A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks

Stefan Schuster[1], David A. Fell[2], and Thomas Dandekar[1,3]

[1]Department of Bioinformatics, Max Delbrück Center for Molecular Medicine, D-13092 Berlin-Buch, Germany. [2]School of Biological and Molecular Sciences, Oxford Brookes University, Oxford OX3 0BP, England. [3]Biocomputing and Structures Program, EMBL, D-69012 Heidelberg, Germany.

Received 30 October 1998; accepted 21 December 1999

A set of linear pathways often does not capture the full range of behaviors of a metabolic network. The concept of 'elementary flux modes' provides a mathematical tool to define and comprehensively describe all metabolic routes that are both stoichiometrically and thermodynamically feasible for a group of enzymes. We have used this concept to analyze the interplay between the pentose phosphate pathway (PPP) and glycolysis. The set of elementary modes for this system involves conventional glycolysis, a futile cycle, all the modes of PPP function described in biochemistry textbooks, and additional modes that are a priori equally entitled to pathway status. Applications include maximizing product yield in amino acid and antibiotic synthesis, reconstruction and consistency checks of metabolism from genome data, analysis of enzyme deficiencies, and drug target identification in metabolic networks.

Keywords: elementary modes, metabolic networks, functional genomics, stochiometric network analysis

Cellular metabolism is often much more plastic than the set of pathways defined in biochemistry textbooks[1,2]. It is not always straightforward to determine the metabolic route that leads from a particular starting material to (a) given product(s), or to decide whether a particular enzyme is essential in the process. These issues have attracted increased interest with the growth of functional genomics[3–8]. For example, DNA arrays allow study of comprehensive patterns of gene expression[6]. Will it be possible to interpret these data in terms of how the metabolic phenotype is changing?

Currently, three different but related approaches are advocated for mapping biochemical networks without preconceptions. Each of these benefits from not requiring kinetic information. First, methods were developed for constructing transformation routes leading from a given substrate to a given product by successive addition of reaction steps[9,10]. Second, it was suggested to use a set of linearly independent basis vectors in flux space[5,11,12]. In other words, such a set contains flux distributions (regarded geometrically as vectors in a space) from which, by adding or substracting multiples of them, all admissible flux distributions can be obtained. However, the choice of the basis vectors to describe this space is not unique. Finally, building on the stoichiometric network analysis of Clarke[13] and earlier work on futile cycles[14], the concept of 'elementary flux modes'[15–17] was introduced. An elementary mode is a minimal set of enzymes that could operate at steady state, with the enzymes weighted by the relative flux they need to carry for the mode to function (Figs 1, 2). 'Minimal' means that if only the enzymes belonging to this set were operating, complete inhibition of one of these would lead to cessation of any steady-state flux in the system. This allows detection of the full set of nondecomposable steady-state flows that the network can support, including cyclic flows. Any steady-state flux pattern can be expressed as a non-negative linear combination of these modes (in a sense, a mixture of pure states) to form a unique set. For a comparison of the three methods, see also ref. 18.

For reasons to be explained later, information on kinetics and regulatory interactions is not needed to define the modes. Thus, it can easily be determined whether a stoichiometrically balanced path exists between a particular set of substrates and products, capturing the essential network character of biochemical transformations. Note that it would not suffice to construct linear paths by following the successive conversions of metabolite molecules, because by-products that cannot be excreted by the cell have to be balanced by additional reactions, which produce further by-products, and so on.

We make a distinction between reversible and irreversible reactions. Irreversibility here means that the net flux always has the same sign under physiological conditions (Fig. 1). Any elementary mode has to use the irreversible reactions in the appropriate direction. In contrast to other approaches[5,9,13], it is unnecessary to split reversible reactions into two oppositely directed irreversible steps, which would increase the number of the fluxes to be computed and require a time-consuming elimination of spurious cycles within reversible reactions. On the other hand, one often not only obtains modes situated on the boundary ("extreme currents" in Clarke's terminology[13]) of the admissible flux region (i.e. defined by flux combinations that correspond to valid steady states of the metabolic network), but also modes interior of this region, that are not independant of the extreme currents in a mathmatical sense, but are also elementary in the sense of not being decomposable[15,16]. The latter are elementary as well in the sense of not being decomposable into simpler modes.

Metabolites are classified as internal or external according to whether or not they are to fulfill the quasi-steady-state condition. In other words, the total rate of production of each internal metabolite equals the total rate of its consumption. In contrast, external metabolites (which are alternatively called pool metabolites, or sources and sinks, and need not be extracellular) do not fulfill this condition because they participate in additional reactions that are not involved in the system under study.

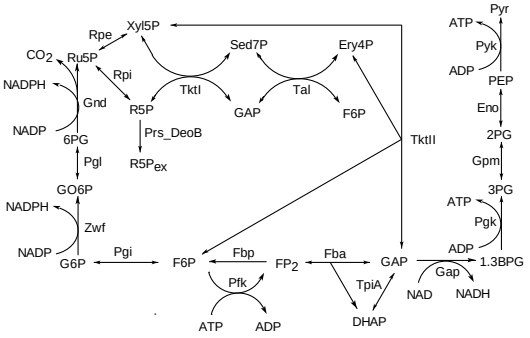

Figure 1. Reaction scheme representing part of monosaccharide metabolism. Reversible reactions are indicated by double arrowheads. The sign definition for the reversible reactions is such that the flux is taken positive if it goes from bottom to top or from left to right in the scheme. Abbreviations are as in Table 1, except for 2PG, 2-phosphoglycerate; 3PG, 3-phosphoglycerate; 6PG, 6-phosphogluconate; 1,3BPG, 1,3-bisphosphoglycerate. $R5P_{ex}$ stands for the ribose (or deoxyribose) moiety involved in RNA, DNA, NAD, ATP etc.

Here, elaborating on our previous work[8,15–17,19], we illustrate the usefulness of elementary mode analysis for systematically organizing and analyzing metabolic networks and for formalizing the concept of a biochemical pathway, with a running example taken from monosaccharide metabolism and discussing other biotechnologically and medically important examples. We show here the additional modes of transformation that can occur in the cell with the enzyme set already well known for glycolysis and PPP. Moreover, in the Computational protocol section, we give a detailed explanation of an efficient algorithm for performing this analysis as applied to the same example.

We explain how elementary mode analysis can test a set of enzymes for production of a desired product and detect non-redundant pathways and enzymes and pathways with maximal molar yields as well as close alternatives. Reconstruction of metabolism from annotated genome sequences and analyzing the impact of enzyme deficiencies are additional applications shown. Metabolic potentiation or reduction of drug effects can also be analyzed by our method, but are not discussed here.

Results

For illustration, we consider the reaction scheme representing part of monosaccharide metabolism shown in Figure 1. Apart from the declaration of internal and external substances, the enzyme reactions and their directionality are the only input required; they are listed on the left side of Table 1. We neglect pyruvate, water dikinase (Pps) (therefore not shown in Fig. 1). Glucose-6-phosphate (G6P) is considered external because it is formed from glucose taken up into the cell and can be isomerized to G1P, which is involved both in glycogen synthesis and glycogen degradation. The right portion of Table 1 shows a list of annotated enzyme genes in *Escherichia coli* and *Treponema pallidum*. In elementary mode analysis, it is useful to lump isoenzymes. For example, there are two different genes both for phosphofructokinase (*pfk*A and *pfk*B) and transketolase (*tkt*A and *tkt*B) present in *E. coli*. As their gene products perform the same conversion, we need not distinguish them in our analysis, although they may have different regulatory properties. The elementary modes are calculated without reference to the special regulatory state in the cell. A similar reasoning applies to TktA and TktB, whereas we do distinguish between the two different *reactions* TktI and TktII. This is always necessary in the case of bi- and multifunctional enzymes because, in principle, the different reactions might participate in different pathways. From this input, our program (see Computational protocol) yields as output the seven elementary modes shown in the upper part of Table 2. These modes, which involve different numbers of enzymes and partly overlap each other (Fig. 2), are amenable to the following biochemical interpretation: (a) Mode 1 represents the usual glycolytic pathway. Note that in *E. coli* and other organisms using the phosphotransferase system (PTS), half the PEP produced is converted to pyruvate by the PTS rather than by Pyk. (b, c) Modes 2 and 3 degrade G6P to pyruvate and CO_2 producing ATP, NADPH, and NADH. (d) Mode 4 converts G6P into ribose-5-phosphate (R5P) and CO_2. It is of importance when the metabolic needs for reducing power for biosynthetic purposes and R5P in nucleotide biosynthesis are balanced. (e) In mode 5, five hexoses are converted into six pentoses. This is important when the need for R5P exceeds that for NADPH. Note that depending on the type of utilization of R5P, additional ATP consumption may occur in modes 4 and 5 (e.g., hydrolysis of ATP to AMP in Prs, but not in DeoB). (f) To mode 6, the term 'pentose phosphate cycle' applies best, because carbons are cycled several times before ending up in CO_2. It is relevant when much more NADPH than R5P is required (like modes 2 and 3), but does not lead to NADH and ATP production. (g) Mode 7 is the futile cycle formed by Pfk and Fbp (see below). Interestingly, modes 3–6 exactly coincide with the four modes of the PPP described in Stryer[20]. Our approach provides an algorithmic method for finding these situations

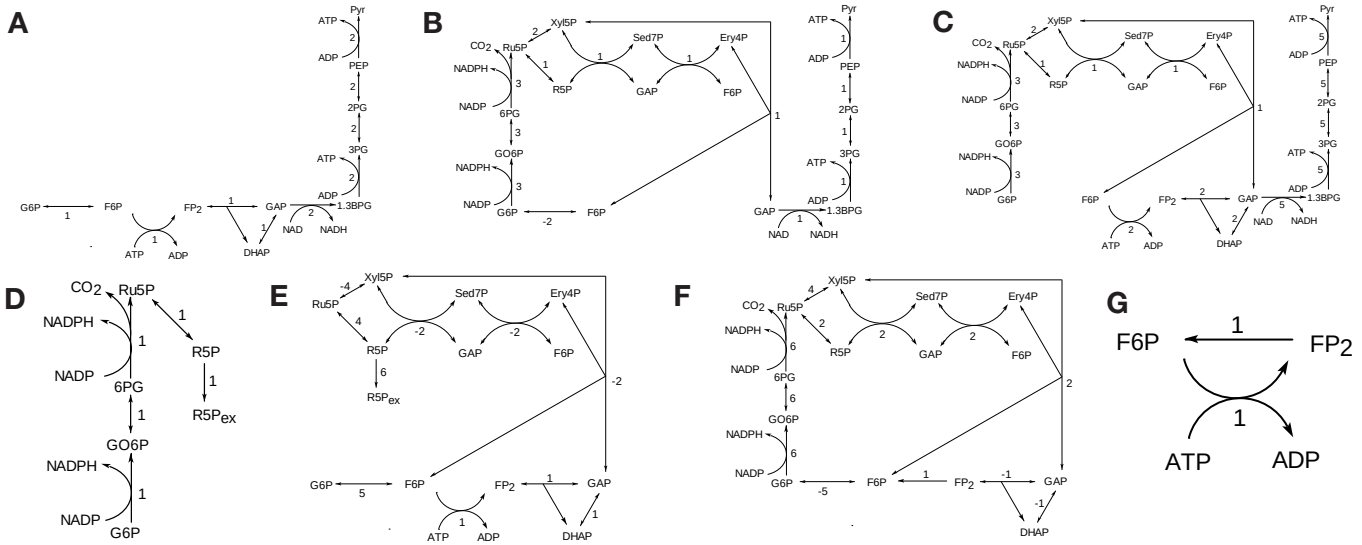

Figure 2. Graphical representation of the elementary modes pertaining to the reaction scheme in Figure 1. (A) through G) correspond to modes (1) through (7) given in the text. The numbers indicate the relative flux carried by the enzymes.

Table 1. Enzyme and reaction input for the glycolysis and *PPP* example shown in Fig. 1[a].

External substances:	Internal substances:
R5Pex, CO2, NADPH, NADP, Pyr, NADH, NAD, Pi, AMP, ADP, ATP, G6P	Ery4P, Sed7P, Xyl5P, R5P, Ru5P, GO6P, P6G, PEP, P2G, P3G, BPG, DHAP, GAP, FP2, F6P

Input of enzymes and reaction equations

	E. coli	*T. pallidum*
Pgi: G6*P* = F6P	EC4025	TP0475
Pfk: F6P + ATP → ADP + FP2	EC3916 (PfkA) EC1723 (PfkB)	TP0108 (PfkA) TP0542 (PfkA, second copy)
Fbp: FP2 → F6P + Pi	EC4232	-
Fba: FP2 = GAP + DHAP	EC2925	TP0662
TpiA: GA*P* = DHAP	EC3919	TP0537
Gap: GAP + Pi + NAD = NADH + BPG	EC1779	TP0844
Pgk: BPG + AD*P* = ATP + P3G	EC2926	TP0538
Gpm: P3G = P2G	EC0755	TP0168
Eno: P2G = PEP	EC2779	TP0817
Pyk: PEP + ADP → ATP + Pyr	EC1854	-
Zwf: G6P + NAD*P* = GO6P + NADPH	EC1852	TP0478
Pgl: GO6P – P6G	EC0766	-
Gnd: P6G + NAD*P* = NADPH + CO2 + Ru5P	EC2029	TP0331
Rpi: Ru5*P* = R5P	EC2914	TP0616
Rpe: Ru5*P* = Xyl5P	EC3386	TP0945
TktI: R5P + Xyl5*P* = GAP + Sed7P	EC2935 (TktA) EC2465 (TktB)	TP0560
Tal: GAP + Sed7*P* = Ery4P + F6P	EC2464	-
TktII: Xyl5P + Ery4*P* = F6P + GAP	EC2935 (TktA) EC2465 (TktB)	TP0560
Prs_DeoB: R5P → R5Pex	EC4383	-
Not shown in Fig. 1:		
Pps: Pyr + AT*P* = PEP + AMP + Pi	EC1702	-
Ppd: Pyr + ATP + Pi = PEP + AMP + PPi	-	TP0746

[a]Abbreviations of enzymes: Eno, enolase; Fba, fructose 1,6-bisphosphate aldolase; Fbp, fructose 1,6-bisphosphatase; Gap, glyceraldehyde 3-phosphate dehydrogenase; Gnd, phosphogluconate dehydrogenase (decarboxylating); Gpm, phosphoglycerate mutase; Pfk, 6-phosphofructokinase; Pgi, phosphoglucoisomerase; Pgk, phosphoglycerate kinase; Pgl, phosphogluconolactonase; Ppd, pyruvate, orthophosphate dikinase; Pps, pyruvate, water dikinase (phosphoenolpyruvate synthase); Pyk, pyruvate kinase; Rpe, ribulose-phosphate 3-epimerase; Rpi, ribose 5-phosphate isomerase; Tal, transaldolase; TktI and II, two functions of transketolase; TpiA, triosephosphate isomerase; Zwf, glucose 6-phosphate dehydrogenase. The symbol Prs_DeoB stands for the enzymes catalysing the first step in the conversion of R5P to "external" ribo- and deoxyribonucleotides. When specifying these enzymes to be 5-phosphoribosyl-1-pyrophosphate synthetase (Prs), phosphopentomutase (DeoB) or others, one has to consider the possible additional consumption of ATP. Several enzymes have isoenzymes with different genome identifiers; these are shown, for illustration, only for Pfk and Tkt.
Abbreviations of metabolites have their meaning usual in biochemistry, for example, G6P for glucose-6-phosphate. 2PG, 3PG, 6PG and 1,3BPG in the diagram are designated P2G, P3G, P6G and BPG above because the programs do not accept metabolite names with leading numerals. Water is not shown as a reactant. Reversible and irreversible reactions are indicated, in the reaction equations, by the symbols = and →, respectively.

and, moreover, it shows that there is an additional route by which G6P can be converted into pyruvate and CO_2 (mode 2). It uses Pgi in the reverse direction instead of Pfk, Fba, and TpiA.

Our method correctly yields the glycolytic pathway and the fructose-1,6-bisphosphate cycle encompassed in the system. The latter is an example of a cycle that has the potential to operate, but whether it actually does is a question of regulation[21,22], whether by allosteric effectors, covalent modification, or differential gene expression. In mammalian liver cells, for example, the recycling between F6P and FP_2 is controlled by an incomplete on/off mechanism[21]. During fasting, Pfk is inactive,

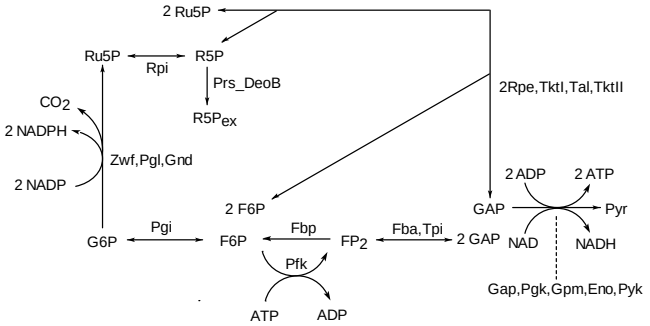

Figure 3. Reduced scheme obtained from the scheme shown in Figure 1 by combining all enzymes that always operate together.

whereas in the fed state, as much as 30% of FP_2 formed by Pfk is converted back to glucose. Such cycles deserve particular extraction because it is of interest to know the percentage of cellular ATP production lost by futile cycling[2,14,21] and, in view of the above-mentioned regulatory properties, to identify potential allosteric enzymes.

It appears to be a matter of taste whether futile cycles can be considered as pathways. Both from the mathematical viewpoint adopted here and in terms of transformation of a substrate into a product (here the "futile" transformation of ATP into ADP), they are pathways. Note that they essentially differ from a cycle made up of the two oppositely directed steps in a reversible reaction because they imply a net transformation and, in physical terms, production of entropy. Conversion of excess R5P to glycolytic intermediates can be modeled by considering the reaction Prs_DeoB (R5P consumption) to be reversible. This gives rise to the same elementary modes as before plus another six modes (modes 8–13 in Table 2).

The function of the nonoxidative part of the PPP is often stated as the transformation of six molecules of pentose into five molecules of hexose[23,24]. This is performed by mode 12. It is the reverse of mode 5, except for the use of Fbp instead of Pfk. Mode 13 realizes an incomplete transformation of riboses into hexoses, where two carbons per molecule of ribose end up in CO_2. Another way to transform R5P to G6P produces pyruvate in addition (mode 8). Modes 9 and 10 are two possibilities for degrading excess R5P. This function of the PPP has been mentioned in biochemistry textbooks[23,25], but no stoichiometry has been given. It is likely to be important, for example, for the digestion of the nucleotides present in the food of carnivores and for microorganisms genetically engineered so as to be able to ferment xylose or other pentoses[26]. The difference between the two modes is that mode 9 produces less ATP than mode 10, but it produces NADPH in addition. In mode 11, both R5P and G6P are consumed to produce pyruvate.

How do the elementary modes behave if additional reactions are included? It can be shown in a general way that the following invariance property holds. For two systems differing only in that the second system involves some additional reactions, all elementary modes of the first system are retained in the second one, which may involve additional modes. An example is provided by the inclusion of Pps in Table 1. When this (reversible) enzyme is considered in addition to the other *E. coli* enzymes listed in Table 1, all the 13 modes given in Table 2 plus eight additional modes are obtained. These eight modes comprise seven modes that coincide with the seven modes producing pyruvate except that Pyk is replaced by Pps in the reverse direction and ADP is, in the overall stoichiometry, partially replaced by AMP plus inorganic phosphate. The eighth additional mode is the futile cycle made up of Pyk and Pps. This invariance only holds, however, if the status of the metabolites (internal or external) is not changed. If it is, the modes may change in a complicated way. In simple systems, however, the number of modes sometimes does not increase but some of them are increased in length. For example, if in an unbranched reaction chain, a reac-

tion removing the product is added and the former product is now considered as an internal metabolite, the mode representing this pathway is simply enlarged by one reaction.

Now we consider *T. pallidum*, the causative agent for syphilis, in which counterparts of many of the *E. coli* enzymes have not been identified (see Table 1). Taking into account that pyruvate, orthophosphate dikinase (Ppd) has been identified in this microorganism, we obtain the elementary mode (Pgi, Pfk, Fba, TpiA, 2Gap, 2Pgk, 2Gpm, 2Eno, –2Ppd) with the overall stoichiometry G6P + 2 AMP + ADP + 2 PP_i + 2 NAD → 3 ATP + 2 NADH + 2 Pyr. As the enzymes Fbp, Pyk, Pgl, Tal, Prs, and DeoB have not yet been identified in this parasite and, in each mode listed in Table 2, at least one of these enzymes is involved, even though some enzymes not included in the one mode given above have been identified, we can deduce that the enzyme set is incomplete and that there must be relevant enzyme activities among the unassigned open reading frames. Some enzymes, or nonorthologous proteins performing the same function, may be identified later by bioinformatics or biochemical methods. Elementary modes analysis can complement these methods[8]. For example, it follows from our analysis that the presence of Tkt makes the presence of some sort of transaldolase very likely, as does the presence of Zwf and Gnd for some sort of phosphogluconolactonase. Another possibility is that similar reactions with different stoichiometries are used (such as the use of Ppd instead of Pyk). An analysis of the modes around glycolysis in *Mycoplasma hominis* and *Methanococcus jannaschii* has been presented earlier[8].

Analysis of another metabolic net, say for the synthesis of a β-lactam antibiotic, would first establish as input (similarly to Table 1) the list of the enzymes present in the organism and the directionality of their reactions. The list would start with the synthesis of the precursor amino acids from the carbon and nitrogen source, through to the nonribosomal condensation of aminoadipic acid, cysteine, and valine, and continuing to the later steps[27] (e.g., lysine aminotransferase and piperidine 6-carboxylic acid dehydrogenase for cephamycin production). Any intended additions or deletions by genetic engineering are included in the list (for instance expressing some of the recently described enzymes involved in methoxylation[27]). The list can then be directly used by our program to calculate the elementary flux modes, producing output similar to that in Table 2. Any observed state of the net, such as the conversion of nutrients into synthesized peptide antibiotic, can only be a linear combination of some number (not necessarily all) of the elementary flux modes calculated, no matter what different growth conditions are used. Care has to be taken if different growth conditions change the directionality of some reactions. In this case and other situations of uncertainty about irreversibility of a reaction, it is always a good option to treat the reaction as reversible. Considering an irreversible reaction reversible can only increase the set of modes, as can be seen in the glycolysis/PPP example with Prs_DeoB. If some enzymes are not expressed in different growth conditions, elementary modes using them will be suppressed. The option obviously exists of calculating the modes only of those enzymes known to be active or expressed under the circumstances.

Knowledge of the elementary modes is useful for determining maximal conversion yields[17,26]. Since any flux pattern is a superposition of elementary modes with non-negative coefficients, the product:substrate ratio (molar yield) of the flux pattern is a weighted average of the yields of the elementary modes involved. For example, consider a pattern that can be decomposed as $\alpha e_1+\beta e_2$ with e_1 and e_2 denoting two modes and α and β denoting positive coefficients. Then, the yield, η, of the flux distribution obeys the relation

$$\eta = (\alpha\eta_1 + \beta\eta_2)/(\alpha + \beta)$$

Since η cannot be greater than the larger of η_1 and η_2, the optimal flux distribution with respect to maximizing the yield must always coincide with an elementary mode. A comparison of this method with the method of linear programming has been given earlier[17].

In agreement with common biochemical knowledge, in the combined glycolysis and PPP system, the best ATP to G6P yield is realized by the glycolytic pathway (mode 1), whereas the best NADPH to G6P yield is achieved by the pentose phosphate cycle (mode 6). In more complex systems, the optimal solution is not so obvious. Recently, we analyzed tryptophan synthesis from glucose and ammonia[17]. In a system comprising 65 reactions, we obtained 26 elementary modes. The two most efficient modes form 105 moles of tryptophan from 233 moles of glucose, giving the theoretical limits for a wide range of conditions investigated experimentally (reviewed in ref. 28).

Discussion

The present analysis sheds new light on the fundamental question of how to define the concept of metabolic pathway[23,25,29,30]. The tricarboxylic acid (TCA) cycle, for example, is usually considered as a biochemical pathway. However, the full TCA cycle is, in *E. coli* and many other bacteria, seen only during aerobic growth on acetate or fatty acids[31]. Instead, under anaerobiosis, its enzymes function as two distinct biosynthetic pathways producing succinyl-CoA and α-ketoglutarate. There is a case for not attempting to decompose metabolism into pathways involving distinct sets of enzymes. Rather, it should be acknowledged that pathways overlap. Modes 2–4, for example, in the

Table 2. Elementary modes of the combined glycolysis and PPP system[a]

Mode	Overall conversion	Participating steps
(1)	G6P + 3 ADP + 2 Pi + 2 NAD → 3 ATP + 2 NADH + 2 Pyr	{Pgi Pfk Fba TpiA 2Gap 2Pgk 2Gpm 2Eno 2Pyk}
(2)	G6P + 2 ADP + Pi + NAD + 6 NADP → 2 ATP + NADH + 6 NADPH + 3 CO2 + Pyr	{-2Pgi Gap 3Zwf 3Pgl 3Gnd Rpi 2Rpe TktI Tal TktII Pgk Gpm Eno Pyk}
(3)	3 G6P + 8 ADP + 5 Pi + 5 NAD + 6 NADP → 8 ATP + 5 NADH + 6 NADPH + 3 CO2 + 5 Pyr	{2Pfk 2Fba 2TpiA 5Gap 3Zwf 3Pgl 3Gnd Rpi 2Rpe TktI Tal TktII 5Pgk 5Gpm 5Eno 5Pyk}
(4)	G6P + 2 NADP → 2 NADPH + CO2 + R5Pex	{Zwf Pgl Gnd Rpi Prs_DeoB}
(5)	5 G6P + ATP → ADP + 6 R5Pex	{5Pgi Pfk Fba TpiA 4Rpi -4Rpe -2TktI -2Tal -2TktII 6Prs_DeoB}
(6)	G6P + 12 NADP → 12 NADPH + Pi + 6 CO2	{-5Pgi -Fba -TpiA 6Zwf 6Pgl 6Gnd 2Rpi 4Rpe 2TktI 2Tal 2TktII Fbp}
(7)	ATP → ADP + Pi	{Pfk Fbp}
(8)	2 ADP + Pi + NAD + 3 R5Pex → 2 ATP + NADH + 2 G6P + Pyr	{-2Pgi Gap -2Rpi 2Rpe TktI Tal TktII Pgk Gpm Eno Pyk -3Prs_DeoB}
(9)	2 ADP + Pi + NAD + 4 NADP + R5Pex → 2 CO2 + 2 ATP + NADH + 4 NADPH + Pyr	{-2Pgi Gap 2Zwf 2Pgl 2Gnd 2Rpe TktI Tal TktII Pgk Gpm Eno Pyk -Prs_DeoB}
(10)	8 ADP + 5 Pi + 5 NAD + 3 R5Pex → 8 ATP + 5 NADH + 5 Pyr	{2Pfk 2Fba 2TpiA 5Gap -2Rpi 2Rpe TktI Tal TktII 5Pgk 5Gpm 5Eno 5Pyk -3Prs_DeoB }
(11)	8 ADP + 5 Pi + 5 NAD + 4 NADP + 2 G6P + R5Pex → 2 CO2 + 8 ATP + 5 NADH + 4 NADPH + 5 Pyr	{2Pfk 2Fba 2TpiA 5Gap 2Zwf 2Pgl 2Gnd 2Rpe TktI Tal TktII 5Pgk 5Gpm 5Eno 5Pyk -Prs_DeoB}
(12)	6 R5Pex → 5 G6P + Pi	{-5Pgi Fbp -Fba -TpiA -4Rpi 4Rpe 2TktI 2Tal 2TktII -6Prs_DeoB}
(13)	8 NADP + 2 R5Pex → 4 CO2 +8 NADPH + G6P + Pi	{-5Pgi Fbp -Fba -TpiA 4Zwf 4Pgl 4Gnd 4Rpe 2TktI 2Tal 2TktII - Prs_DeoB}

[a]The enzymes involved in the elementary modes are given in braces, weighted with their fractional flux. Negative values indicate the reaction is used in the reverse sense. When the reaction Prs_DeoB is considered irreversible, only the upper seven modes are obtained.

system studied above would then be considered as equally entitled to pathway status as modes 1 or 5, for example, rather than being combinations of other pathways (glycolysis and the PPP).

Selkov and coworkers[29] define a pathway as a set of oriented reactions interacting under given physiological conditions via simple or apparently simple intermediates, with a compound considered as a simple intermediate if there is a single pair of production and consumption reactions determining its concentration. In Voet and Voet[25], metabolic pathways are defined in a similar way—as a "series of consecutive enzymatic reactions that produce specific products." Although this applies in many cases, there are counter-examples. In glycolysis, for instance, glyceraldehyde-3-phosphate is produced both by Fba and TpiA, so that it is not a "simple" intermediate in the above sense. In the PPP, not only consecutive enzymes, but also the parallel reactions of Rpi and Rpe are needed to guarantee functioning in modes 2, 3, and 6 (so that Ru5P is not a "simple" intermediate), while in mode 5, these two enzymes can indeed be considered sequential. These examples show that to find nonredundant steps in a metabolic network (which is important in the identification of enzymes to target in order to suppress pathogenic agents), it is not sufficient simply to look for metabolites that are produced by only one reaction. The significant innovation in elementary modes is to introduce the requirement of simplicity implied by nondecomposability. For example, both the modes 1 and 2 produce pyruvate from G6P and, hence, so does their union. But this complex set of enzymes is certainly not a pathway because it is decomposable into two simpler, functional sets.

In interpreting elementary modes, one should keep in mind that they represent idealized situations. They are the simplest flux distributions the cell can realize and are invariant to changes in the properties of regulation (e.g., feedback loops). Sometimes, the metabolic state will approach a single mode closely. For example, glycolysis is much more active than the PPP enzymes in mature yeast cells producing ethanol[6], whereas mode 4 in the above example corresponds approximately to the flux distribution found experimentally in hybridoma and tumor cells[32,33], in which the measured flux through PgiA is very small. These cells need both NADPH and ribose for biosynthetic purposes because of their rapid growth and proliferation. Controlling the intake of thiamine (the cofactor of Tkt) or treating cancer patients with antithiamine analogs has therefore been proposed[33]. Mode 6 is mainly used in cells with a high demand for NADPH for the biosynthesis of fatty acids and steroids, such as in adipose tissue and mammary gland[23,34].

Another example is the human parasite *Trypanosoma brucei*, for which we determined the elementary modes on the basis of a model developed by Bakker and colleagues[35] for drug target selection. All three modes obtained are in agreement with experimental observations: degradation of glucose into pyruvate under aerobic conditions[36], conversion of glucose into equimolar amounts of pyruvate and glycerol under anaerobic conditions[36], and uptake of glycerol and conversion into pyruvate under inhibition of glucose catabolism[37]. It was shown experimentally that at in vivo oxygen tensions, the first mode is operative almost exclusively, while by decreasing oxygen tension, the first mode can gradually be transformed into the second mode via flux distributions representing mixtures of these modes[36]. In many cell types, however, demands are mixed so that the flux pattern is a superposition of several elementary modes[15,16], such as in riboflavin-producing *Bacillus subtilis*[2]. Nevertheless, kinetic factors and regulatory structure can only select from the possibilities described by the modes, so it is a useful conceptual approach to consider these ideal, extreme situations, of which all possible real states are composed. Since the percentage contribution of each mode is determined by kinetics and regulation, it can be changed by genetic engineering.

Elementary modes analysis has been applied to the computation of maximal conversion yields[17,26]. Importantly, the enzyme set used as input may comprise additional enzymes to be introduced by genetic manipulation. If several or all the enzymes corresponding to a particular mode were amplified, that mode could be scaled up to dominate the flux pattern. Thus, an elementary mode identifies a set of enzymes and relative activities for the 'universal method' developed by Kacser and Acerenza[38], although to complete their implementation, the initial flux distribution would also be needed. Alternatively, all other modes could be suppressed by inhibiting several enzymes that are not used by the best mode. If insertion of any particular enzyme were an obstacle, modes not requiring it could be calculated. In the example based on tryptophan synthesis, Zwf is required in one of the optimal modes but another mode with the same tryptophan yield on glucose (0.451) can be obtained without this enzyme, as well as several other modes with only slightly smaller yield. The next optimal mode allows a yield of 0.449. In general, our method allows one to detect suboptimal solutions, which might sometimes be easier to realize biotechnologically (e.g., smaller number of enzymes or cheaper starting material).

Liao and colleagues[26] have successfully constructed an *E. coli* strain with an improved yield of DAHP, a precursor for aromatic amino acid synthesis from carbohydrate, and have shown that their results are consistent with an elementary modes analysis. This illustrates application of elementary modes to studying the optimization of yields in biotechnology. Recently, the method has also been used in the detection and analysis of the routes of sugar metabolism in sugarcane plants[39]. There are five cyclic modes and nine productive modes, three of which perform glucose degradation and six realize sucrose accumulation. Modification of growth media in order to maximize conversion ratios is also of interest: the utilization of different carbon or nitrogen sources can be assessed by specifying different external metabolites. Moreover, it has been suggested[40] that detection of elementary modes could be applicable in flux-balance analysis. A similar approach has also been applied to elucidating the evolution of metabolic networks[41].

The method is also applicable to drug target identification. Given a relatively well-annotated parasite genome such as that of *Mycobacterium tuberculosis*[7], the metabolic routes remaining after blocking of a drug target can be calculated. This allows one to identify the most vulnerable sites in the metabolic network where few or no metabolic side routes are available and furthermore to analyze the effect of drug combinations, for instance, in the treatment of tuberculosis[7]. Another example is from *T. brucei* (example above). One of the modes calculated shows that consumption of glycerol might provide an escape from therapies against the parasite's glucose catabolism. Moreover, the flux modes can be determined that are enhanced when an enzyme activity increases during development of (multi-)drug resistance and hence the appropriate targets that remain.

In a biotechnological context, the approach does allow examination of whether deletion of a particular gene could prevent a metabolic network from continuing to function. This is also related to the enzyme deficiencies important in medicine. For example, glucose-6-phosphate dehydrogenase (Zwf) defects (favism) belong to the most widespread enzymopathies in humans and cause side effects in drug therapy[42]. From Table 2, it can be seen that modes 2, 3, 4, 6, 9, 11, and 13 are inoperative if Zwf is completely deficient. Pandolfi and coworkers[43] noted that, having knocked out Zwf in mouse cells, there was still a pathway leading from glucose to ribose via the Tkt and Tal reactions (Mode 5, Table 2). A related application is in guiding metabolic reconstruction[4]. In addition to the earlier example (*T. pallidum*), some assertions in the literature about possible bypasses of enzyme deficiencies, for example in *Mycoplasma hominis*, which probably lacks Pfk, Fba, and Zwf, have been shown to be incorrect by our analysis[17]. An example where the proposal can be shown to be tenable is provided by a mode bypassing fumarase via aspartate aminotransferase and aspartase in *Haemophilus influenzae*[3].

As mentioned, an alternative proposal has been to use a set of

linearly independent basis vectors for defining biochemical pathways[5,11,12]. However, this method is less powerful than elementary modes in predicting the capabilities of metabolic genotypes because of the nonuniqueness of the basis vectors (even after transforming them into vectors complying with the irreversibility constraints[5]) and because it is not restricted to non-negative linear combination. Moreover, it is less suitable for predicting the effects of enzyme deficiencies or knockout mutations. A chromaffin cell might contain, for example, among others, the enzymes phenylalanine 4-monooxygenase (Ph4H), tyrosine aminotransferase (AttY), and tyrosine 3-monooxygenase (Ty3H). A valid basis is given by (Ph4H, AttY), converting phenylalanine first to tyrosine and then degrading it to 4-hydroxyphenylpyruvate, and (−AttY, Ty3H), which transforms 4-hydroxyphenylpyruvate via tyrosine to L-dihydroxyphenyl-alanine (L-DOPA). Both of these basis vectors appear to drop out in a system deficient in AttY, for instance because of drug inhibition of this enzyme. However, the mode (Ph4H, Ty3H) would still operate in the system, although it was not in the basis because it is the sum of the two basis vectors. This has the biochemical consequence that the synthesis of L-DOPA from phenylalanine nevertheless continues. In larger networks, where automated searching would be essential, this example points to phenotypes correctly predicted by elementary mode analysis, but not by basis vector approaches. When no elementary mode exists without a given enzyme, there really is no possible flux distribution in the system deficient in this enzyme. This is also relevant for the interpretation of missing or unidentified genes in a genome.

One of the great current challenges in bioinformatics is to correlate the simple linear world of nucleotide sequences with the nonlinear world of cellular function. In order for these correlations to be performed in a systematic, ultimately automated way, it will be necessary for cellular function to be describable in a more analytical manner. Elementary mode analysis clarifies the textbook notion of biochemical pathway, applying a stringent mathematical apparatus. The calculated modes reveal an important basis for the surprising adaptability and robustness[30] of metabolic networks and allow at the same time a more specific targeting and modification of pathways. For biochemical function, we suggest that elementary modes represent a more systematic and analytical definition of a biochemical pathway than has previously existed.

Computational protocol

The elementary modes for reaction systems of any complexity can be detected by an algorithm to be outlined and illustrated in the following. A more mathematical description as well as references to related methods have been given earlier[15,16]. Even if all reactions are irreversible, this algorithm is faster than the method proposed by Clarke[13] (for a comparison, see ref. 19). The algorithm has been implemented as computer programs in Smalltalk (program EMPATH, John Woods, Oxford) and C (program METATOOL, Pfeiffer et al.[44]). Both are available from ftp://bmshuxley.brookes.ac.uk/pub/mca/software/ibmpc. The programs start from a list of reaction equations and a declaration of reversible and irreversible reactions and of internal and external metabolites (see Table 1). As in several other metabolic simulators, this list is automatically translated into a stoichiometry matrix (for an explanation of this and related terms, see http://www.biologie.huberlin.de/biophysics/Theory/tpfeiffer/metatool.html). This matrix is then transposed and augmented with the identity matrix, to give a matrix called the initial tableau. From this, further tableaux are consecutively computed by pairwise linear combination of rows so that the columns of the transposed stoichiometry matrix become null vectors successively. This procedure corresponds to ensuring the steady-state constraint is satisfied for each metabolite taken in turn.

Consider, for example, the reaction scheme shown in Fig. 1. It includes 15 internal metabolites and 19 reactions. Accordingly, its stoichiometry matrix has dimension 15×19. Before computing the elementary modes, it is convenient (but not necessary) to reduce this matrix by lumping those reactions that necessarily operate together. In the considered system, it can easily be seen that the sets {Gap, Pgk, Gpm, Eno, Pyk} and {Zwf, Pgl, Gnd} constitute such sequences. An algorithm for detecting such sets of enzymes in systems of any complexity has been developed and included into the program METATOOL[44]. Applying this algorithm to the system considered reveals another two sequences: {Fba, TpiA} and {2 Rpe, TktI, Tal, TktII}. The factor 2 means that the flux through Rpe is, in any steady state of the system, twice as large as the flux through TktI, Tal and TktII. Lumping the reactions in any one sequence gives the reduced system shown in Fig. 3. It encompasses fewer metabolites than the original system because the metabolites situated within a reaction sequence can be omitted. The initial tableau of the reduced system reads

(1)

$$\mathbf{T}^{(i)} = \left(\begin{array}{ccccc|ccccc} 0 & 0 & 1 & 0 & 0 & 1 & 0 & \cdots & 0 & \\ 0 & -1 & 0 & 2 & 0 & 0 & 1 & \cdots & 0 & \text{reversible} \\ -1 & 0 & 0 & 0 & 1 & 0 & 0 & \cdots & 0 & \text{reactions} \\ -2 & 0 & 2 & 1 & -1 & 0 & 0 & \cdots & 0 & \\ \hline 0 & 0 & 0 & -1 & 0 & 0 & 0 & \cdots & 0 & \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & \cdots & 0 & \text{irreversible} \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & \text{reactions} \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & \cdots & 1 & \end{array}\right)$$

The five columns on the left-hand side correspond to the metabolites Ru5P, FP_2, F6P, GAP, and R5P. The nine rows correspond to the reactions or reaction sequences Pgi, {Fba, TpiA}, Rpi, {2Rpe, TktI, Tal, TktII}, {Gap, Pgk, Gpm, Eno, Pyk}, {Zwf, Pgl, Gnd}, Pfk, Fbp, and Prs_DeoB, of which the latter five are irreversible (lower part of the matrix).

The entries with row numbers 1, 2, 5, 7, 8, and 9 in the first column are zeros. Hence, it is not necessary to combine them with other rows. Instead, they can be transferred into the "reversible" and "irreversible" parts of the next tableau. In addition, a "reversible" row in the next tableau arises from subtracting the fourth row from twice the third row. Moreover, appropriate linear combinations of the third and sixth rows and of the fourth and sixth rows yield "irreversible" rows. In general, linear combinations of two rows corresponding to the same type of directionality (reversible or irreversible) go into the part of the respective type in the new tableau, whereas linear combinations of rows corresponding to different types go into the "irreversible" part because they contain at least one irreversible reaction. Care must be taken that "irreversible" rows can enter a linear combination only with a positive coefficient in order that all modes use the irreversible reactions in the right direction.

For the system under consideration, we obtain the following tableau $\mathbf{T}^{(1)}$:

(2)

$$\mathbf{T}^{(1)} = \left(\begin{array}{ccccc|cccccccc} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & -1 & 3 & 0 & 0 & 2 & -1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 2 & 0 & 0 \end{array}\right)$$

The right-hand side part of the tableau keeps track of the linear combinations performed.

Now we perform linear combination of rows to ensure that the second column of the result will be zero. The rows numbered 1, 3, 4, 7, 8, and 9 go straight into the next tableau because their respective second elements are zero already. The next tableau reads

(3)

$$\mathbf{T}^{(2)} = \left(\begin{array}{ccccc|cccccccc} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & -1 & 3 & 0 & 0 & 2 & -1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & 2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -2 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right)$$

In the course of the algorithm, calculation of nonelementary modes, duplicate modes, and flux modes violating the sign restriction for the irreversible reactions must be avoided. To this end, three conditions are checked. First, a pair of rows is combined only if

(4) $\quad S(m_i^{(j)}) \cap S(m_k^{(j)}) \not\subset S(m_j^{(j+1)})$,

for all row indices *l* belonging to the respective part (reversible or irreversible) of the new tableau as it has been compiled until that moment. denotes the *i*th row in the right-hand side submatrix of tableau $\mathbf{T}^{(j)}$ and $S(m_i^{(j)})$ is the set of positions of zeroes in this row. This set contains information about which enzymes are not used in a mode under consideration. For example, tableau $\mathbf{T}^{(3)}$ contains, in its right-hand side part, the row

(5) $\quad m_i^{(3)} = (0\,0\,1\,0\,0\,1\,0\,0\,0)$

as this row comes straight from tableau $\mathbf{T}^{(2)}$. For this row,

(6) $\quad S(m_i^{(3)}) = \{1, 2, 4, 5, 7, 8, 9\}$.

When constructing $\mathbf{T}^{(3)}$, a candidate pair for linear combination comprises the second and sixth rows of $\mathbf{T}^{(2)}$. However, as $S(m_2^{(2)}) \cap S(m_6^{(2)}) = \{1, 2, 5, 7, 8, 9\}$, which is a subset of the set given in Eq. (6), these rows must not be combined. If we did combine them, we would obtain a row that, after scaling, equals the row given in Eq. (5). Linear combination of the seventh and eighth rows is forbidden for a similar reason. An example where condition (4) prevents nonelementary modes from being calculated occurs in the construction of $\mathbf{T}^{(4)}$. For the rows (0 0 0 -1 3 | 2 0 2 -1 0 0 0 0 0) and (0 0 0 5 -1 | 0 2 0 1 0 2 2 0 0) situated in $\mathbf{T}^{(3)}$, we have $S(m_i^{(3)}) \cap S(m_k^{(3)}) = \{5, 8, 9\}$ and a row transferred earlier into $\mathbf{T}^{(4)}$ reads (0 0 0 0 6 | 5 1 4 -2 0 0 1 0 0). If we did combine the two rows, we would obtain (0 0 0 0 7 | 5 1 5 -2 0 1 1 0 0), which comprises fewer zeros than the row already situated in $\mathbf{T}^{(4)}$.

The second condition says that "irreversible" rows can only be added rather than subtracted (cf. above). For example, rows numbered 6 and 8 in $\mathbf{T}^{(2)}$ must not be combined because they both comprise some entries belonging to irreversible reactions and contain positive entries in the third position. For the system considered, these two conditions are relevant only from tableau $\mathbf{T}^{(3)}$ on. In total, seven linear combinations are allowed, giving rise to a tableau $\mathbf{T}^{(3)}$ comprising 11 rows.

It may occur that upon constructing a new tableau, some row that has been correctly computed turns out to be nonelementary because some other row, which is computed later, comprises more zero positions. This can be avoided by using a third test criterion. If any pair of rows pass condition (4) and are combined and added to the tableau, all the rows $m_i^{(j+1)}$ previously added to the new tableau are checked to ensure that:

(7) $\quad S(m_i^{(j+1)}) \cap S(m_i^{(j)}) \notin S(m_k^{(j)})$.

According to our experience, this criterion is rarely violated. To save computational time, it is therefore sufficient to apply it only upon computing the final tableau. For the glycolysis/PPP example, criterion (7) is never violated, whereas condition (4) helps us avoid computing 11 and 9 irrelevant rows in $\mathbf{T}^{(4)}$ and $\mathbf{T}^{(5)}$, respectively. Finally, we obtain

(8)

$$\mathbf{T}^{(6)} = \left(\begin{array}{ccccc|cccccccc} 0 & \cdots & & \cdots & 0 & 1 & 1 & 0 & 0 & 2 & 0 & 1 & 0 & 0 \\ \vdots & & & & \vdots & -2 & 0 & 1 & 1 & 1 & 3 & 0 & 0 & 0 \\ & & & & & 0 & 2 & 1 & 1 & 5 & 3 & 2 & 0 & 0 \\ & & & & & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ & & & & & 5 & 1 & 4 & -2 & 0 & 0 & 1 & 0 & 6 \\ \vdots & & & & \vdots & -5 & -1 & 2 & 2 & 0 & 6 & 0 & 1 & 0 \\ 0 & \cdots & & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right)$$

This example shows that the number of rows may decrease or increase in the course of the algorithm. The rows in the right-hand side submatrix of the final tableau represent the elementary modes, all of which are irreversible. Note that some of their entries correspond to lumped reaction sequences. These modes can, however, be easily translated into the modes in terms of the original set of 19 reactions, as given in the upper part of Table 2. The output of the above-mentioned computer programs has a form similar to Table 2, rather than the abstract form of a matrix. The running time of the programs is less than 1 s for the system considered on a usual PC or workstation.

Acknowledgments

The authors are indebted to the anonymous referees for very helpful comments. We would like to thank Dr. Peer Bork (Heidelberg) and Thomas Pfeiffer (Berlin) for stimulating discussions and the DFG and BMBF (Germany) for generous support.

1. Yarmush, M.L. & Berthiaume, F. Metabolic engineering and human disease. *Nat. Biotechnol.* **15**, 525–528 (1997).
2. Sauer, U. et al. Metabolic fluxes in riboflavin-producing *Bacillus subtilis*. *Nat. Biotechnol.* **15**, 448–452 (1997).
3. Tatusov, R.L. et al. Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*. *Curr. Biol.* **6**, 279–291 (1996).
4. Bork, P. et al. Predicting function: from genes to genomes and back. *J. Mol. Biol.* **283**, 707–725 (1998).
5. Schilling, C.H. & Palsson, B.O. The underlying pathway structure of biochemical reaction networks. *Proc. Natl. Acad. Sci. USA* **95**, 4193–4198 (1998).
6. DeRisi, J.L., Iyer, V.R. & Brown, P.O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
7. Cole, S.T. et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998)
8. Dandekar, T., Schuster, S., Snel, B., Huynen, M. & Bork P. Pathway alignment: application to the comparative analysis of glycolytic enzymes, *Biochem. J.* **343**, 115–124 (1999).
9. Seressiotis, A. & Bailey, J. E. MPS: an algorithm and data base for metabolic pathway synthesis. *Biotechn. Lett.* **8**, 837–842 (1986).
10. Mavrovouniotis, M.L., Stephanopoulos, G. & Stephanopoulos, G. Computer-aided synthesis of biochemical pathways. *Biotechnol. Bioeng.* **36**, 1119–1132 (1990).
11. Fell, D.A. in *Modern trends in biothermokinetics* (eds Schuster, S., Rigoulet, M., Ouhabi, R. & Mazat, J.-P.) 97–101 (Plenum, New York, NY; 1993).
12. Simpson, T.W., Colón, G.E. & Stephanopoulos, G. Two paradigms of metabolic engineering applied to amino acid biosynthesis. *Biochem. Soc. Trans.* **23**, 381–387 (1995).
13. Clarke, B.L. Complete set of steady states for the general stoichiometric dynamical system. *J. Chem. Phys.* **75**, 4970–4979 (1981).
14. Leiser, J. & Blum, J.J. On the analysis of substrate cycles in large metabolic systems. *Cell Biophys*. **11**, 123–138 (1987).
15. Schuster, S. & Hilgetag, C. On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* **2**, 165–182 (1994).
16. Schuster, S., Hilgetag, C., Woods, J.H. & Fell, D.A. in *Computation in cellular and molecular biological systems* (eds Cuthbertson, R., Holcombe, M. & Paton, R.) 151–165 (World Scientific, Singapore, 1996).
17. Schuster, S., Dandekar, T. & Fell, D. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol*. **17**, 53–60 (1999).
18. Schilling, C.H., Schuster, S., Palsson, B.O. & Heinrich R. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* **15**, 296–303 (1999).
19. Schuster, R. & Schuster, S. Refined algorithm and computer program for calculating all non-negative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed. *Comp. Appl. Biosci.* **9**, 79–85 (1993).
20. Stryer, L. *Biochemistry* (Freeman, New York, NY; 1995).
21. Hers, H.G. & Hue, L. Gluconeogenesis and related aspects of glycolysis. *Annu. Rev. Biochem.* **52**, 617–653 (1983).
22. Fell, D. *Understanding the control of metabolism* (Portland Press, London; 1997).
23. Yudkin, M. & Offord, R. *A guidebook to biochemistry* (Cambridge University Press, Cambridge; 1980).
24. Meléndez-Hevia, E., Waddell, T.G. & Montero, F. Optimization of metabolism: the evolution of metabolic pathways toward simplicity through the game of the pentose phosphate cycle. *J. Theor. Biol.* **166**, 201–220 (1994).
25. Voet, D. & Voet, J.G. *Biochemistry* (John Wiley, New York, NY; 1997).
26. Liao, J.C., Hou, S.-Y. & Chao, Y.-P. Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnol. Bioeng.* **52**, 129–140 (1996).
27. Martin, J.F. New aspects of genes and enzymes for beta-lactam antibiotic biosynthesis. *Appl. Microbiol. Biotechnol.* **50**, 1–15 (1998).
28. Frost, J.W. & Draths, K.M. Biocatalytic syntheses of aromatics from D-glucose: renewable microbial sources of aromatic compounds. *Annu. Rev. Microbiol.* **49**, 557–579 (1995).
29. Selkov, E. Jr., Grechkin, Y., Mikhailova, N. & Selkov, E. MPW: the metabolic pathways database. *Nucleic Acids Res.* **26**, 43–45 (1998).
30. Hartwell, L. A robust view of biochemical pathways. *Nature* **387**, 855–857 (1997).
31. Cronan Jr., J.E. & LaPorte, D. in *Escherichia coli and Salmonella. Cellular and molecular biology*, Vol. I (ed. Neidhardt, F.C.) 206–215 (ASM Press, Washington, DC; 1996).
32. Bonarius, H.P.J. et al. Metabolic flux analysis of hybridoma cells in different culture media using mass balances. *Biotechn. Bioeng.* **50**, 299–318 (1996).
33. Boros, L.G. et al. Nonoxidative pentose phosphate pathways and their direct role in ribose synthesis in tumors: is cancer a disease of cellular glucose metabolism? *Med. Hypoth.* **50**, 55–59 (1998).
34. Smith, E.L. et al. *Principles of biochemistry. General aspects* (McGraw-Hill, New York, NY; 1983).
35. Bakker, B.M., Michels, P.A.M., Opperdoes, F.R. & Westerhoff, H.V. Glycolysis in bloodstream form *Trypanosoma brucei* can be understood in terms of the kinetics of the glycolytic enzymes. *J. Biol. Chem*. **272**, 3207–3215 (1997).
36. Eisenthal, R. & Panes, A. The aerobic/anaerobic transition of glucose metabolism in *Trypanosoma brucei*. *FEBS Lett*. **181**, 23–27 (1985).
37. Kiaira, J.K. & Njogu, M.R. Oligomycin-sensitivity of hexose-sugar catabolism in the bloodstream form of *Trypanosoma brucei brucei*. *Biotechnol. Appl. Biochem*. **20**, 347–356 (1994).
38. Kacser, H. & Acerenza, L. A universal method for achieving increases in metabolite production. *Eur. J. Biochem*. **216**, 361–367 (1993).
39. Rohwer, J.M. & Hofmeyr, J.–H.S. in *Technological and medical implications of metabolic control analysis* (eds Cornish-Bowden, A. & Cárdenas, M.L.) 73–79 (Kluwer Academic Publishers Dordrecht; 2000).
40. Bonarius, H.P.J., Schmid, G. & Tramper, J. Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends Biotechn.* **15**, 308–314 (1997).
41. Nuño, J.C., Sánchez-Valdenebro, I., Pérez-Iratxeta, C., Meléndez-Hevia, E. & Montero, F. Network organization of cell metabolism: monosaccharide interconversion. *Biochem. J.* **324**, 103–111 (1997).
42. Ruwende, C. et al. Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. *Nature* **376**, 246–249 (1995)
43. Pandolfi, P.P. et al. Targeted disruption of the housekeeping gene encoding glucose 6-phosphate dehydrogenase (G6PD): G6PD is dispensable for pentose synthesis but essential for defense against oxidative stress. *EMBO J.* **14**, 5209–5215 (1995).
44. Pfeiffer, T., Sánchez-Valdenebro, I., Nuño, J.C., Montero, F. & Schuster, S. METATOOL: For studying metabolic networks, *Bioinformatics* **15** (1999) 251–257.