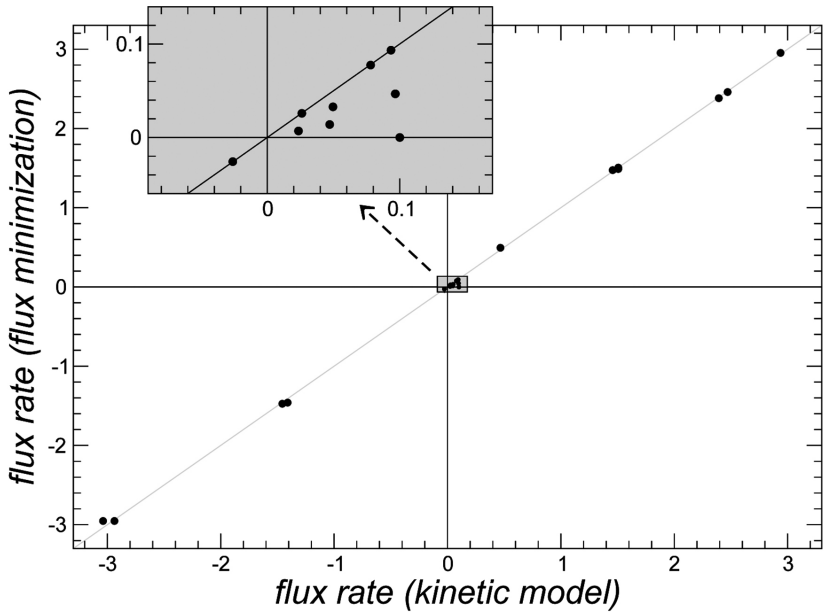


Figure 6. A.. B-network. Scattergram illustrating the correlation of experimental flux values [24] with flux values predicted by *MinMode composition.* Significant deviations between experimental flux values with flux values predicted by global optimization are displayed in grey.B. E-network. Scattergram illustrating the correlation of flux values calculated by means of a kinetic model [30] with flux values predicted by *MinMode composition*.

To check the feasibility of the *MinMode composition* for the E-network we compared the predicted fluxes with those calculated by means of a validated kinetic model [30]. Figure 6B reveals an almost perfect prediction for the larger fluxes and acceptable deviations for the small fluxes. Taken together, the quality of flux predictions based on the MinMode decomposition was equivalent with the quality achieved by global optimization [31].

Similarity analysis of *MinModes*

Vectors forming a basis in strict mathematical sense have to be orthogonal, i. e. independent from each other. This criterion does not hold for *MinMode*s. As demonstrated above with the *MinMode*s of the E-network, *MinMode*s belonging to different metabolic outputs, e. g. production of ATP and 2,3DPG, may be very similar. Thus it is practically impossible to conclude from observed changes of fluxes, which of the two target fluxes have changed. Therefore we represent those *MinMode*s exhibiting a strong similarity by a single *Principal MinMode* and we use the smaller set of such *Principal MinMode*s for the decomposition. To quantify the similarity of two *MinMode*s, Pearson's correlation coefficient turns out not to be a reliable measure because the components of flux vectors are not normally distributed (they contain many zero-fluxes and many tightly related flux values owing to the flux-

balance conditions). We decided to quantify the similarity between two arbitrary flux modes MM and MM' by a similarity index (si) defined as the (relative) number of components having the same sign in both modes:

$$\mathrm{si}\left(\mathrm{MM},\mathrm{MM}'\right)=\frac{1}{\mathrm{N_r}}\sum_{\mathrm{i=1}}^{\mathrm{N_r}}\rho\left(\mathrm{MM_i},\mathrm{MM_i'}\right),\quad \rho\left(\mathrm{MM_i},\mathrm{MM_i'}\right)=\begin{cases}1 \text{ if } \mathrm{sgn}(\mathrm{MM_i})=\mathrm{sgn}(\mathrm{MM_i'})\\ 0 \text{ else}\end{cases} \qquad (6)$$

In Equation 6, sgn(x) denotes the sign-function (sgn(x) = +1, -1 or 0 if $x > 0$, $x < 0$ or $x = 0$). The sum in Equation 6 runs over all components except those two referring to the target fluxes generating the two *MinMode*s MM and MM', respectively. The similarity indices form the (symmetric) *MinMode* similarity matrix of *MinMode*s. Figure 7 shows the *MinMode* similarity matrices for the two exemplary networks. Values larger than the arbitrarily chosen threshold value of 0.9 (i.e. 90% of the fluxes in the two

Methylobacterium

		1 acetylCoA	2 aKG	3 ATP	4 methyleneH4F	5 Ery4P	6 formylH4F	7 Glc6P	8 glycine	9 NADPH	10 OAA	11 PentoP	12 PEP	13 PHB	14 pyruvate	15 serine	16 Succ	17 TP
1	acetylCoA		0.73	0.79	0.80	0.64	0.76	0.61	0.67	0.81	0.84	0.61	0.72	0.85	0.72	0.67	0.76	0.68
2	aKG	0.73		0.55	0.57	0.83	0.56	0.80	0.85	0.63	0.87	0.80	0.91	0.75	0.91	0.85	0.55	0.87
3	ATP	0.79	0.55		0.93	0.51	0.87	0.47	0.52	0.85	0.63	0.47	0.59	0.72	0.59	0.52	0.97	0.55
4	methyleneH4F	0.80	0.57	0.93		0.51	0.93	0.47	0.52	0.81	0.64	0.47	0.59	0.71	0.59	0.52	0.93	0.55
5	Ery4P	0.64	0.83	0.51	0.51		0.47	0.96	0.85	0.60	0.77	0.96	0.89	0.71	0.89	0.85	0.51	0.93
6	formylH4F	0.76	0.56	0.87	0.93	0.47		0.44	0.49	0.79	0.61	0.44	0.55	0.67	0.55	0.49	0.87	0.51
7	Glc6P	0.61	0.80	0.47	0.47	0.96	0.44		0.84	0.59	0.75	1.00	0.85	0.67	0.85	0.84	0.47	0.89
8	glycine	0.67	0.85	0.52	0.52	0.85	0.49	0.84		0.64	0.80	0.84	0.91	0.72	0.91	1.00	0.52	0.89
9	NADPH	0.81	0.63	0.85	0.81	0.60	0.79	0.59	0.64		0.68	0.59	0.68	0.79	0.68	0.64	0.83	0.64
10	OAA	0.84	0.87	0.63	0.64	0.77	0.61	0.75	0.80	0.68		0.75	0.85	0.75	0.85	0.80	0.60	0.81
11	PentoP	0.61	0.80	0.47	0.47	0.96	0.44	1.00	0.84	0.59	0.75		0.85	0.67	0.85	0.84	0.47	0.89
12	PEP	0.72	0.91	0.59	0.59	0.89	0.55	0.85	0.91	0.68	0.85	0.85		0.79	1.00	0.91	0.59	0.96
13	PHB	0.85	0.75	0.72	0.71	0.71	0.67	0.67	0.72	0.79	0.75	0.67	0.79		0.79	0.72	0.69	0.75
14	pyruvate	0.72	0.91	0.59	0.59	0.89	0.55	0.85	0.91	0.68	0.85	0.85	1.00	0.79		0.91	0.59	0.96
15	serine	0.67	0.85	0.52	0.52	0.85	0.49	0.84	1.00	0.64	0.80	0.84	0.91	0.72	0.91		0.52	0.89
16	Succ	0.76	0.55	0.97	0.93	0.51	0.87	0.47	0.52	0.83	0.60	0.47	0.59	0.69	0.59	0.52		0.55
17	TP	0.68	0.87	0.55	0.55	0.93	0.51	0.89	0.89	0.64	0.81	0.89	0.96	0.75	0.96	0.89	0.55	

Erythrocyte

		1 ATPase	2 DPGm	3 GSHox	4 PRPPs
1	ATPase		0.97	0.50	0.77
2	DPGm	0.97		0.40	0.77
3	GSHox	0.50	0.40		0.37
4	PRPPs	0.77	0.77	0.37	

Figure 7. Similarity matrices for the *MinModes* of the B- and E-network.
Similarity between two MinModes was assessed by the similarity index (si) defined in Equation 5. si-values larger than 0.9 are indicated in grey.

MinModes under comparison are either both zero or point in the same direction) are marked in grey. We note that three pairs of *MinMode*s calculated for the B-network exhibit perfect similarity (si = 1). For example, the *MinMode*s associated with the production of the metabolites Glc6P and PentoseP do not differ either in the zero fluxes or in the directionality of the non-zero fluxes. Plotting the components of these two *MinMode*s against each other also reveals strong similarity (Fig. 8A). This is due to the fact that the pentose phosphates are formed in the pentose phosphate pathway which branches from glucose-6-phosphate, i.e. production of pentose phosphates necessarily involves the production of glucose-6-phosphate.

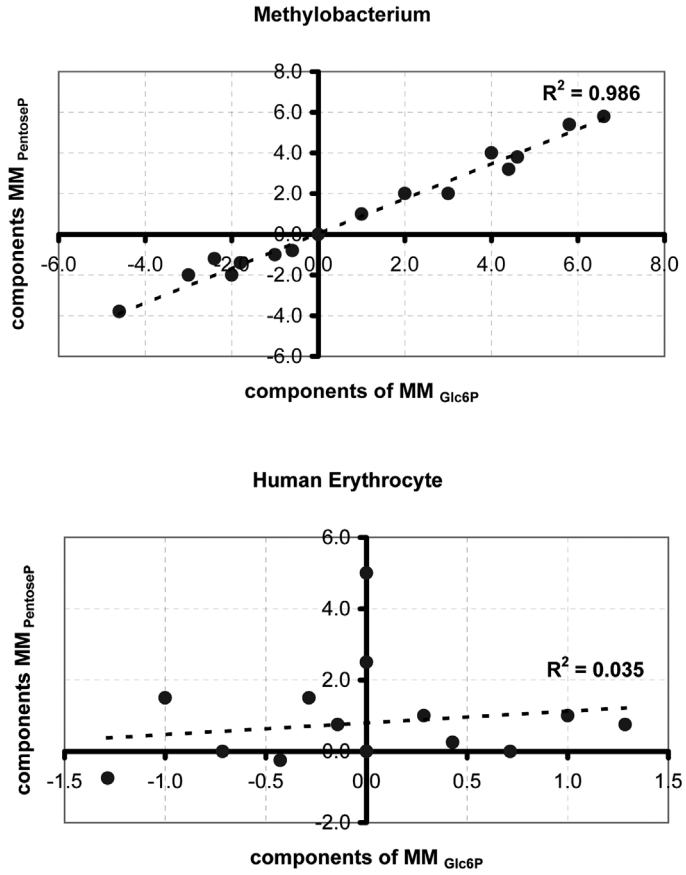


Figure 8. Scattergram illustrating the correlation between the *MinModes* for the formation of Glc6P and pentoseP.

The important point is, however, that with methanol as carbon source a large part of the network is used to produce glucose-6-phosphate. In contrast, in the E-network with glucose as substrate only two reactions are needed (Glct and HK) to form glucose-6-phosphate. As a consequence, the two *MinMode*s associated with the production of Glc6P and PentoseP are completely different (Fig. 8B). This shows that the similarity of two *MinMode*s associated with the production of a given pair of metabolites depends strongly upon the architecture of the network and the available extracellular substrates. From the two similarity matrices shown in Fig. 7 we can identify those target fluxes which are simultaneously affected by changes in the level of active enzymes. For example, in the case of ATP and 2,3DPG we will not be able to discriminate which cellular requirements have changed with respect to those two target functions by only inspecting the internal fluxes.

Based on the similarity matrix, *MinMode*s can be grouped into clusters encompassing all *MinMode*s with sufficiently high mutual similarity. Figure 9 shows the result of a cluster analysis of the *MinModes* for the B-network, performed by using the furthest neighbour method, i.e. measuring the overall similarity of all *MinMode*s assembled in a cluster by the smallest pair-wise similarity index.

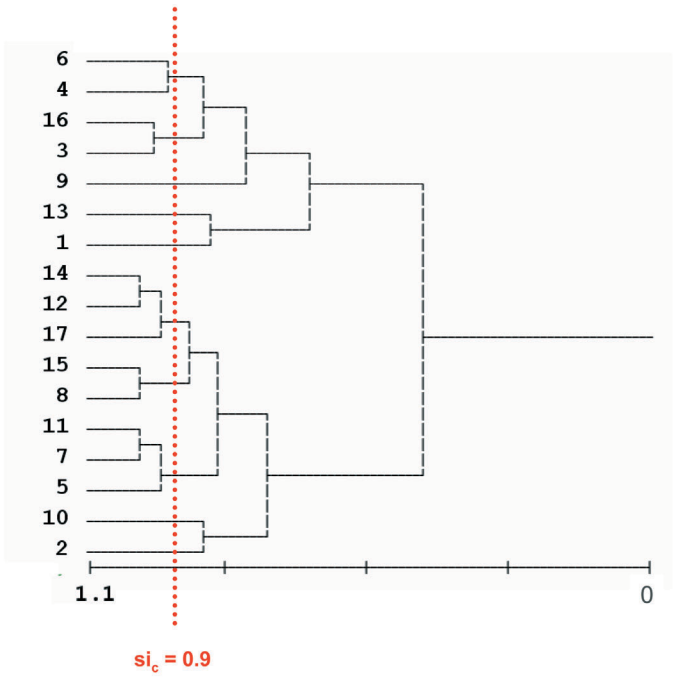


Figure 9. Dendogram illustrating the clustering of *MinModes* for the B- and E-network.
Cluster analysis was performed on the basis of the similarity matrices shown in Fig. 7 using the closest neighbour method, i.e. the smallest similarity index for all pairs of *MinModes* falling into one cluster is larger than the critical value indicated on the horizontal axis.

Extraction of *Principal MinMode*s

Using *MinModes* for the decomposition of flux distributions it seems feasible to represent those *MinModes* comprising a large degree of similarity by a single *Principal MinMode*. *Principal MinModes* exhibit a lesser degree of similarity and thus allow a more unambiguous decomposition. To this end, we have to define a cut-off value (si_c) for *MinMode* similarity. *MinModes* assembled in a cluster possessing a cluster similarity larger than this cut-off value are lumped together (as a linear combination of its elements) to a single *Principal MinModes* (PMMs). Note that *Principal MinModes* does not satisfy the MinMode definition. Further *Principal MinMode*s are given in terms of those *MinMode*s which

do not fall into clusters with sufficiently high cluster similarity. The *Principal MinMode*s obtained by this procedure for the two exemplary networks at a cut-off value of $si_c = 0.9$ are depicted in Fig. 10.

Principal MinModes - Methylobacterium

PMM $_1$	$MM_5 + MM_7 + MM_{11}$	Ery4P, Glc6P, PentoseP
PMM $_2$	$MM_{12} + MM_{14} + MM_{17}$	PEP, Pyruvate, TP
PMM $_3$	$MM_8 + MM_{15}$	glycine, serine
PMM $_4$	$MM_4 + MM_6$	methyleneH4F, formylH4F
PMM $_5$	$MM_3 + MM_{16}$	Succ, ATP
PMM $_6$	MM_1	acetylCoA
PMM $_7$	MM_2	aKG
PMM $_8$	MM_9	NADPH
PMM $_9$	MM_{10}	OAA
PMM $_{10}$	MM_{13}	PHB

Principal MinModes - Erythrocyte

PMM 1	$MM_1 + MM_2$	DPGM, ATPase
PMM 2	MM_3	GSHox
PMM 3	MM_4	PRPPs

Figure 10. Definition of Principal *MinModes.*
MinModes falling into clusters with minimal similarity of 0.9 (see Fig. 9) have been lumped into a single Principal MinMode.

There are 10 PMMs (instead of 17 MMs) for the B-network and 3 PMMs (instead of 4) for the erythrocyte network. Obviously, the number of *Principal MinMode*s depends on the choice of the cut-off value si_c for mutual *MinMode* similarity. At $si_c = 0.8$ we would get only 3 PMMs for the B-network.

Flux changes induced by changes of enzyme levels: simulated gene expression

To study the changes of stationary fluxes accompanying changes of enzyme levels we used our comprehensive kinetic model of the erythrocyte metabolism to calculate stationary states at various enzyme levels. Variations in the amount of an enzyme were accomplished by varying its maximal velocity. We considered the following two extreme cases of gene expression. *Random gene expression* was simulated by multiplying the actual v_{max} values of all enzymes with factors randomly chosen within the interval [0.1, 10.0]. *'On demand' gene expression* (see Fig. 11 for a detailed explanation) was simulated by first changing the load

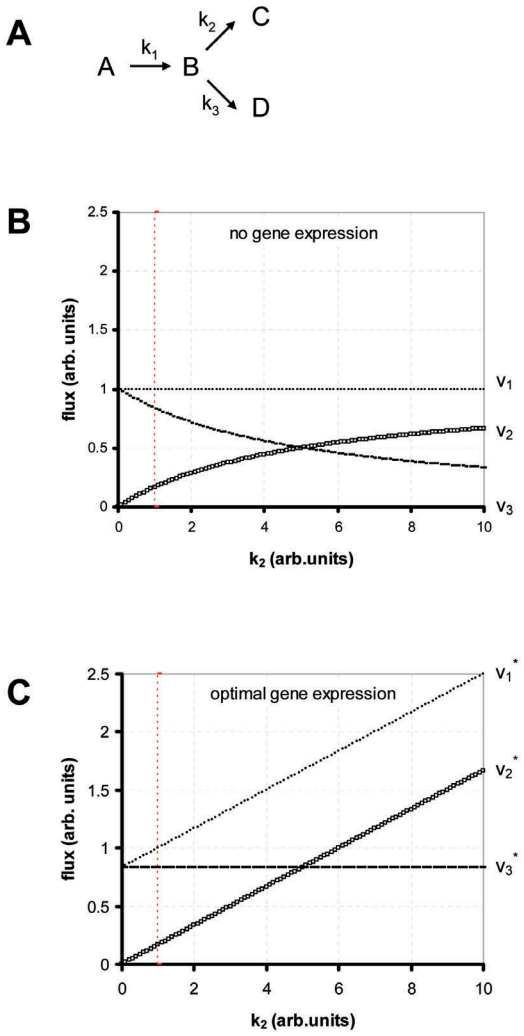


Figure 11. Illustration of the hypothesis on optimal gene expression.
The simplistic network in panel A is composed of three monomolecular reactions. The function of the network is to convert substrate A into two different end products C and D. Under non-saturating conditions the fluxes are given by $v_1 = k_1$ [A], $v_2 = k_2$ [B] and $v_3 = k_3$ [B] whereby the rate constants k_2 and k_3 represent the load parameters and the rate constant k_1 is proportional to the concentration of the enzyme catalysing the first reaction.. Metabolic steady states of the system are defined by the flux-balance condition $v_1 = v_2 + v_3$. Thus, at fixed concentration of the substrate [A]= 1, the output fluxes read $v_2 = \frac{k_1 k_2}{k_2 + k_3}$ and $v_3 = \frac{k_1 k_3}{k_2 + k_3}$.

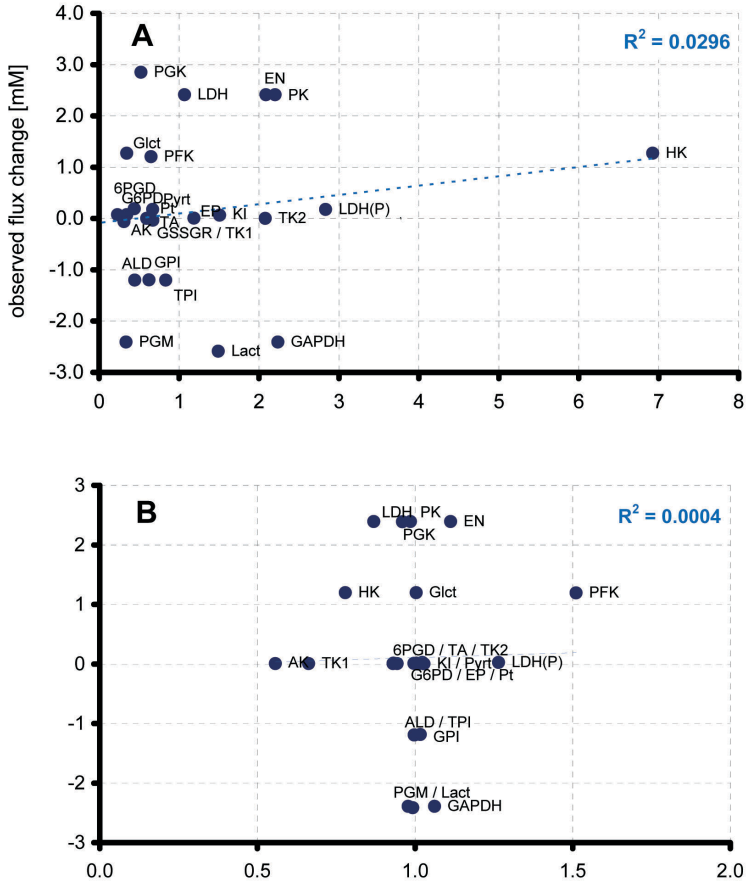
Panel B illustrates the dependence of the three stationary fluxes on the load parameter k_2 at fixed values $k_3 = 5$ of the second load parameter and $k_1 = 1$ of the rate constant for the first reaction. Increasing values of the load parameter k_2 result in a decrease of flux v_3 whereby the increase of flux v_2 is sub-linear, i.e. the control coefficient $C_2 = \frac{\partial \ln(v_2)}{\partial \ln(k_2)}$ of the output flux v_2 with respect to the load parameter k_2 is smaller than unity. This sub-optimal behaviour becomes successively pronounced with increasing values of k_2. Optimal gene expression is hypothesized both to accomplish a maximal response of the output flux v_2 towards changes of the load parameter k_2 such that the flux-control coefficient becomes unity, $C_2 \rightarrow 1$, and to prevent a change of the other (independent) output flux v_3. This can be achieved by variable expression of enzyme catalysing the reaction $A \rightarrow B$, i.e. adapting the rate constant k_1 to the load parameters according to $k_1 = \gamma (k_2 + k_3)$. The corresponding fluxes v_i^* $(i = 1,2,3)$ in the presence of optimal gene expression are shown in panel C whereby the proportionality constant was put $\gamma = {}^1/_6$ so that the load characteristics without and with gene expression match at $k_2 = 1$ (indicated by the dashed vertical line).

parameter k_i for target fluxes v_i by a given factor $\eta = k_i'/k_i$ where k_i' is the new value of the load parameter. This change of the load parameter implies a change of the target flux to the new value v_i' but, in general, this change is smaller than η, i.e. the flux-control coefficient $C_i = \frac{1}{\eta}\frac{v_i'}{v_i}$ of the chosen target flux with respect to the load parameter is smaller than unity.

New v_{max} values of all enzymes were than determined fulfilling two criteria: (i) the change of the chosen target flux was also η-fold, i.e. the flux-control coefficient of this target flux with respect to the load parameter was unity, $C_i = 1$, and (ii) the other target fluxes remained at their initial value. This simulated mode of gene expression assures high selectivity in the cellular response towards changes of the metabolic load.

For these two modes of simulated gene expression, changes in the steady-state fluxes were expressed as difference between new and initial flux values and these changes were plotted against the changes in the v_{max} values of the catalysing enzymes expressed as fold changes calculated by dividing the new v_{max} value by its initial value. The scattergrams in Fig. 12A-C and the associated measures of determination (R^2) reveal poor correlations not only for the random case but also for the two cases of 'on demand' gene expression optimizing the response of the metabolic system towards an increase in the load parameters k_{ATPase} for the energy consumption and k_{GSHox} for the consumption of GSH. This finding is in clear contrast to the well-known linear relationship between flux rate and enzyme activity holding for isolated reactions at fixed concentration of the reactants. In a reaction network, however, changes in the activity of a single enzyme remain not restricted to changes in the rate of the corresponding reaction but give rise to changes of all fluxes owing to the coupling of the reactions through shared reactants and allosteric effectors. This way a local perturbation of a single enzyme propagates through the whole network and the resulting

new steady-state flux distribution depends on the specific kinetic properties of all enzymes in the network. Hence, from the kinetic point of view, a simple correlation between changes of enzyme levels and changes of the associated fluxes indeed cannot be expected. In the following paragraph we propose a method to exploit information on changes of enzyme concentrations to arrive at better predictions of flux changes in the network.



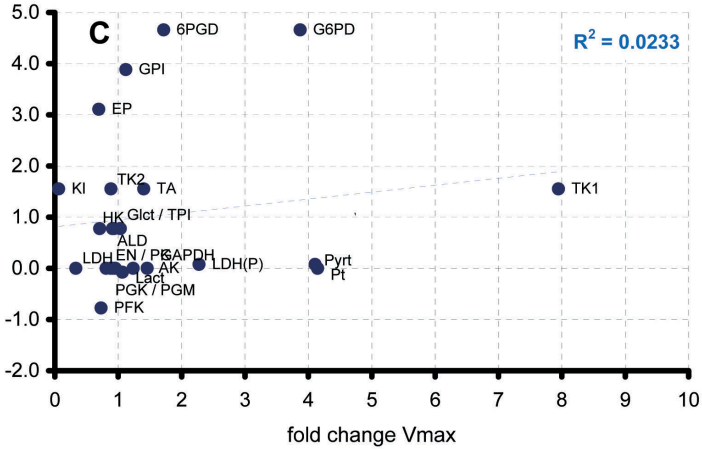


Figure 12. Scattergram illustrating the correlation between flux changes and changes of v_{max} values for the E-network. v_{max} values were changes either randomly (A) or applying an 'on demand' gene expression strategy (details given in the main text). 'Observed' flux changes were calculated by using the kinetic model [30] (B) Simulated 'on demand' gene expression accompanying an increase of the load parameter k_{ATPase} for ATP consumption by a factor of $\eta = 2$. (C) Simulated 'on demand' gene expression accompanying an increase of the load parameter k_{GSHox} for GSH consumption by a factor of $\eta = 100$.

Predicting flux changes from changes of enzyme levels by using the flux decomposition into *Principal MinModes*

Changes in the stationary fluxes are not independent – they are coupled through the balance conditions which hold even if there are changes in the amount of enzymes due to variable gene expression. For example, given that the postulated metabolic network of the erythrocyte is correct the fluxes through the glucose transporter (Glct) and the hexokinase (HK) have to be equal, and the same also holds for any flux changes through these two reactions. Looking at the data in Fig. 10 not in X → Y direction but in Y → X direction one would expect the data points for Glct and HK to coincide if there was a perfect correlation of flux changes with changes of enzyme activities. The scattergram in Fig. 12B ('on demand' gene expression at higher energetic load k_{ATPase}) shows that the v_{max} changes for these two proteins are different. Although there is no change in the activity of the glucose transporter and even a decrease (!) in the activity of the hexokinase (HK) the flux through both reactions has increased by a factor of about 1.2. This is a pure kinetic effect brought about by a lowered intracellular glucose concentration due to activation (higher expression) of the phosphofructokinase (PFK), one of the key regulatory glycolytic enzymes. This example illustrates that some of the flux changes in the network are due to kinetic effects induces by changes in the expression of those enzymes (as the PFK) exerting the dominant control over the desired changes in the metabolic output of the network. Regarding the problem of predicting flux changes from changes of enzyme levels we have to conclude that only some of the observed changes in enzyme activities are indicative for changes of the associated

fluxes. For the considered example, the increase in the activity of the PFK actually reflects an increase in the flux through this enzyme whereas the unaltered activity of glucose transporter (Glct) does not.

Let us consider the pool of correlated fluxes, i.e. which are related to each other by fixed ratios in any conceivable flux distributions due to flux balance conditions. Given that within this pool there exists at least one flux for which the change is not kinetically determined but predominantly due to a change of the enzyme level. If so, it would make sense to represent all fluxes belonging to this pool by a single representative flux and to correlate its change with the average observed changes in the activities of the associated enzymes. Such a constrained correlation analysis can be accomplished by approximating the unknown vector $\Delta\mathbf{v}$ of flux changes by a linear decomposition into *Principal MinModes* $\mathbf{PMM_i}$ (see Equation 5) the components of which automatically obey the flux-balance conditions:

$$\Delta\mathbf{v} = \sum_i \Delta\alpha_i \, \mathbf{PMM}_i \tag{7}$$

The coefficients $\Delta\alpha_i$ (i = 1,2,..., number of PMMs) of this decomposition are than determined by maximizing the correlation between $\Delta\mathbf{v}$ and the observed changes of enzyme activities $\Delta\mathbf{A} = \left(E_1^*/E_1, E_2^*/E_2, ..., E_{ne}^*/E_{ne} \right)$, i.e.

$$\underset{(\alpha_i \geq 0)}{\mathrm{MAX}} \, \mathrm{Corr}\left(\Delta\mathbf{v}, \Delta\mathbf{A}\right) \tag{8}$$

Here, fold-changes in the v_{max} values are taken as measure for changes in the enzyme amounts. Solving the optimization problem (8) we obtain a prediction of the flux changes in the network. The value of the coefficient $\Delta\alpha_i$ indicates the changes of the target fluxes associated with the *Principal MinMode* $\mathrm{PMM_i}$. Hence, the set of coefficients $\Delta\alpha_i$ allow direct inferences to be made on those changes in the metabolic output the network which have provoked the observed changes in the enzyme levels.

This procedure was applied to the three "expression patterns" illustrated in Fig. 13A-C. The similarity index used to extract *Principal MinModes* for the erythrocyte network was put to 0.9 resulting in the following 3 *Principal MinModes* given in Fig. 10.

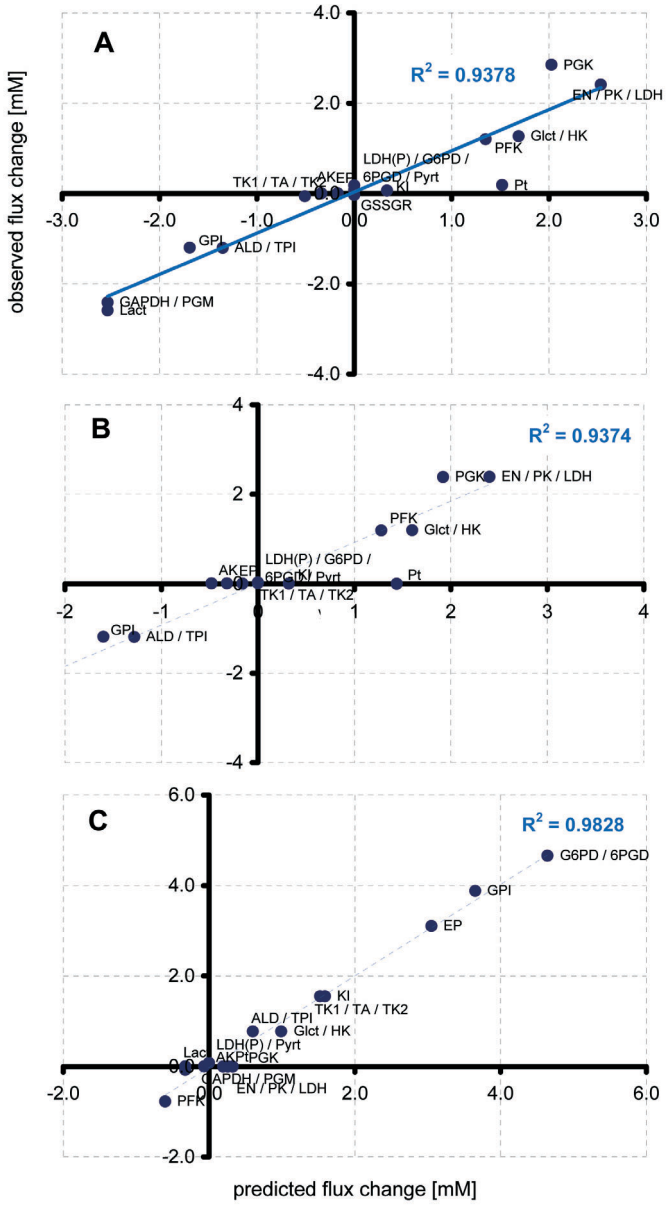


Figure 13. Scattergram illustrating the correlation between 'observed' flux changes and flux changes obtained by maximizing the correlation between changes of v_{max} values and the *MinMode* decomposition (6) based on the three Principal MinModes for the E-network given in Fig. 10. The three simulated cases A-C are explained in Fig. 11. (A) Random gene expression. Estimated values for the decomposition coefficients: $\Delta\alpha_1 = 0.08$, $\Delta\alpha_2 = 0.0$, $\Delta\alpha_3 = 0.06$. (B) 'On demand' gene expression at increased load parameter k_{ATPase}. $\Delta\alpha_1 = 0.07$, $\Delta\alpha_2 = 0.0$, $\Delta\alpha_3 = 0.0$, (C) 'On demand' gene expression at increased load parameter k_{GSHox}. $\Delta\alpha_1 = 0.0$, $\Delta\alpha_2 = 0.72$, $\Delta\alpha_3 = 0.0$.

Figure 13 shows the "observed" (= simulated) flux changes plotted against predicted flux changes obtained as solution of the optimization problem (7). Even for the simulated case of *random gene expression* (Fig. 13A) the concordance between predicted and observed flux changes is surprisingly good. The values of the coefficients α_1, α_2 and α_3 solving the optimization problem (7) for the three cases of simulated gene expression are given in the legend of Fig. 13. According to these values the changes in the maximal enzyme activities for case B and C were clearly identified as resulting from an increase in either the energetic or oxidative load.

Putting these findings together we may conclude that the proposed strategy of:

- decomposing the flux distribution into minimal flux modes
- lumping these *MinModes* together to redundant-free *Principal MinModes*,
- expressing the unknown flux changes as linear combination of *Principal MinModes*
- and determining the unknown coefficients of this linear combination by maximizing the correlation with observed changes of enzyme level (=v_{max} values)

provides a powerful means of predicting flux changes in the metabolic network as well as those changes in the output of the metabolic network having caused these flux changes.

Discussion

As demonstrated for the exemplary network considered herein, the projection of the global flux-minimized steady-state solution onto the convex basis of elementary modes resulted in a manageable set of elementary flux modes with non-vanishing coefficients. However, these "basic" elementary modes were difficult to assign to a specific metabolic output. Besides this, there are some further shortcomings rendering the convex basis of elementary modes unsuitable for the decomposition of flux distributions into functionally interpretable modes. First, the determination of the convex basis is not unique [8]. Thus, choosing another convex basis of elementary modes, their physiological interpretation can be vastly different [19]. Second, the coefficients for the non-negative linear decomposition are also not unique (in our computational protocol we have chosen the coefficients with minimal 1-norm) and thus their absolute values do not allow conclusions to be drawn with respect to the relative importance of the various elementary modes. Third, the (subjective) decision on reversible and irreversible reactions deeply affects the set of elementary modes and thus the convex basis. In the B-network, 43 chemical reactions are considered irreversible. One might argue that potentially every chemical reaction can be reversed by increasing the concentration of the products and/or reduction of the concentration of the substrates. Setting all reactions as reversible renders many elementary modes *a priori* physiologically irrelevant. Elementary modes appearing implausible to the biochemist are, for example,

inner cycles that operate without exchange of matter with the environment or modes encompassing reactions with flux directions that are not in concordance with their (large) change of standard free energy.

In this work we introduced the concept of minimal flux modes (*MinMode*s), defined as flux minimized steady-state flux distributions enabling the production of a single metabolite. The production of a certain subset of metabolites defines the functionally relevant output of the network. The introduction of *MinMode*s means to decompose this output into separate contributions. The flux cone spanned by the set of *MinMode*s is a true subset of the mode space. For the Methylobacterium model the convex dimension of the *MinMode* space is 17, whereas the convex basis of elementary modes comprises 7033 vectors. The full vector space spanned by linear combinations of the convex basis with real number coefficients has the dimension 29. Hence the set of *MinMode*s is not complete, i. e. an arbitrary steady flux distribution cannot be exactly decomposed into a linear combination of *MinMode*s. On the other hand, *MinMode*s are attractive because of their clear assignment to specific output reactions. Because the *MinMode*s are biochemically feasible, the reduced flux cone spanned by the *MinMode*s is also feasible as a whole, in contrast to the complete flux cone. Elementary modes and minimal fluxes as introduced in this paper represent two different methodological concepts that ultimately have the same goal: To decompose the fluxes in the metabolic network into a set of more simple but physiologically relevant flux modes. Elementary mode analysis starts with a complete set of all possible and not further decomposable routes ("top down" approach). The resulting huge set of such elementary modes has to be reduced to a physiologically relevant and numerically tractable set by imposing additional constraints [17 – 19]. In contrast, minimal mode analysis starts with a very small set of modes each of them connected with one physiologically relevant output of the network ("bottom up" approach). Thus, the number of *MinMode*s cannot be larger than the number of metabolites occurring in the network. However, *MinMode*s allow only an approximate description of the true flux distribution as they do not form a complete basis in strict mathematical sense. Nevertheless, the striking advantage of the *MinMode* concept is its applicability to very large whole-cell networks (> 500 reactions) where the effective handling of elementary modes can be clearly ruled out for computational reasons.

The set of *MinMode*s can be investigated by similar methods as used for analyses based on elementary modes. For example, the frequency with which a reaction has a non-zero flux within the set of *MinMode*s can be taken to rank the functional relevance of reactions in the network. Such analyses may give insight into the evolution of metabolic networks: Reactions with many non-zero fluxes in the *MinMode*s may represent the ancient part of the network responsible for some basal functions. This part of the network was then successively complemented by reactions and pathways connected to more specific functions and thus being less represented in the *MinMode*s. It has to be noted that disabling reactions with a high number of non-zero fluxes in the set of *MinMode*s does not necessarily lead to lethality because besides the *MinMode*s (being special flux distributions) alternative routes may exist. For example, formate dehydrogenase (reaction 6) has a high participation frequency but was shown to be not essential for growth on methanol [35]. On the other hand, reactions which upon exclusion from the network (by setting the flux to zero) do not

allow the determination of a *MinMode* for each output metabolite are to be considered essential. Omitting non-essential reactions merely leads to a different pattern of *MinMode*s. If a mutant lacking a reaction predicted to be not essential is not able to survive or shows reduced growing capabilities, an important physiological function of this reaction has failed to be noticed. Therefore, relating the observed phenotype of knock-out mutants with network based classifications of essential and non-essential reactions may provide a valuable heuristics to unravel the physiological importance of metabolites.

Interestingly there are fifteen reactions of the B-network that do not contribute to any flux mode. Assuming the synthesis of biomass precursors to be the sole cellular function of the network, these 15 reactions should be abdicable for cellular growth using methanol as the only carbon source. Three of these reactions contribute to the degradation of the storage metabolite PHB. The enzymes PHB depolymerase, β-hydroxybutyrate dehydrogenase and acetoacetate-succinyl-CoA transferase, able to catalyse the conversion of PHB into acetoac-CoA are not required for cellular growth. However, this holds true only for environmental conditions where enough substrate is available. In a starvation phase, the apparently dispensable PHB degradation becomes important. Another example is the aldolase reaction converting 2 triose phosphates into Fru-1,6-BP (reaction 34). The flux through this reaction in every *MinMode* is zero. Thus, synthesis of Fru-1,6-BP as an intermediate seems to be not necessary for the fulfilment of the assumed metabolic tasks under the given environmental conditions. There are two possible explanations for this redundancy. Either, this reaction is required for enhanced stability and robustness of the network and is abdicable under conditions where the enzymes catalysing alternative routes are expressed or its metabolic task is not required under the given conditions. Or, Fru-1,6-BP is required for other biochemical processes not considered yet, either as a reactant in reactions not included in the network or as a regulatory metabolite. The latter explanation is very likely because knowledge of the full spectrum of metabolites necessary to ensure all cellular activities will be incomplete. For example, it has become known quite recently that presence of Fr-1,6-BP is an absolute requirement for lactate dehydrogenase activity in *Lactobacillus casei* [36] and a similar regulatory function of this metabolite cannot be excluded in *Methylobacterium extorquens*. Therefore, to take into account a potential role of all metabolites in cellular functionality a complete set of *MinMode*s should be constructed enabling the production of all metabolites occurring in the network. If this more systematic approach is applied, only 6 of the 15 previously unused reactions remain. In case the model is correct this redundancy should be explained solely by network robustness to mutations and changed environmental conditions. Mutants not able to catalyse these reactions should grow normally. The example of α-KG dehydrogenase (reaction 34), one of the 6 abdicable reactions, supports this hypothesis. The enzyme catalyses the conversion of α-KG into Succ-CoA as part of the citric acid cycle. For growth on C1 resources this cycle is partly repressed and accomplishes an assimilatory role [24, 37]. It could be shown that a lack of this enzyme does not influence the growth behaviour of *Methylobacterium extorquens* (while growing on methanol only) [38]. Thus, the classification of reactions into those which are essential and non-essential has to be considered with precaution because such a classification depends strongly upon the specific external conditions as well as on the knowledge of the physiological functions that metabolites or reactions may have.

One may wonder whether the proposed computational approach to construct optimal flux distributions for each output variable of the network is more or less feasible than the calculation of an optimized flux distribution meeting all target flux simultaneously. The good concordance between experimentally determined fluxes and calculations based on *MinMode compositions* (see Fig. 6) may suggest that optimization of metabolic networks with respect to single target reactions has been an important goal of natural evolution. Considering that the relative importance of target fluxes may vary depending on the specific external conditions of the cell (at the extreme one target reaction as, for example, the production of glutathione in the presence of oxidative stress, might transiently overshoot all others) such a strategy appears to be not implausible because it allows independent regulation of different metabolic outputs. Moreover, adding new reactions (and thus functionality) to an already existing network during the course of natural evolution should not comprise already achieved optimality. Of course, it is too early to make a sound judgment, so that further applications of the proposed methods to other, more complex networks are needed. Based on the same assumption of a minimized total sum of fluxes, the flux distributions obtained by global and single target optimization show of course only little differences. Thus the new approach results in a flux distribution that is just as good as the one obtained by the previous approach. The Minimal Flux Modes 11 few differences can be illustrated by a simple network (Fig. 14A), consisting of 5 reactions v1,v2,v3,v4,v5 where v4 and v5 are considered as target reactions. Two minimal flux modes can be calculated for this network. For the *MinMode* that produces metabolite A by realizing the flux v4 = 1, we obtain the *MinMode* ~v =(10010), whereas for the realization of the flux v5 = 1, the *MinMode* is ~v =(01001). The sum of fluxes in each *MinMode* is 2. The combination of *MinMode*A and *MinMode*B results in the flux distribution ~v =(11011) with the sum of fluxes being 4. The basic flux minimization approach does not presume costs for enzyme synthesis or different activity levels for an enzyme. Therefore, demanding non-zero fluxes through both target reactions v4 and v5 gives the same flux distribution. Thus, single optimization and global optimization may lead to equal solutions as long as no currency metabolites (as ATP, NADH etc.) are involved. In the case, where v4 is an ATP consuming and v3 an ATP producing reaction (Fig. 14B), additional fluxes for ATP balancing become necessary (v6,v7). Here, the sum of fluxes for the single optimized solution would be 4 for *MinMode*A.

In combination with *MinMode*B the sum of fluxes is now 6. Global optimization yields a suboptimal path for the synthesis of B that contains a non-zero flux for the reaction v3, resulting in the vector ~v =(2011100) with a sum of fluxes equal to 5. Therefore, for a global optimization a suboptimal (carbon) route for one metabolite can be chosen if it is of advantage for the synthesis of another target metabolite.

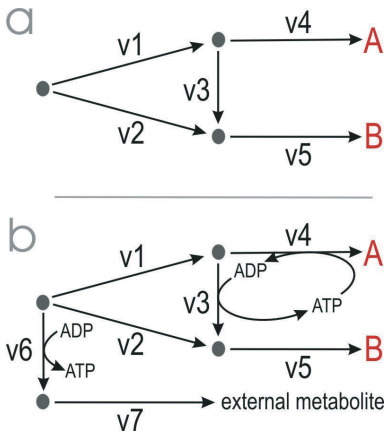


Figure 14. A simplistic network illustrating network effects which may cause differences between fluxes calculated by global minimization and MinMode decomposition.

The basic idea behind the concept of *MinMode*s presented in this paper is to interpret the flux distribution in a metabolic network as a superposition of various flux modes each related to one of the many functional requirements that the cells has to fulfil simultaneously but with different relative intensities. Hence, any metabolic status can be represented in terms of the coefficients entering the linear combination of *MinMode*s to the overall flux distribution. The use of those coefficients simplifies the flux-balance approach considerably and makes it possible to relate observed changes in metabolic fluxes directly to changes in the functionality of the cell.

As demonstrated for the *MinMode*s of the two exemplary networks considered, the *MinMode*s may exhibit a remarkable degree of similarity. Generally, *MinMode*s associated with target reactions located in close vicinity, i.e. belonging to the same pathway, should give rise to similar *MinMode*s. Thus, to employ *MinMode*s as a sort of 'basic vectors' it seems feasible to lump together similar *MinMode*s into a single *Principal MinMode*s. On one hand this leads to a further reduction of the set of relevant *MinMode*s, on the other it also reduces the clear-cut physiological interpretation of these *Principal MinMode*s as they are not associated with only one target flux but a certain group of target fluxes. In this article the definition of such *Principal MinMode*s was accomplished by clustering the *MinMode*s on the basis of a similarity index that counts the number of pair-wise fluxes pointing in the same direction or being zero. As with any statistical procedure, it is finally left to the user to define the minimal degree of similarity that has to be present among all *MinMode*s lumped together to a single *Principal MinMode*s.

In the last part of this article we have used the decomposition of flux changes into *Principal MinMode*s to predict changes in metabolic flux rates from observed changes of enzyme levels. It has to be noted that these results are based on simulated 'gene expression' experiments where we changed the maximal activities of erythrocyte enzymes and calcu-

lated the associated flux changes by means of a comprehensive kinetic model available for the E-network. These simulations required some *a priori* assumptions to be made about the regulatory principles underlying variable gene expression. These principles are still widely unknown. However, there is increasing theoretical and experimental evidence [39, 40] that temporal gene expression is an important means of cells to adapt their protein synthesizing capacity to changing external conditions such that the required metabolic output is achieved with high efficiency. This plausible strategy was the rational behind the simulations of 'on demand' gene expression which assures a high response of the network to changes in the demand of specific target reactions whereas the fluxes through other target reaction are kept at constant values. As expected, for the both extreme cases – random and 'on demand' gene expression – there was no significant correlation between changes in enzyme activities and changes in flux rates through the corresponding reactions. This theoretical finding, questions the naive interpretation of changes in gene expression profiles as to directly reflect changes in the activity of the underlying pathways.

Using the *MinMode* decomposition of the unknown flux changes as a side constraint and determining the coefficients of this decomposition to provide a maximal correlation with observed changes of enzyme activities, we obtained a significantly better prediction of flux changes. A further benefit of this strategy is that the values of the coefficients directly indicate the changes in the target fluxes that have elicited the changes in enzyme activities. This way it should be possible to make inferences on the functional strategy of cells just employing information of changes in enzyme levels.

#	enzyme	reaction	k_i	DPGM	ATPase	GSHox	PRPPS
1	Glct	Gluc(out) → Gluc	1	1	1	1	1
2	HK	Gluc + ATP → Glc6P + ADP	3900	1	1	1	1
3	GPI	Fru6P → Glc6P	2.55	-1	-1	1	-1
4	PFK	Fru6P + ATP → Fru1,6P + ADP	100000	1	1	0	1
5	ALD	DHAP + GraP → Fru1,6P	8.77	-1	-1	0	-1
6	TPI	GraP → DHAP	24.6	-1	-1	0	-1
7	GAPDH	1,3PG + NADH → GraP + Pi + NAD	5210	-1	-1	-1	-1
8	PGK	1,3PG + ADP → 3PG + ATP	1455	0	1	1	1
9	DPGM	1,3PG → 2,3PG	100000	1	0	0	0
10	DPGase	2,3PG → 3PG + Pi	100000	1	0	0	0
11	PGM	2PG → 3PG	6.9	-1	-1	-1	-1
12	EN	2PG → PEP	1.7	1	1	1	1
13	PK	PEP + ADP → Pyr + ATP	13790	1	1	1	1
14	LDH	Pyr + NADH → Lac + NAD	9090	1	1	1	1
15	LDH(P)	Pyr + NADPH → Lac + NADP	1420	0	0	0	0
16	ATPase	ATP → ADP + Pi	100000	0	1	0	0
17	AK	2 ADP –¿ ATP + AMP	0.64	0	0	-1	-1
18	G6PD	G6P + NADP → 6PG + NADPH	2000	0	0	1	0
19	6PGD	6PG + NADP → Ru5P + CO2 + NADPH	141.7	0	0	1	0
20	GSSGR	GSSG + NADPH → 2 GSH + NADP	3417.8	0	0	1	0
21	GSHox	GSH → GSSG	100000	0	0	1	0
22	EP	Ru5P → X5P	2.7	0	0	1	-1
23	KI	Ru5P → R5P	3	0	0	1	1
24	TK1	X5P + R5P → GraP + S7P	1.05	0	0	1	-1
25	TA	S7P + GraP → E4P + Fru6P	1.05	0	0	1	-1
26	PRPPS	R5P + ATP → AMP + PrPP	100000	0	0	1	1
27	TK2	X5P + E4P → GraP + Fru6P	1.2	0	0	1	-1
28	Pt	Pi(out) → Pi	1	0	0	1	1
29	Lact	Lac(out) → Lac	1	-1	-1	-1	-1
30	Pyrt	Pyr(out) → Pyr	1	0	0	0	0

Table 2. Model scheme and the corresponding minimal flux modes (MinModes) of the E-network.

References

[1] Clarke, B.L. (1988) Stoichiometric network analysis. *Cell Biophys.* **12:**237–253.

[2] Forster, J., Famili, I., Fu, P., Palsson, B.O., Nielsen, J. (2003) Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome Res.* **13:**244–253.

[3] Price, N.D., Papin, J.A., Schilling, C.H., Palsson, B.O. (2003) Genome-scale microbial in silico models: the constraints-based approach. *Trends Biotechnol.* **21:**162-169.

[4] Reed, J.L., Vo, T.D., Schilling, C.H., Palsson, B.O. (2003) An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biol.* **4:**R54.

[5] Schilling, C.H., Covert, M.W., Famili, I., Church, G.M., Edwards, J.S., Palsson, B.O. (2002) Genome-scale metabolic model of Helicobacter pylori 26695. *J. Bacteriol.* **184:**4582–4593.

[6] Schilling, C.H., Letscher, D., Palsson, B.O. (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theoret. Biol.* **203:**229**-**248.

[7] Schuster, S., Hilgetag, C., Woods, J.H., Fell, D.A. (2002) Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol.* **45:**153**-**181.

[8] Wagner, C., Urbanczik, R. (2005) The geometry of the flux cone of a metabolic network. *Biophys J.* **89:**3837–3845.

[9] Papin, J.A., Price, N.D., Wiback, S.J., Fell, D.A., Palsson, B.O. (2003) Metabolic pathways in the post-genome era. *Trends Biochem. Sci.* **28:**250–258.

[10] Papin, J.A., Price, N.D., Edwards, J.S., B, B.Ø. P. (2002) The genome-scale metabolic extreme pathway structure in Haemophilus influenzae shows significant network redundancy. *J. Theoret. Biol.* **215:**67**-**82.

[11] Price, N.D., Papin, J.A., Palsson, B.Ø. (2002) Determination of redundancy and systems properties of the metabolic network of Helicobacter pylori using genome-scale extreme pathway analysis. *Genome Res.* **12:**760-769.

[12] Beard, D.A., Babson, E., Curtis, E.. Qian, H. (2004) Thermodynamic constraints for biochemical networks. *J. Theoret. Biol.* **228:**327–333.

[13] Beard, D.A., Liang, S.-d., Qian, H. (2002) Energy balance for analysis of complex metabolic networks. *Biophys. J.* **83:**79-86.

[14] Klamt, S., Gilles, E.D. (2004) Minimal cut sets in biochemical reaction networks. *Bioinformatics* **20:**226-234.

[15] Gagneur, J., Klamt, S. (2004) Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics* **5:**175.

[16] Klamt, S., Stelling, J. (2002) Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.* **29:**233-236.

[17] Covert, M.W., Palsson, B.O. (2003) Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J. Theoret. Biol.* **221:**309-325.

[18] Schwartz, J.-M., Kanehisa, M. (2005) A quadratic programming approach for decomposing steady-state metabolic flux distributions onto elementary modes. *Bioinformatics* **21:**Suppl 2, ii204-ii205.

[19] Schwartz, J.-M., Kanehisa, M. (2006) Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis, *BMC Bioinformatics* **7:**186.

[20] Cornish-Bowden, A., Cárdenas, M.L. (2002) Metabolic balance sheets, *Nature* **420:**129–130.

[21] Holzhütter, H.-G. (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur. J. Biochem.* **271:**2905–2922.

[22] Holzhütter, H.-G. (2006) The generalized flux-minimization method and its application to metabolic networks affected by enzyme deficiencies, *Biosystems* **83:**98–107.

[23] Holzhütter, S., Holzhütter, H.G. (2004) Computational design of reduced metabolic networks. *Chembiochem* **5:**1401–1422.

[24] Dien, S.J. V., Lidstrom, M.E. (2002) Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph Methylobacterium extorquens AM1, with application to reconstruction of C(3) and C(4) metabolism. *Biotechnol. Bioeng.* **78:**296–312.

[25] Dien, S.J. V., Strovas, T., Lidstrom, M.E. (2003) Quantification of central metabolic fluxes in the facultative methylotroph methylobacterium extorquens AM1 using 13C-label tracing and mass spectrometry. *Biotechnol. Bioeng.* **84:**45–55.

[26] Hoffmann, S., Hoppe, A., Holzhütter, H.G. (2006). Composition of metabolic flux distributions by functionally interpretable minimal flux modes (MinModes). Paper presented at the *Genome Informatics*.

[27] Dien, S.J. V. personal communication.

[28] Vorholt, J. personal communication.

[29] Heinrich, R., Holzhütter, H.G., Schuster, S. (1987) A theoretical approach to the evolution and structural design of enzymatic networks: linear enzymatic chains, branched pathways and glycolysis of erythrocytes. *Bull. Math. Biol.* **49:**539–595.

[30] Schuster, R., Holzhütter, H.G. (1995) Use of mathematical models for predicting the metabolic effect of large-scale enzyme activity alterations. Application to enzyme deficiencies of red blood cells. *Eur. J. Biochem.* **229:**403–418.

[31] Holzhütter, S., Holzhütter, H.-G. (2004) Computational design of reduced metabolic networks. *Chembiochem* **5:**1401-1422.

[32] CPLEX. http://www.ilog.com/products/cplex/.

[33] Klamt, S., Stelling, J., Ginkel, M., Gilles, E.D. (2003) FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics* **19:**261–269.

[34] Pfeiffer, T., Sánchez-Valdenebro, I., Nuño, J.C., Montero, F., Schuster, S. (1999) METATOOL: for studying metabolic networks. *Bioinformatics* **15:**251–257.

[35] Chistoserdova, L., Laukel, M., Portais, J.-C., Vorholt, J.A., Lidstrom, M.E. (2004) Multiple formate dehydrogenase enzymes in the facultative methylotroph Methylobacterium extorquens AM1 are dispensable for growth on methanol. *J. Bacteriol.* **186:**22–28.

[36] Arai, K., Hishida, A., Ishiyama, M., Kamata, T., Uchikoba, H., Fushinobu, S., Matsuzawa, H., Taguchi, H. (2002) An absolute requirement of fructose 1,6-bisphosphate for the Lactobacillus casei L-lactate dehydrogenase activity induced by a single amino acid substitution. *Protein Eng.* **15:**35–41.

[37] Dien, S.J. V., Marx, C.J., O'Brien, B.N., Lidstrom, M.E. (2003) Genetic characterization of the carotenoid biosynthetic pathway in Methylobacterium extorquens AM1 and isolation of a colorless mutant. *Appl. Environ. Microbiol.* **69:**7563–7566.

[38] Chistoserdova, L., Chen, S.-W., Lapidus, A., Lidstrom, M.E. (2003) Methylotrophy in Methylobacterium extorquens AM1 from a genomic point of view. *J. Bacteriol.* **185:**2980–2987.

[39] Klipp, E., Heinrich, R., Holzhutter, H.G. (2002) Prediction of temporal gene expression. Metabolic opimization by re-distribution of enzyme activities. *Eur. J. Biochem.* **269:**5406–5413.

[40] Zaslaver, A., Mayo, A.E., Rosenberg, R., Bashkin, P., Sberro, H., Tsalyuk, M., Surette, M.G., Alon, U. (2004) Just-in-time transcription program in metabolic pathways. *Nature Genet.* **36:**486- 491.