Analysis of structural robustness of metabolic networks

T. Wilhelm, J. Behre and S. Schuster

Abstract: We study the structural robustness of metabolic networks on the basis of the concept of elementary flux modes. It is shown that the number of elementary modes itself is not an appropriate measure of structural robustness. Instead, we introduce three new robustness measures. These are based on the relative number of elementary modes remaining after the knockout of enzymes. We discuss the relevance of these measures with the help of simple examples, as well as with larger, realistic metabolic networks. Thereby we demonstrate quantitatively that the metabolism of *Escherichia coli*, which must be able to adapt to varying conditions, is more robust than the metabolism of the human erythrocyte, which lives under much more homeostatic conditions.

1 Introduction

A striking feature of living organisms is their homeostasis; they are, within some range, robust to external (e.g. temperature, food supply) and internal perturbations (e.g. spontaneous mutations). For example, many knockout mutants of micro-organisms are still able to grow, some with almost the same growth rate as the wild type. This has been shown, for example, by a systematic study on single knockout mutants of virtually all genes in baker's yeast [1, 2]. In many cells, there are parallel and thus redundant metabolic pathways. For example, the pentose phosphate pathway circumvents the upper part of glycolysis. Phosphoglycerate kinase in human erythrocytes can be bypassed via the Rapoport-Luebering shunt [3]. Both bypasses imply, however, a loss in ATP production. Often, redundancy in metabolism cannot be seen as easily as in these examples. To understand robustness in complex systems such as metabolic networks, theoretical tools are needed [4–8].

Studies on robustness have manifold applications. In biotechnology, specific enzymes can be knocked out to suppress futile cycles. For example, Rohwer and Botha [9] detected five futile cycles in the sucrose-accumulating sugar cane tissue. Another goal pursued in biotechnology is to suppress the synthesis of undesired products, e.g. caffeine in tea leaves [10]. It is then of interest to know which products can still be synthesised by the mutants [9, 11]. Studies on network redundancy are relevant also in medicine: How robust is human metabolism against enzyme deficiencies [12, 13]? In drug target identification, it is most suitable to find an enzyme that is non-redundant in the pathogenic micro-organism, while redundant in the human host, so that the latter is not perturbed too much. This has been analysed

for the causative agent of the African sleeping disease, *Trypanosoma brucei* [14, 15].

Robustness is generally defined as the insensitivity of a system to changes in parameters [7]. These can be parameters determined by the surroundings of the organism or by internal fluctuations. Different types of robustness can be distinguished. For example, the negative feedback loops present in many biochemical pathways such as the synthesis routes of numerous amino acids make the production rate robust to changes in the demand of the product. This can be viewed as dynamic robustness. Here, however, we analyse structural robustness. We deal with the question as to whether a cell can tolerate the elimination of some enzymes by mutations. Structural robustness is necessarily linked with redundancy because the network has no other possibility in responding to a knockout than to use alternative routes.

Current methods for the modelling of metabolism have various strengths and shortcomings. Specifically, dynamic simulation of metabolic and regulatory networks [4, 14, 16, 17] meets difficulties as the necessary mechanistic detail and kinetic parameters are rarely available. In contrast, methods for analysing the topological structure of metabolic networks such as metabolic pathway analysis [11, 18–21] only require knowledge of the stoichiometric coefficients and the directionality of reactions, which is available in many cases from the literature or on-line databases.

A central concept in metabolic pathway analysis is that of elementary flux modes [18, 19]. An elementary mode is a minimal set of enzymes that can operate at steady state, such that all irreversible reactions involved are used in the appropriate direction. The enzymes are weighted by the relative flux they carry. Any flux distribution in the living cell is a superposition of elementary modes. This concept, which takes into account relevant stoichiometric and thermodynamic constraints, has been applied to a number of biochemical networks of increasing complexity (e.g. [6, 9, 22, 23]). Elementary mode analysis appears to be well-suited to characterise network redundancy because each elementary mode is non-redundant. By examining which of these modes form the same products from the same substrates, one can detect parallel routes. Using metabolic pathway analysis software, such as METATOOL [24] or FluxAnalyzer [25], this can be performed in an automated way. However, it should be mentioned that the computation

114

Syst. Biol., Vol. 1, No. 1, June 2004

of elementary modes in larger networks meets the problem of combinatorial explosion [26]. Therefore, additional intellectual effort is necessary to formulate a feasible problem for a sub-network of interest.

The total number of elementary modes for given conditions has been used as a quantitative measure of network flexibility and as an estimate of fault-tolerance [6, 27]. Redundancy has also been analysed by Palsson *et al.* [20, 21]. They used the concept of extreme pathways, which is closely related to that of elementary modes. (For a comparison of the two concepts, see [28] and [29].) The simulations by Papin *et al.* [20] for *Haemophilus influenzae* showed that there was an average of 37 extreme pathways corresponding to the same input/output regime, when the network was used to produce a single amino acid. Price *et al.* [21] found by model calculations that the synthesis of amino acids and ribonucleotides in *Helicobacter pylori* is less redundant than in *H. influenzae*. In these papers, the number of extreme pathways with the same overall stoichiometry (in terms of initial substrates and final products) is used as a measure of redundancy. A slightly different approach was suggested by Oancea [30] and Çakır *et al.* [31]. They assessed the importance of each enzyme by the number of elementary modes in which it is involved or, conversely, by the number of modes remaining in the system deficient in the enzyme under study.

The structural robustness of microbial metabolism has been studied *in silico* using elementary mode analysis by Stelling *et al.* [6]. The following criterion was used: if, after deletion of an enzyme, at least one elementary mode allowing a positive growth yield remained, the mutant was predicted to be viable. To characterise robustness in more detail, the maximum possible growth yield was plotted versus the number of elementary modes remaining in various single mutants (Fig. 2*b* in [6]). This plot shows that the growth yield remains fairly constant even if the number of elementary modes drops significantly. Only when the latter number is very low, is the network no longer able to sustain growth.

In the present paper, we start from the reasoning that robustness is not perfectly identical with redundancy. In defining structural robustness, one should compare the entire system with a mutated system. By only considering the number of elementary modes, it is unclear what happens if an enzyme is knocked out. As we will show in Section 2 by way of an example, systems with the same number of elementary modes can have different robustnesses.

We propose three new measures of metabolic network robustness. They are based on elementary flux modes and take into account the effect of enzyme knockout. We illustrate the approach by simple hypothetical examples and two more complex systems describing central metabolism of *Escherichia coli* and the metabolism in human erythrocytes.

2 Measures of network robustness

Consider the two simple networks shown in Fig. 1. Both involve the same number (two) of elementary modes. However, the system in Fig. 1*a* is less robust because knockout of enzyme 1 deletes two elementary modes, while deleting only one mode in the system shown in Fig. 1*b*. Thus, the number of elementary modes (or extreme pathways) should not be used as a measure of robustness.

To characterise the structural robustness to the knockout (deficiency) of one enzyme, E_i, the ratio, $z^{(i)}/z$, between the number of elementary modes remaining after knockout $z^{(i)}$,
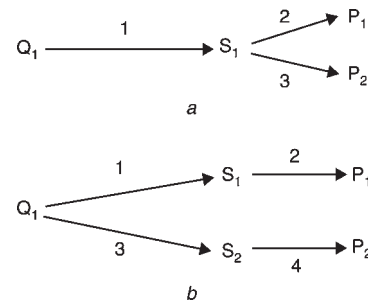

Fig. 1 *Simple hypothetical networks illustrating that the number of elementary modes is not an appropriate measure of robustness. All reactions are considered irreversible for simplicity's sake. S_i, internal metabolites; Q_k, substrates; P_k, products*

a The pathways producing P_1 and P_2 share one enzyme
b The pathways producing P_1 and P_2 do not share any enzyme

and the number in the unperturbed network z, can be used. This gives a normalised value between 0 and 1. The extreme values are reached when no elementary mode is left (zero robustness) and when all elementary modes remain (complete robustness). This definition is actually related to the earlier suggestion to characterise the importance of each enzyme by the number of modes in which it is involved [30, 31].

To quantify the global robustness of the entire network, the arithmetic mean of all these numbers can be taken

$$R_1 = \frac{\sum_{i=1}^{r} z^{(i)}}{r \cdot z} \qquad (1)$$

where r denotes the total number of reactions in the system. This quantity is again between 0 and 1.

Consider the simple example shown in Fig. 2. It gives rise to four elementary modes: $\{E_1, E_2, E_4\}$, $\{E_3, E_4\}$, $\{E_5, E_6\}$, and $\{E_5, E_7\}$. Knockout of enzymes E_4 or E_5 causes two elementary modes to disappear, while knockout of one of the remaining enzymes causes only one elementary mode to drop out (that is, three to remain). Accordingly, the robustness of this system is:

$$R_1 = \frac{2 \times 2 + 5 \times 3}{7 \times 4} \cong 0.679 \qquad (2)$$

This means that knockout of one enzyme implies that, on average, two-thirds of the pathways in the system are still present. For the systems shown in Figs. 1*a* and 1*b*, R_1 can readily be computed to be $1/3$ and $1/2$, respectively.

Special attention should be paid to reactions that are at thermodynamic equilibrium and hence, have zero net flux at any steady state of the system. Such reactions have been termed 'strictly detailed balanced reactions' and can be detected by analysing the nullspace matrix and checking a generalised Wegscheider condition [32]. They are not involved in any elementary mode. Therefore, their robustness measure $z^{(i)}/z$ equals one. However, for most

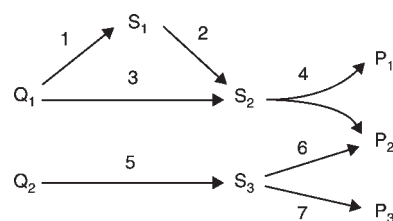**Fig. 2** *Simple hypothetical network used for illustrating the proposed robustness measures. All reactions are considered irreversible. S_i, internal metabolites; Q_k, substrates; P_k, products*

Syst. Biol., Vol. 1, No. 1, June 2004

115

applications, such enzymes should be excluded from the robustness analysis because they do not contribute to any net conversion. Otherwise, the robustness of a system would change as the number of strictly detailed balanced reactions increases because their knockout does not affect the number of elementary modes while the total number of reactions r, increases. Such a dependency of robustness on strictly detailed balanced reactions is not, however, biologically meaningful. Interestingly, if they are excluded, the robustness measure can never be exactly equal to one, since knockout of an enzyme then always deletes at least one elementary mode.

In (1), no difference is made between different products of the metabolic network. For example, the contribution to the robustness measure is the same when an elementary mode producing ATP drops out and when a mode producing NADPH is eliminated. For some applications, it is certainly of interest to distinguish between different products. To cope with such situations, we propose two further definitions.

We consider the sub-network consisting of all elementary modes producing a certain essential product P_k and apply definition (1). There may be other products (excreted byproducts such as CO_2 or ethanol) which are not included in the calculation. Let us consider the example of Fig. 2. Two elementary modes, consisting of four enzymes, produce P_1. Note that both of these start from Q_1, so that they refer to the same overall stoichiometry, which would be of importance for the redundancy definition of Papin *et al.* [20]. For our definition, however, we only consider the products. The robustness concerning product P_1 is:

$$R_1^{(1)} = \frac{1+1+1+0}{4 \times 2} = 0.375 \qquad (3)$$

For the other products we obtain $R_1^{(2)} = 11/18$ and $R_1^{(3)} = 0$. The latter result is a special case of the general fact that if $R_1^{(k)} = 0$, then each enzyme of the elementary modes producing P_k is an essential enzyme as defined by Klamt and Gilles [8].

Our second measure of metabolic network robustness is based on the assumption that all of the considered products are essential for the organism. That is, the organism is no longer viable as soon as one essential product cannot be produced anymore. The measure is defined as the minimal robustness concerning an essential product:

$$R_2 = \min\left\{R_1^{(1)}, R_1^{(2)}, \ldots, R_1^{(n)}\right\} \qquad (4)$$

For the example of Fig. 2, we obtain $R_2 = 0$, because $R_1^{(3)} = 0$.

However, it may occur that the robustness of one product is quite low, but that most of the random mutations would concern elementary modes producing other, more robust products. To take this into account we propose the arithmetic mean of the particular product robustness values as a third network robustness measure

$$R_3 = \frac{\sum_i R_1^{(i)}}{n} \qquad (5)$$

with n denoting the number of essential products. The example network of Fig. 2 has the robustness $R_3 \cong 0.329$. For special purposes, it could be sensible to consider other mean values or some special weights for the particular product robustness values. Here we just cope with the simplest generic measures.

3 Realistic examples: central metabolisms of *E. coli* and human erythrocytes

As a proof of concept, we study two realistic metabolic networks that have about the same number of elementary modes. The first illustrative example is the metabolism in human erythrocytes, which is a favourite subject of modelling studies [12, 13, 16, 17, 31, 33, 34]. The network is fully described in Appendix 7.1. It is related to a model used earlier [16, 31, 33]. The choice of external metabolites and exchange reactions is similar to that used by Wiback and Palsson [34]. The scheme comprises glycolysis, the pentose phosphate pathway, glutathione oxidation/reduction and adenine nucleotide metabolism. We consider glucose, NAD, NADP, NADH, NADPH, NH_3, CO_2, and 2,3-diphosphoglycerate (2,3DPG) as external metabolites (substrates and products). Moreover, pyruvate, lactate, inosine, adenosine, adenine and hypoxanthine are considered to be exchanged across the cell membrane by reactions of the form S = S_EXT with S_EXT considered to be external. The consumption of ATP in various cell processes is modelled as a reaction ATP = ADP_EXT + Pi_EXT. Wiback and Palsson [34] considered even more of such exchange reactions (with some of them describing consumption inside the cell), which increases the number of elementary modes. Although it is questionable whether such exchange reactions are relevant in erythrocytes, we use the network in this configuration because it gives rise to about the same number of elementary modes as the *E. coli* model described below. ATP, NADPH, 2,3DPG and hypoxanthine are here considered as essential products. The functions of ATP and NADPH as energy and redox 'currencies', respectively, are well known. 2,3DPG is an effector of oxygen binding to haemoglobin [3, 17]. In addition to oxygen transport, an important function of human red blood cells is the transport of purine bases from organs with excess purine to organs in which purines are required [35]. Hypoxanthine is the relevant product excreted by erythrocytes in this context. Our present choice of external metabolites increases the number of elementary modes from 21 [33] to 667.

The second network describes the central metabolism of *E. coli* (Appendix 7.2) and is based on a model published by Stelling *et al.* [6]. The original model involves substrate uptake (including the phosphotransferase system), central carbon metabolism (including glycolysis, the pentose phosphate pathway, tricarboxylic acid cycle and glyoxylate shunt), monomer and precursor synthesis (including the synthesis of all proteinogenic amino acids and the nucleotide triphosphates), polymer synthesis, biomass production (growth) and the excretion of several byproducts such as CO_2 and formate. Further details are given in Stelling *et al.* [6] and in the on-line material mentioned therein. Since we consider four relevant products of erythrocyte metabolism, we take the same number of products in the *E. coli* model. The original model [6] involves biomass and ATP as the relevant products. Here, by way of example, we consider different combinations of four amino acids as relevant products (see Table 1). Biomass production, the remaining amino acids, as well as the other precursors of biomass are excluded for simplicity's sake. To obtain about the same number of elementary modes, we take acetate as the only substrate. Other substrates (glycerol, glucose, or succinate) would yield many more elementary modes.

The results of our calculations are shown in Table 1. For the combination alanine, arginine, asparagine, and histidine in the *E. coli* model, the number of elementary modes

116

Syst. Biol., Vol. 1, No. 1, June 2004

Table 1: Values of the three robustness measures for the metabolic systems in human erythrocytes and *E. coli* under study

Metabolic network/ essential products	Number of elementary modes	R_1	R_2	R_3
Human erythrocyte ATP, hypoxanthine, NADPH, 2,3DPG	667	0.3834	0.3401	0.3607
E. coli				
Ala, Arg, Asn, His[§]	667	0.5084	0.3207	0.4295
Arg, Asn, His, Ile	656	0.5211	0.3451	0.4427
Arg, Asn, Ile, Leu	567	0.5479	0.4763	0.4964
Arg, Asn, Leu, Pro	540	0.5360	0.4586	0.4836
His, Ile, Leu, Lys	802	0.5112	0.3482	0.4437
Ile, Leu, Pro, Val	597	0.5488	0.4675	0.5058

[§] Amino acids are indicated in the usual three-letter code

exactly equals that in the erythrocyte model. Moreover, we have chosen another five combinations with about the same number of elementary modes. All the values of robustness R_1 lie between 0.38 and 0.55, implying that the number of elementary modes remaining after knockout of one enzyme is about one-third to one-half of the total number. The values of the other robustness measures are in a similar range, notably between 0.32 and 0.51. It can be seen that for all three robustness measures, the values for the erythrocyte model are lower than for the *E. coli* network, with one exception: R_2 is higher for the erythrocyte than for the combination Ala, Arg, Asn, His in *E. coli*. Apart from that, most of the values are markedly lower, especially in the case of measure R_1.

4 Conclusions

In this paper, we have proposed three new measures of network robustness. They are based on a comparison of the numbers of elementary flux modes in the unperturbed situation and after knockout of one enzyme, averaged over all enzymes.

For the measures R_2 and R_3 one has to consider all essential products separately (see Figs. 1 and 2). Note that in the general definition of elementary modes, only a distinction between internal and external metabolites is made. The latter include substrates and products. Under different conditions, one and the same external metabolite (e.g. ethanol in the case of yeast) may occur as a substrate or as a product. The robustness measures, however, should be calculated for one specific situation. In another situation, the set of essential products needs to be newly defined.

Ebenhöh and Heinrich [36] distinguish between strong and weak robustness. A metabolic network is strongly robust against a certain mutation (which in their analysis can be a knockout, change or even addition of an enzyme) if it can still produce the same products. It is weakly robust if it is not strongly robust but can still produce at least one product (not necessarily one of the original products). In our approach, a strongly robust network would have a robustness measure R_2 greater than zero (with all products taken as essential products), while for a weakly robust network, R_2 would be zero.

To illustrate the applicability of the proposed definitions, we have calculated the structural robustness measures for a model of human erythrocyte metabolism and for a model of the central metabolism of *E. coli*. The robustness of the former system is markedly lower than that of the latter system for all robustness measures and for all but one combination of products (four amino acids) of the *E. coli* system. This is in accordance with common biochemical knowledge saying that erythrocytes have a much simpler and thus, less robust metabolism than *E. coli*. The latter must be able to adapt to different conditions such as the human intestine, water of varying purity outside the human body, etc. In contrast, human erythrocytes live under relatively constant conditions in the blood. Our results are in agreement with experimental observations showing that most enzyme deficiencies, such as those of hexokinase, pyruvate kinase, or glucose-6-phosphate dehydrogenase, lead to severe diseases [37], while single-gene deletions of the majority of *E. coli* enzymes do not entail inviability. Interestingly, the calculated values of R_1 in *E. coli* ranging from 0.51 to 0.55 correspond well with the fraction 0.56 of viable mutants found by Stelling *et al.* [6] by literature search. It is worthwhile studying this relationship in more detail, both, experimentally and theoretically.

For the two considered systems, the robustness values R_1 are between 0.38 and 0.55. As most single mutants of *E. coli* are still able to grow [6], this implies that about one-third of the elementary modes in *E. coli* are often sufficient to sustain growth.

To demonstrate the applicability of our definitions, we have considered two systems with about the same number of elementary modes and the same number of biologically relevant products. This was to show that the robustness (according to our definitions) can differ nevertheless and hence, that the number of elementary modes is not an appropriate measure of robustness. However, since the proposed measures are normalised quantities, one can use them, in general, also to compare the robustness of systems with different numbers of elementary modes and different numbers of essential products. For example, we also analysed the original erythrocyte model of Joshi and Palsson [16] giving rise to 21 elementary modes [33]. We took hypoxanthine excretion and sodium/potassium transport as essential functions. This yielded the following robustness values: $R_1 = 0.4424$, $R_2 = 0.2029$, $R_3 = 0.2056$. Their difference to the values for *E. coli* is even more pronounced. In further investigations, we will apply the measures to *E. coli* under different conditions as well as to the metabolism in other cell types and other organisms. As R_2 and R_3 are defined as the minimum and average values of the same set of quantities, we obviously have $R_2 \leq R_3$. In all our numerical calculations, moreover, $R_3 < R_1$. It is worth trying to prove this relation analytically in a general way.

In future work, it will be of interest to apply the present method for comparing two micro-organisms or different groups of products in the same micro-organism, e.g. amino acid metabolism and lipid metabolism in *E. coli*. The question arises whether the present approach is scalable to larger, e.g. genome-wide, networks. This meets the above-mentioned problem of combinatorial explosion of elementary modes. However, since our robustness measures are ratios of numbers of elementary modes, it is of great interest to see whether the ratios representing the robustness measures can be computed directly without computing the elementary modes themselves. Another option is to decompose complex networks into smaller, tractable sub-networks [20, 26]. Approaches based on linear programming [5, 38] scale up more easily to larger systems. However, they only take into account the optimal situation

rather than all possible flux distributions in the system and thus, can hardly cope with network flexibility.

The definitions introduced here essentially take into account single mutants. In future work, it would be of interest to extend the analysis to double and multiple mutants. So far, there are only a few modelling studies on such mutants [38]. Klamt and Gilles [8] introduced the concept of minimal cut sets. These are minimal sets of enzymes whose deletion or complete inhibition prevents the operation of a target reaction under study. If the smallest of these sets involves, for example, two enzymes, any single mutant can still sustain the target reaction, but some double mutant cannot. Klamt and Gilles [8] introduced a fragility coefficient as the reciprocal of the average size of all minimal cut sets in which an enzyme E_i participates. A network fragility coefficient F, was defined as the average fragility coefficient over all enzymes. This coefficient takes into account both single and multiple knockouts. On the other hand, the fragility coefficient is based on an all-or-none decision whether or not a product can still be synthesised, while our measures have the advantage that the number of elementary modes is considered. Moreover, they are easier to compute. Our preliminary calculations show that for many simple systems (e.g. the systems shown in Figs. 1a and b), robustness R_1 and the coefficient F, calculated by taking all output reactions as target reactions, add up to one (note that high fragility implies low robustness), while for more complex systems such as *E. coli* and erythrocyte metabolisms, R_1 is larger than $1 - F$. For example, for the complete *E. coli* system (including biomass and ATP production) with acetate as the only substrate, our calculations give a value of $R_1 = 0.4084$, while F = 0.783 [8]. It will be of interest to elucidate the interrelation between these coefficients in more detail.

Further possibilities of extending the proposed definitions include the introduction of a weighted mean of product robustness values, since normally, different products are not equally important for the organism (e.g. ATP and 2,3DPG). For example, the weighting factors could be the normalised numbers of elementary modes producing the product in question or the normalised numbers of enzymes involved. The latter option appears to be sensible if each enzyme is subject to failure with the same probability.

In future studies, it will be of interest to analyse the change of robustness of metabolism during biological evolution [36]. While one would assume an increase in robustness, the opposite change can have happened as well, for example, in the evolution of intracellular parasites. In this context, it is worth studying the evolution of enzymes with broad substrate specificity. It has been argued that enzymes with high specificity have developed from low-specificity ancestors during biological evolution [39]. One reason for this development may be an increase in robustness. Two specialised enzymes cause the system to have a greater robustness than one less specific enzyme because the knockout of the latter would generally delete more pathways.

5 Acknowledgments

6 References

1 Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., *et al.*: 'Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis', *Science*, 1999, **285**, pp. 901–906

2 Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., *et al.*: 'Functional profiling of the *Saccharomyces cerevisiae* genome', *Nature*, 2002, **418**, pp. 387–391

3 Bossi, D., and Giardina, B.: 'Red cell physiology', *Mol. Asp. Med.*, 1996, **17**, pp. 117–128

4 Barkai, N., and Leibler, S.: 'Robustness in simple biochemical networks', *Nature*, 1997, **387**, pp. 913–917

5 Edwards, J.S., and Palsson, B.O.: 'Robustness analysis of the *Escherichia coli* metabolic network', *Biotechnol. Prog.*, 2000, **16**, pp. 927–939

6 Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E.D.: 'Metabolic network structure determines key aspects of functionality and regulation', *Nature*, 2002, **420**, pp. 190–193

7 Morohashi, M., Winn, A.E., Borisuk, M.T., Bolouri, H., Doyle, J., and Kitano, H.: 'Robustness as a measure of plausibility in models of biochemical networks', *J. Theor. Biol.*, 2002, **216**, pp. 19–30

8 Klamt, S., and Gilles, E.D.: 'Minimal cut sets in biochemical reaction networks', *Bioinformatics*, 2004, **20**, pp. 226–234

9 Rohwer, J.M., and Botha, F.C.: 'Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data', *Biochem. J.*, 2001, **358**, pp. 437–445

10 Kato, M., Mizuno, K., Crozier, A., Fujimura, T., and Ashihara, H.: 'Caffeine synthase gene from tea leaves', *Nature*, 2000, **406**, pp. 956–957

11 Mavrovouniotis, M.L., Stephanopoulos, G., and Stephanopoulos, G.: 'Computer-aided synthesis of biochemical pathways', *Biotechnol. Bioeng.*, 1990, **36**, pp. 1119–1132

12 Schuster, R., and Holzhütter, H.G.: '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.*, 1995, **229**, pp. 403–418

13 Martinov, M.V., Plotnikov, A.G., Vitvitsky, V.M., and Ataullakhanov, F.I.: 'Deficiencies of glycolytic enzymes as a possible cause of hemolytic anemia', *Biochim. Biophys. Acta.*, 2000, **1474**, pp. 75–87

14 Bakker, B.M., Mensonides, F.I.C., Teusink, B., Van Hoek, P., Michels, P.A.M., and Westerhoff, H.V.: 'Compartmentation protects trypanosomes from the dangerous design of glycolysis', *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, pp. 2087–2092

15 Wagner, C.: 'Systembiologie gegen Parasiten', *BioWorld*, 2004, **9**, (1), pp. 2–4

16 Joshi, A., and Palsson, B.O.: 'Metabolic dynamics in the human red cell. Part I. A comprehensive kinetic model', *J. Theor. Biol.*, 1989, **141**, pp. 515–528

17 Mulquiney, P.J., and Kuchel, P.W.: 'Model of 2,3-biphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: equations and parameter refinement', *Biochem. J.*, 1999, **342**, pp. 581–596

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

19 Schuster, S., Fell, D.A., and Dandekar, T.: 'A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks', *Nature Biotechnol.*, 2000, **18**, pp. 326–332

20 Papin, J.A., Price, N.D., Edwards, J.S., and Palsson, B.O.: 'The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy', *J. Theor. Biol.*, 2002, **215**, pp. 67–82

21 Price, N.D., Papin, J.A., and Palsson, B.O.: 'Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis', *Genome Res.*, 2002, **12**, pp. 760–769

22 Poolman, M.G., Fell, D.A., and Raines, C.A.: 'Elementary modes analysis of photosynthate metabolism in the chloroplast stroma', *Eur. J. Biochem.*, 2003, **270**, pp. 430–439

23 Carlson, R., and Srienc, F.: 'Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: Identification of reactions', *Biotechnol. Bioeng.*, 2004, **85**, pp. 1–19

24 Pfeiffer, T., Sánchez-Valdenebro, I., Nuño, J.C., Montero, F., and Schuster, S.: 'METATOOL: For studying metabolic networks', *Bioinformatics*, 1999, **15**, pp. 251–257

25 Klamt, S., Stelling, J., Ginkel, M., and Gilles, E.D.: 'FluxAnalyzer: Exploring structure, pathways, and fluxes in balanced metabolic networks by interactive flux maps', *Bioinformatics*, 2003, **19**, pp. 261–269

26 Dandekar, T., Moldenhauer, F., Bulik, S., Bertram, H., and Schuster, S.: 'A method for classifying metabolites in topological pathway analyses based on minimization of pathway number', *BioSystems*, 2003, **70**, pp. 255–270

27 Çakir, T., Kirdar, B., and Ülgen, K.Ö.: 'Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks', *Biotechnol. Bioeng.*, 2004, **86**, pp. 251–260

28 Klamt, S., and Stelling, J.: 'Two approaches for metabolic pathway analysis?', *Trends Biotechnol.*, 2003, **21**, pp. 64–69

29 Palsson, B.O., Price, N.D., and Papin, J.A.: 'Development of network-based pathway definitions: the need to analyze real metabolic networks', *Trends Biotechnol.*, 2003, **21**, pp. 195–198

30 Oancea, I.: 'Topological analysis of metabolic and regulatory networks

118

Syst. Biol., Vol. 1, No. 1, June 2004

by decomposition methods'. PhD thesis, Humboldt University, Berlin, Germany, 2003

31 Çakir, T., Tacer, C.S., and Ülgen, K.Ö.: 'Metabolic pathway analysis of enzyme-deficient human red blood cells', *BioSystems*, 2004, in press

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

33 Schuster, S., Fell, D.A., Pfeiffer, T., Dandekar, T., and Bork, P.: 'Elementary modes analysis illustrated with human red cell metabolism' in Larsson, C., Påhlman, I.-L., and Gustafsson, L., (Eds.): 'BioThermoKinetics in the Post Genomic Era' (Chalmers, Göteborg, 1998), pp. 332–339

34 Wiback, S.J., and Palsson, B.O.: 'Extreme pathway analysis of human red blood cell metabolism', *Biophys. J.*, 2002, **83**, pp. 808–818

35 Salerno, C., and Giacomello, A.: 'Hypoxanthine-guanine exchange by intact human erythrocytes', *Biochem.*, 1985, **24**, pp. 1306–1309

36 Ebenhöh, O., and Heinrich, R.: 'Stoichiometric design of metabolic networks: multifunctionality, clusters, optimization, weak and strong robustness', *Bull. Math. Biol.*, 2003, **65**, pp. 323–357

37 Scriver, C.R., and Sly, W.L., (Eds.): 'The Metabolic and Molecular Bases of Inherited Disease' (McGraw-Hill, New York, 1995), vol. III

38 Edwards, J.S., and Palsson, B.O.: 'Systems properties of the *Haemophilus influenzae* Rd metabolic genotype', *J. Biol. Chem.*, 1999, **274**, pp. 17410–17416

39 Kacser, H., and Beeby, R.: 'Evolution of catalytic proteins or on the origin of enzyme species by means of natural selection', *J. Mol. Evol.*, 1984, **20**, pp. 38–51

7 Appendix

7.1 Input file of the erythrocyte system for the program METATOOL

The identifiers have the following meaning: ENZREV, reversible enzymes; ENZIRREV, irreversible enzymes; METINT, internal metabolites; METEXT, external metabolites; CAT, catalyzed reactions.

-ENZREV
PGI ALD TPI GAPDH PGK PGM EN LD PGL RPI XPI TKI TA TKII PRM PNPase ApK PRY:TP LAC:TP ATP:EX INO:TP ADO:TP ADE:TP

-ENZIRREV
HK PFK DPGM DPGase PK G6PDH PDGH PRPPsyn HGPRT AdPRT IMPase AMPDA AMPase ADA AK HX:TP

-METINT
G6P F6P FDP DHAP GA3P 13DPG 3PG 2PG PEP PYR LAC 6PGL 6PGC RL5P X5P R5P ADP Pi S7P E4P PRPP IMP R1P HX INO ADE ADO AMP ATP

-METEXT
NAD NADP NADH NADPH H NH3 CO2 H2O PYR_EXT LAC_EXT ADP_EXT Pi_EXT INO_EXT ADO_EXT HX_EXT ADE_EXT 23DPG

-CAT
HK : 1 ATP = 1 G6P + 1 ADP + 1 H .
PGI : 1 G6P = 1 F6P .
PFK : 1 F6P + 1 ATP = 1 FDP + 1 ADP + 1 H .
ALD : 1 FDP = 1 DHAP + 1 GA3P .
TPI : 1 DHAP = 1 GA3P .
GAPDH : 1 GA3P + 1 NAD + 1 Pi = 1 13DPG + 1 NADH + 1 H .
PGK : 1 13DPG + 1 ADP = 1 3PG + 1 ATP .
DPGM : 1 13DPG = 1 23DPG + 1 H .
DPGase : 1 23DPG + 1 H2O = 1 3PG + 1 Pi .
PGM : 1 3PG = 1 2PG .
EN : 1 2PG = 1 PEP + 1 H2O .
PK : 1 PEP + 1 ADP + 1 H = 1 PYR + 1 ATP .
LD : 1 PYR + 1 NADH + 1 H = 1 LAC + 1 NAD .
G6PDH : 1 G6P + 1 NADP = 1 6PGL + 1 NADPH + 1 H .
PGL : 1 6PGL + 1 H2O = 1 6PGC + 1 H .
PDGH : 1 6PGC + 1 NADP = 1 RL5P + 1 NADPH + 1 CO2 .
RPI : 1 RL5P = 1 R5P .
XPI : 1 RL5P = 1 X5P .

TKI : 1 X5P + 1 R5P = 1 GA3P + 1 S7P .
TA : 1 GA3P + 1 S7P = 1 F6P + 1 E4P .
TKII : 1 X5P + 1 E4P = 1 F6P + 1 GA3P .
PRPPsyn : 1 R5P + 1 ATP = 1 PRPP + 1 AMP .
PRM : 1 R1P = 1 R5P .
HGPRT : 1 PRPP + 1 HX + 1 H2O = 1 IMP + 2 Pi .
AdPRT : 1 PRPP + 1 ADE + 1 H2O = 1 AMP + 2 Pi .
PNPase : 1 INO + 1 Pi = 1 R1P + 1 HX .
IMPase : 1 IMP + 1 H2O = 1 INO + 1 H + 1 Pi .
AMPDA : 1 AMP + 1 H2O = 1 IMP + 1 NH3 .
AMPase : 1 AMP + 1 H2O = 1 ADO + 1 H + 1 Pi .
ADA : 1 ADO + 1 H2O = 1 INO + 1 NH3 .
AK : 1 ADO + 1 ATP = 1 AMP + 1 ADP .
ApK : 2 ADP = 1 AMP + 1 ATP .
PRY:TP : 1 PYR_EXT = 1 PYR .
LAC:TP : 1 LAC_EXT = 1 LAC .
ATP:EX : 1 ATP = 1 ADP_EXT + Pi_EXT .
INO:TP : 1 INO_EXT = 1 INO .
ADO:TP : 1 ADO_EXT = 1 ADO .
HX:TP : 1 HX = 1 HX_EXT .
ADE:TP : 1 ADE_EXT = 1 ADE .

7.2 Input file of the E. coli system for the program METATOOL adapted from [6]

The identifiers have the same meaning as given in Appendix 7.1. All the 20 proteinogenic amino acids are here indicated as internal metabolites. In the six different versions used in the calculations, different sets of four amino acids are set to external status (see main text).

-ENZREV
CO2_ex G6P::F6P F16P::T3P DHAP::G3P G3P::DPG DPG::3PG 3PG::2PG 2PG::PEP Cit::ICit ICit::alKG SuccCoA::Succ Fum::Mal Mal::OxA G6P::PGlac AcCoA::Adh Adh::Eth Rl5P::X5P Rl5P::R5P Transket1 Transaldo Transket2 AcCoA::AcP AcP::Ac Pyr::Lac NADHDehydro TransHydro ATPSynth MTHF_Synth

-ENZIRREV
O2_up N_up S_up DHAP::Glyc3P Lac_ex Eth_ex Ac_ex Ac_up Form_ex F16P::F6P F6P::F16P PEP::PYR Pyr::PEP PYR::AcCoA AcCoA::Cit alKG::SuccCoA Succ::Fum Fum::Succ ICit::Glyox Glyox::Mal PGlac::PGluc PGluc::Rl5P OxA::PEP PEP::OxA Pyr::Form Oxidase ATPdrain Chor_Synth PRPP_Synth Ala_Synth Val_Synth Leu_Synth Asn_Synth_1 Asp_synth Asp::Fum Asp:: AspSAld AspSAld::HSer Lys_Synth Met_Synth Thr_Synth Ile_Synth His_Synth Glu_synth Gln_Synth Pro_Synth Arg_Synth Trp_Synth Tyr_Synth Phe_Synth Ser_ Synth Gly_Synth Cys_Synth rATP_Synth rGTP_Synth rCTP_Synth rUTP_Synth dATP_Synth dGTP_Synth dCTP_Synth dTTP_Synth mit_FS_Synth UDPGlc_Synth CDPEth_Synth OH_myr_ac_Synth C14_0_FS_Synth CMP_KDO_Synth NDPHep_Synth TDPGlcs_Synth UDP_NAG_Synth UDP_NAM_Synth di_am_pim_Synth ADPGlc_Synth Mal::Pyr Pyr::Ac Ac::AcCoA

-METINT
G6P F6P F16P DHAP Glyc3P G3P DPG 3PG 2PG PEP Pyr AcCoA Cit ICit alKG SuccCoA Succ Fum Mal OxA Glyox R5P Rl5P E4P X5P S7P PGlac PGluc ATP NADH NADPH QuiH2 H_ex O2 CO2 N S AcP Ac Form Lac Adh Eth Chor PRPP MTHF AspSAld HSer rATP rGTP rCTP rUTP dATP dGTP dCTP dTTP mit_FS UDPGlc CDPEth OH_myr_ac C14_0_FS CMP_KDO NDPHep TDPGlcs UDP_NAG UDP_NAM di_am_pim ADPGlc Cys Asp Glu Phe Gly Ile Lys Leu Met Pro Gln Ser Thr Val Trp Tyr Ala His Asn Arg

Syst. Biol., Vol. 1, No. 1, June 2004

119

-METEXT
O2_ext N_ext CO2_ext Lac_ext Eth_ext Ac_ext Form_ext
ATP_ext

-CAT
O2_up : O2_ext = 1 O2 .
N_up : N_ext = 1 N .
CO2_ex : 1 CO2 = CO2_ext .
S_up : 4 ATP + 4 NADPH = 1 S .
DHAP::Glyc3P : 1 DHAP + 1 NADH = 1 Glyc3P .
Lac_ex : 1 Lac = Lac_ext .
Eth_ex : 1 Eth = Eth_ext .
Ac_ex : 1 Ac = Ac_ext .
Ac_up : Ac_ext = 1 Ac .
Form_ex : 1 Form = Form_ext .
G6P::F6P : 1 G6P = 1 F6P .
F16P::F6P : 1 F16P = 1 F6P .
F6P::F16P : 1 F6P + 1 ATP = 1 F16P .
F16P::T3P : 1 F16P = 1 DHAP + 1 G3P .
DHAP::G3P : 1 DHAP = 1 G3P .
G3P::DPG : 1 G3P = 1 DPG + 1 NADH .
DPG::3PG : 1 DPG = 1 3PG + 1 ATP .
3PG::2PG : 1 3PG = 1 2PG .
2PG::PEP : 1 2PG = 1 PEP .
PEP::PYR : 1 PEP = 1 Pyr + 1 ATP .
Pyr::PEP : 1 Pyr + 2 ATP = 1 PEP .
PYR::AcCoA : 1 Pyr = 1 AcCoA + 1 NADH + 1 CO2 .
AcCoA::Cit : 1 AcCoA + 1 OxA = 1 Cit .
Cit::ICit : 1 Cit = 1 ICit .
ICit::alKG : 1 ICit = 1 alKG + 1 NADPH + 1 CO2 .
alKG::SuccCoA : 1 alKG = 1 SuccCoA + 1 NADH + 1 CO2 .
SuccCoA::Succ : 1 SuccCoA = 1 Succ + 1 ATP .
Succ::Fum : 1 Succ = 1 Fum + 1 QuiH2 .
Fum::Succ : 1 Fum + 1 QuiH2 = 1 Succ .
Fum::Mal : 1 Fum = 1 Mal .
Mal::OxA : 1 Mal = 1 OxA + 1 NADH .
ICit::Glyox : 1 ICit = 1 Succ + 1 Glyox .
Glyox::Mal : 1 AcCoA + 1 Glyox = 1 Mal .
G6P::PGlac : 1 G6P = 1 PGlac + 1 NADPH .
AcCoA::Adh : 1 AcCoA + 1 NADH = 1 Adh .
Adh::Eth : 1 NADH + 1 Adh = 1 Eth .
PGlac::PGluc : 1 PGlac = 1 PGluc .
PGluc::Rl5P : 1 PGluc = 1 Rl5P + 1 NADPH + 1 CO2 .
Rl5P::X5P : 1 Rl5P = 1 X5P .
Rl5P::R5P : 1 Rl5P = 1 R5P .
Transket1 : 1 R5P + 1 X5P = 1 G3P + 1 S7P .
Transaldo : 1 G3P + 1 S7P = 1 F6P + 1 E4P .
Transket2 : 1 E4P + 1 X5P = 1 F6P + 1 G3P .
OxA::PEP : 1 OxA + 1 ATP = 1 PEP + 1 CO2 .
PEP::OxA : 1 PEP + 1 CO2 = 1 OxA .
AcCoA::AcP : 1 AcCoA = 1 AcP .
AcP::Ac : 1 AcP = 1 ATP + 1 Ac .
Pyr::Form : 1 Pyr = 1 AcCoA + 1 Form .
Pyr::Lac : 1 Pyr + 1 NADH = 1 Lac .
NADHDehydro : 1 NADH = 1 QuiH2 + 2 H_ex .
Oxidase : 1 QuiH2 + 0.5 O2 = 2 H_ex .
TransHydro : 1 NADH + 1 H_ex = 1 NADPH .
ATPSynth : 3 H_ex = 1 ATP .
ATPdrain : 1 ATP = ATP_ext .
Chor_Synth : 2 PEP + 1 E4P + 1 ATP + 1 NADPH =
    1 Chor .
PRPP_Synth : 1 R5P + 2 ATP = 1 PRPP .
MTHF_Synth : 1 ATP + 1 NADPH = 1 MTHF .
Ala_Synth : 1 Pyr + 1 Glu = 1 alKG + 1 Ala .
Val_Synth : 2 Pyr + 1 NADPH + 1 Glu = 1 alKG +
    1 CO2 + 1 Val .
Leu_Synth : 2 Pyr + 1 AcCoA + 1 NADPH + 1 Glu =
    alKG + 1 NADH + 2 CO2 + 1 Leu .

Asn_Synth_1 : 2 ATP + 1 N + 1 Asp = 1 Asn .
Asp_synth : 1 OxA + 1 Glu = 1 alKG + 1 Asp .
Asp::Fum : 1 Asp = 1 Fum + 1 N .
Asp::AspSAld : 1 ATP + 1 NADPH + 1 Asp = 1 AspSAld .
AspSAld::HSer : 1 NADPH + 1 AspSAld = 1 HSer .
Lys_Synth : 1 di_am_pim = 1 CO2 + 1 Lys .
Met_Synth : 1 SuccCoA + 1 MTHF + 1 HSer + 1 Cys =
    1 Pyr + 1 Succ + 1 N + 1 Met .
Thr_Synth : 1 ATP + 1 HSer = 1 Thr .
Ile_Synth : 1 Pyr + 1 NADPH + 1 Glu + 1 Thr = 1 alKG +
    1 CO2 + 1 N + 1 Ile .
His_Synth : 1 ATP + 1 PRPP + 1 Gln = 1 alKG +
    2 NADH + 1 His .
Glu_synth : 1 alKG + 1 NADPH + 1 N = 1 Glu .
Gln_Synth : 1 ATP + 1 N + 1 Glu = 1 Gln .
Pro_Synth : 1 ATP + 2 NADPH + 1 Glu = 1 Pro .
Arg_Synth : 1 AcCoA + 4 ATP + 1 NADPH + 1 CO2 +
    1 N + 1 Asp + 2 Glu = 1 alKG + 1 Fum+
    1 Ac + 1 Arg .
Trp_Synth : 1 Chor + 1 PRPP + 1 Gln + 1 Ser =
    1 G3P + 1 Pyr + 1 CO2 + 1 Glu + 1 Trp .
Tyr_Synth : 1 Chor + 1 Glu = 1 alKG + 1 NADH +
    1 CO2 + 1 Tyr .
Phe_-
Synth : 1 Chor + 1 Glu = 1 alKG + 1 CO2 + 1 Phe .
Ser_Synth : 1 3PG + 1 Glu = 1 alKG + 1 NADH + 1 Ser .
Gly_Synth : 1 Ser = 1 MTHF + 1 Gly .
Cys_Synth : 1 AcCoA + 1 S + 1 Ser = 1 Ac + 1 Cys .
rATP_Synth : 5 ATP + 1 CO2 + 1 PRPP + 2 MTHF +
  2 Asp + 1 Gly + 2 Gln = 2 Fum + 1 NADPH +
  2 Glu + 1 rATP .
rGTP_Synth : 6 ATP + 1 CO2 + 1 PRPP + 2 MTHF +
    1 Asp + 1 Gly + 3 Gln = 2 Fum + 1 NADH+
    1 NADPH + 3 Glu + 1 rGTP .
rCTP_Synth : 1 ATP + 1 Gln + 1 rUTP = 1 Glu + 1 rCTP .
rUTP_Synth : 4 ATP + 1 N + 1 PRPP + 1 Asp =
    1 NADH + 1 rUTP .
dATP_Synth : 1 NADPH + 1 rATP = 1 dATP .
dGTP_Synth : 1 NADPH + 1 rGTP = 1 dGTP .
dCTP_Synth : 1 NADPH + 1 rCTP = 1 dCTP .
dTTP_Synth : 2 NADPH + 1 MTHF + 1 rUTP = 1 dTTP .
mit_FS_Synth : 8.24 AcCoA + 7.24 ATP+
    13.91 NADPH = 1 mit_FS .
UDPGlc_Synth : 1 G6P + 1 ATP = 1 UDPGlc .
CDPEth_Synth : 1 3PG + 3 ATP + 1 NADPH + 1 N =
    1 NADH + 1 CDPEth .
OH_myr_ac_Synth : 7 AcCoA + 6 ATP + 11 NADPH =
    1 OH_myr_ac .
C14_0_FS_Synth : 7 AcCoA + 6 ATP + 12 NADPH =
    1 C14_0_FS .
CMP_KDO_Synth : 1 PEP + 1 R5P + 2 ATP =
    1 CMP_KDO .
NDPHep_Synth : 1.5 G6P + 1 ATP = 4 NADPH +
  1 NDPHep .
TDPGlcs_Synth : 1 F6P + 2 ATP + 1 N = 1 TDPGlcs .
UDP_NAG_Synth : 1 F6P + 1 AcCoA + 1 ATP +
    1 Gln = 1 Glu + 1 UDP_NAG .
UDP_NAM_Synth : 1 PEP + 1 NADPH + 1 UDP_NAG =
    1 UDP_NAM .
di_am_pim_Synth : 1 Pyr + 1 SuccCoA + 1 NADPH+
    1 AspSAld + 1 Glu = 1 alKG+
    1 Succ + 1 di_am_pim .
ADPGlc_Synth : 1 G6P + 1 ATP = 1 ADPGlc .
Mal::Pyr : 1 Mal = 1 Pyr + 1 NADH + 1 CO2 .
Pyr::Ac : 1 Pyr = 1 QuiH2 + 1 CO2 + 1 Ac .
Ac::AcCoA : 2 ATP + 1 Ac = 1 AcCoA .

120

Syst. Biol., Vol. 1, No. 1, June 2004