Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering

Stefan Schuster, Thomas Dandekar and David A. Fell

Rational metabolic engineering requires powerful theoretical methods such as pathway analysis, in which the topology of metabolic networks is considered. All metabolic capabilities in steady states are composed of elementary flux modes, which are minimal sets of enzymes that can each generate valid steady states. The modes of the fructose-2,6-bisphosphate cycle, the combined tricarboxylic-acid-cycle–glyoxylate-shunt system and tryptophan synthesis are used here for illustration. This approach can be used for many biotechnological applications such as increasing the yield of a product, channelling a product into desired pathways and in functional reconstruction from genomic data.

Genetic manipulation of cells has the potential to enable the enhancement of yield and rate of metabolic production by restructuring or deregulating metabolic networks. This has led to the concept of metabolic engineering within biotechnology[1–4], which includes extending existing pathways to achieve the synthesis of novel products, redirecting metabolite fluxes towards a desired product and accelerating or bypassing steps exerting high flux control (bottlenecks). To achieve such modifications, powerful methods of recombinant DNA technology (including promoter engineering, site-directed mutagenesis, domain swapping, high-copy-number expression vectors and regulatory cassettes) have been developed[5]. However, the term 'metabolic engineering' implies analysing metabolic states and rationally designing and predicting changes, for which appropriate tools are required. Those available include metabolic-flux balancing[4,6–8], metabolic-control theory[9,10] and optimization methods[9,11–13].

In general, there are two main approaches to metabolic engineering. First, some enzyme(s) can be blocked completely or removed by knocking out the corresponding genes, or new enzyme(s) can be inserted. Thus, the topology of the reaction network is changed, and this can be supported by theoretical analysis. Second, enzyme activities can be changed gradually or regulatory interactions can be altered. This requires much more extensive data and sophisticated theoretical tools because complex kinetic information is involved.

The structural (topological) analysis of metabolic pathways only uses information about the stoichiometric structure and the reversibility or irreversibility of reactions. In a sense, it is a more qualitative approach, but is far from trivial. The set of pathways in a biochemistry textbook often do not capture the true behaviour of a metabolic network; the modules of metabolism are much more varied than this. To represent fundamental routes in biochemical networks without preconceptions, three different (but mathematically related) approaches have been proposed: (1) a method for constructing, by computer, simple routes in biochemical systems leading from given substrate(s) to given product(s)[14–16]; (2) the use of a set of linearly independent basis vectors in flux space[17,18]; and (3) the concept of 'elementary flux modes'[19,20] or 'basic reaction modes'[21]. This enables the detection of fundamental steady-state flow situations in which the network under study can operate, including cyclic flows. Any steady-state flux pattern can be expressed as a non-negative linear combination of these modes.

S. Schuster (schuster@bp.biologie.hu-berlin.de) and T. Dandekar (dandekar@embl-heidelberg.de) are at the Department of Bioinformatics, Max Delbrück Centre for Molecular Medicine, D-13125 Berlin-Buch, Germany and at the Biocomputing Unit, EMBL, D-69012 Heidelberg, Germany. D. A. Fell (daf@hms.brookes.ac.uk) is at the School of Biological and Molecular Sciences, Oxford Brookes University, Oxford, UK OX3 0BP.

Theoretical methods

In biochemical modelling, a technical distinction is usually made between external and internal metabolites (see Glossary). The reversibility or irreversibility of enzymatic reactions is known in most cases, even when the kinetic parameters are unknown. (Reversible reactions can be directed in either direction under physiological conditions, such as the reactions shared by glycolysis and gluconeogenesis; irreversible reactions include those of most kinases and phosphatases.) A great many biochemical systems are known to attain stationary states and, even if the system behaves in an oscillatory way, the time average of the reaction rates must still fulfil a balance equation.

A 'flux mode' is defined as a steady-state flux distribution in which the proportions of fluxes are fixed (Fig. 1). For example, x units of flux might go through reactions 1 and 2, while y units are carried by reactions 3 and 4 (with the ratio of x to y specified). The physical units of flux will remain unspecified unless measured flux data are available.

A flux mode is called 'elementary' if it is nondecomposable. For example, the flux mode in Fig. 1 is not elementary, because it can be decomposed into a mode in which reactions 1 and 2 carry x units of flux

 0167-7799/99/$ – see front matter PII: S0167-7799(98)01290-6

Glossary

Elementary mode Minimal set of enzymes that could operate at steady state with all irreversible reactions proceeding in the appropriate direction.
Topology of a metabolic network Set of enzymes and metabolites, their connections and the stoichiometry and directionality of the reactions.
Irreversible reaction A reaction in which the rate of the forward reaction is always (in all physiological conditions) so much higher than the rate of the reverse reaction that the latter is relatively negligible.
Yield of a biotransformation Ratio indicating how many moles of product are produced per mole of substrate used.
External metabolites Metabolites buffered by connection to reservoirs. They can be considered to be sources and sinks (nutrients and waste products, stored or excreted products, or precursors for further transformations).
Internal (balanceable) metabolites Metabolites only participating in reactions of the model. The formation of each is exactly balanced by its consumption (steady-state assumption).
Linear combination (superposition) Sum of several terms, each of which may be weighted by an appropriate coefficient.

each while reactions 3 and 4 are blocked, and a mode in which y units of flux are attached to reactions 3 and 4 while reactions 1 and 2 are 'silent'; these two modes are elementary because they cannot be further decomposed (Box 1). Note that the elementary modes in this example do not overlap, which is of interest for the decomposition of the network into subsystems exerting limited control on each other[9]. In general, however, elementary modes do overlap, as in the examples below. Generally speaking, an elementary flux mode is a minimal set of enzymes that can operate at steady state with all irreversible reactions proceeding in the direction prescribed thermodynamically. The enzymes are weighted by the relative flux that they carry[9,20]; examples of their mathematical properties are presented in Box 2.

An illustrative operational definition can be demonstrated by the following: (1) start from a given

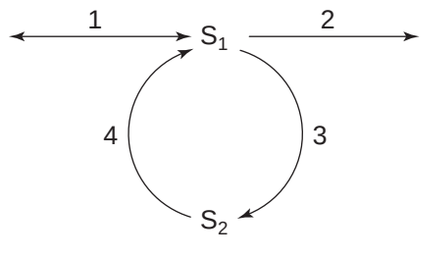

Figure 1
Simple example reaction system representing the fructose-2,6-bisphosphate cycle. Enzymes: 1, phosphoglucoisomerase; 2, phosphofructokinase 1; 3, phosphofructokinase 2; 4, fructose-2,6-bisphosphatase. The reactions of enzymes 2–4 are considered to be irreversible. Metabolites: S_1, fructose-6-phosphate; S_2, fructose-2,6-bisphosphate. ATP, ADP and inorganic phosphate (not shown) are treated as external substances.

Box 1. Calculating elementary modes

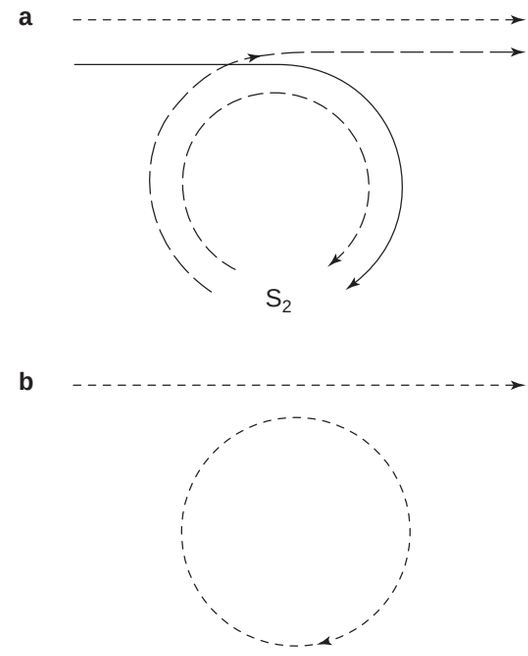

Figure i
Graphical illustration of the calculation of elementary modes for the scheme shown in Fig. 1. (**a**) Preliminary modes after the first step of the algorithm. (**b**) Final result.

Consider a hypothetical situation in which all metabolites are external (i.e. where all concentrations are 'clamped'). In this case, every reaction represents an elementary mode on its own. In each step of the algorithm, preliminary elementary modes have to be linearly combined to give new preliminary elementary modes in the next system, in which another metabolite has been internalized, that is, in which its balance equation at steady state is taken into account. For the example reaction system shown in Fig. 1, in the first step of the algorithm, it is required that metabolite S_1 should be made to fulfil the balance equation at steady state. This condition can be satisfied by many different flux distributions between the four reactions. Because elementary modes are, by definition, as simple as possible, only pairwise combinations of reactions are considered. For example, S_1 can be at steady state when reactions 1 and 3 carry the same flux and reactions 2 and 4 carry no flux at all. In total, four possible combinations (Fig. ia) can be obtained. By internalizing S_2, the final set of two modes shown in Fig. ib is obtained. In this step, it is important not to combine the preliminary mode leading from the input to S_2 with the one leading from S_2 to the output because the resulting mode would involve a loop and, hence, would clearly be non-elementary. The two resulting modes are irreversible; one of these is a futile cycle. In steady state, the irreversible reactions 2–4 force the reversible reaction 1 (phosphoglucoisomerase) to operate in the direction of synthesis of S_1.
A special condition is included in the algorithm to guarantee that non-elementary modes are excluded at each stage, so that a time-consuming, exhaustive search is unnecessary. This condition is, in fact, the crux of the algorithm. A related algorithm for constructing reaction mechanisms[16] involves a different routine, in which pathways are classified according to their length and those pathways that involve subpathways of shorter length are discarded.

metabolic system; (2) block an enzyme by the addition of an excess amount of an enzyme-specific inhibitor; (3) determine whether there is still some flow going through the system; and (4) block a second enzyme, and so on. An elementary mode is reached when the inhibition of a further, still active, enzyme leads to cessation of any steady-state flux in the system. The set of enzymes (or genes) that can be deleted simultaneously has been termed the dispensable set[21].

An algorithm for detecting all elementary modes for systems of arbitrary complexity has been developed recently[20], based on the Gauss–Jordan method for solving linear equation systems but considerably more complex. This algorithm, which only uses the reaction equations and the information about which of these are irreversible (Box 1), has been implemented by computer programs in three different languages: (1) ELMO (C. Hilgetag) in Turbo-Pascal; (2) EMPATH (J. Woods) in Smalltalk; and (3) METATOOL (T. Pfeiffer) in C. The latter two programs are available from ftp://bmsdarwin.brookes.ac.uk/pub/software/ibmpc.

The methods developed by Seressiotis and Bailey[14] and Mavrovouniotis *et al.*[15,16] differ in several aspects from the method based on elementary modes. They start from a distinction between required reactants, allowed reactants (that might or might not be consumed by the pathway), intermediates, required products and allowed (by)products. They are thus more goal-oriented, in that they immediately consider specific biotransformations. By contrast, in the elementary-modes approach, the distinction made is only between intermediates and external substances. Accordingly, it is a more integrative approach to exploring the full potential of all basic biotransformations within the considered system (e.g. a microorganism) without any preconceptions or biases. This enables any steady-state flux pattern to be expressed as a linear combination of these modes. The method can also be used to detect all cyclic pathways in the system without the need for explicitly indicating the external substances driving these cycles.

Box 2. Mathematical implications

In contrast to the basis vectors in linear algebra (i.e. the vectors that span a given space), the elementary mode vectors, $e^{(k)}$, are uniquely determined (up to multiplication by a non-negative real number). Any real flux distribution can be represented as a superposition of these vectors with non-negative coefficients ($\lambda_k \geq 0$) (Eqn i)

$$V = \lambda_1 e^{(1)} + \lambda_2 e^{(2)} + \ldots \quad \text{(i)}$$

where $V = (v_1, v_2, \ldots)$ and is the vector of net reaction rates. For the system shown in Fig. 1, the elementary modes can be written in vector notation as (1 1 0 0) and (0 0 1 1), while any steady-state flux mode has the form (x x y y) with non-negative x and y. This can be written as a linear combination of the two elementary modes. The restriction that the coefficients λ_k be non-negative is necessary and sufficient to guarantee that all linear combinations comply with the sign restriction, $v_i \geq 0$, for the irreversible reactions. However, this implies that, in some situations, more vectors are needed to span the admissible flux region than there are in a set of linearly independent basis vectors[9,19,20].

As flux modes can be arbitrarily scaled, this can be done in such a way that the sum of all coefficients λ_k equals unity. Mathematically, the linear combination in Eqn i is then called a convex combination. As the stoichiometric conversion ratio for any given pair of externals obeys this combination rule as well, the ratio realized by a non-elementary mode cannot be better than that realized by an elementary mode.

The tricarboxylic acid cycle and adjacent reactions

In the reaction scheme of the tricarboxylic acid (TCA) cycle, glyoxylate shunt and adjacent amino acid metabolism (Fig. 2), all cofactors (such as ATP and NAD) are considered as external, as are 2-phosphoglycerate (PG), NH_3 and CO_2 (that is, their concentrations are assumed to be unaffected by the reactions in the scheme). All of the enzymes shown are present in wild-type *Escherichia coli*. By contrast, alanine aminotransferase appears to be absent from *E. coli*; its function is probably carried out by the valine–pyruvate aminotransferase and the branched-chain amino acid aminotransferase. Valine, in turn, gets its amino group from glutamate, again through the branched-chain amino acid aminotransferase. This scheme gives rise to 16 elementary modes (Table 1), all of which are irreversible. The system has been chosen to be fairly simple, to illustrate the concept. A more realistic scheme would include features such as consumption of phosphoenolpyruvate (PEP) by the phosphotransferase system and the distinction between succinate dehydrogenase and fumarate reductase.

Different methods of processing the substrates in this network become apparent from this analysis.

- Mode 15 is the usual (catabolic) tricarboxylic acid cycle.
- Flux-distribution 6 is a combination of the glyoxylate shunt with part of the TCA cycle. The oxaloacetate (OAA) produced is used in equal proportions by citrate synthase and PEP carboxykinase (Pck). This mode has also been found and discussed by Liao *et al.*[21], who considered all the reactions to be irreversible.
- Mode 4 represents alanine formation from glycolytic intermediates and NH_3.
- Mode 5 is a cycle performing a transhydrogenation from NADPH to NADH.
- Modes 3 and 7 perform the synthesis of aspartate with two different yields – one molecule of Asp per molecule of PG used and 0.5 Asp per PG, respectively. The OAA needed for aspartate formation can be regenerated in two different ways: either by the carboxylation of PEP or via the glyoxylate shunt. Mode 3 corresponds to the flux distribution reported for lysine overproduction in *Corynebacterium glutamicum*[1]. Modes 3 and 7 are also equivalent to the pathways obtained by Mavrovouniotis *et al.* for the synthesis of lysine via aspartate[15], apart from the fact that OAA is regenerated by the carboxylation of PEP rather than pyruvate; notice that pyruvate carboxylase is absent from *E. coli*. The additional pathways computed by Mavrovouniotis *et al.*[15] can be obtained by this algorithm when the enzymes lactate–malate transhydrogenase (not present in *E. coli*) and lactate dehydrogenase are included, and aspartase is made reversible.

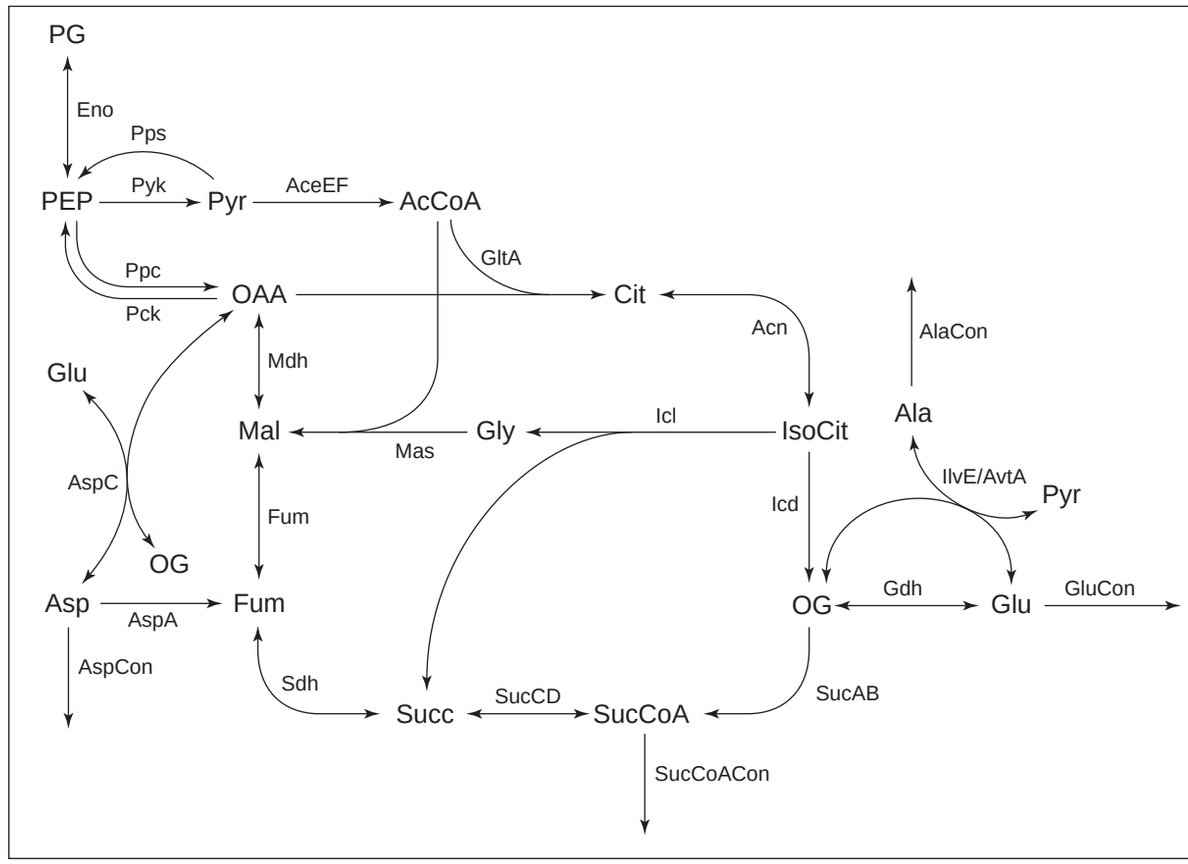

Figure 2

Reaction scheme consisting of the tricarboxylic acid cycle, glyoxylate shunt and some adjacent reactions of amino acid metabolism in *Escherichia coli*. Abbreviations of metabolites: AcCoA, acetyl-CoA; Ala, alanine; Asp, aspartate; Cit, citrate; Fum, fumarate; Glu, glutamate; Gly, glyoxylate; IsoCit, isocitrate; Mal, malate; OAA, oxaloacetate; OG, 2-oxoglutarate; PEP, phosphoenolpyruvate; PG, 2-phosphoglycerate; Pyr, pyruvate; Succ, succinate; SucCoA, succinyl-CoA. Abbreviations of enzymes: AceEF, pyruvate dehydrogenase; Acn, aconitase; AspA, aspartase; AspC, aspartate aminotransferase; Eno, enolase; Fum, fumarase; Gdh, glutamate dehydrogenase; GltA, citrate synthase; Icd, isocitrate dehydrogenase (in *E. coli* with cofactors NADP/NADPH); Icl, isocitrate lyase; Mas, malate synthase; IlvE/AvtA, branched-chain amino acid aminotransferase/valine-pyruvate aminotransferase; Mdh, malate dehydrogenase; Pck, PEP carboxykinase (in *E. coli* with cofactors ADP/ATP); Ppc, PEP carboxylase; Pps, PEP synthetase; Pyk, pyruvate kinase; Sdh, succinate dehydrogenase; SucAB, 2-oxoglutarate dehydrogenase; SucCD, succinyl-CoA synthetase (in *E. coli* with cofactors ADP/ATP); AlaCon, AspCon, GluCon and SucCoACon, consumption of alanine, aspartate, glutamate and succinyl-CoA, respectively. Reversible reactions are indicated by double arrow-heads.

- Modes 12 and 14 represent glutamate synthesis with different molar yields. OAA is regenerated via PEP carboxylase (Ppc) and the glyoxylate shunt, respectively.
- The formation of haem or other substances from succinyl-CoA can occur in six different modes. Three (modes 9, 11 and 16) involve the glyoxylate shunt and the rest use Ppc to replenish the TCA-cycle intermediates. Under anaerobic conditions, the TCA cycle in *E. coli* functions as two anabolic pathways – a reductive pathway that produces succinyl-CoA [via malate (mode 10) or aspartate (mode 8)] and an oxidative pathway (related to mode 13) producing 2-oxoglutarate[7,22]. These pathways do not form a cycle, because 2-oxoglutarate dehydrogenase is severely repressed during anaerobic growth. The superposition of modes 10 and 13 is related to the 'horseshoe' structure proposed for the evolution of the TCA cycle[23].
- Modes 1 and 2 are the futile cycles formed by Pck and Ppc and by pyruvate kinase and PEP synthetase, respectively. In *E. coli* wild-type strains and regulatory mutants, no significant futile cycling is detected, probably because of tight regulation[6,7,24]. However, as demonstrated for the Pck–PEP-carboxylase cycle, substrate cycles can be induced artificially by over-expression of the enzymes involved[24].

Growth on acetate in the absence of glucose can be modelled by considering enolase irreversible in the backward direction and including acetyl-CoA synthetase. In this case, there are 11 elementary modes, one of which has the overall stoichiometry.

$$3\ ATP + FAD + 2\ NAD + 2\ acetate \rightarrow ADP + 2\ AMP + FADH_2 + 2\ NADH + CO_2 + PG$$

and uses, among others, the reactions of isocitrate lyase and malate synthase. As PG can be further converted into glucose, this mode represents gluconeogenesis from acetate. This is in accordance with textbook knowledge stating that the glyoxylate shunt is necessary for growth on acetate.

It is usually believed that isocitrate dehydrogenase (Icd), which is inhibited by phosphorylation, is

Table 1. Elementary modes of the combined TCA cycle and glyoxylate shunt system in the output format of the program EMPATH

Mode	Overall reaction and the enzymes involved
1	$ATP \rightarrow ADP$ {Pck Ppc}
2	$ADP \rightarrow AMP$ {Pyk Pps}
3	$NH_3 + NADPH + CO_2 + PG \rightarrow Asp_{ex} + NADP$ {Eno AspC AspCon Gdh Ppc}
4	$ADP + NH_3 + NADPH + PG \rightarrow Ala_{ex} + ATP + NADP$ {Eno Pyk Gdh IlvE/AvtA AlaCon}
5	$NADPH + NAD \rightarrow NADP + NADH$ {Fum Mdh AspC AspA Gdh}
6	$ADP + FAD + 4\ NAD + PG \rightarrow ATP + FADH_2 + 4\ NADH + 3\ CO_2$ {Eno 2Pyk 2AceEF GltA Acn Sdh Fum 2Mdh Icl Mas Pck}
7	$2\ ADP + NH_3 + FAD + NADPH + 4\ NAD + 2\ PG \rightarrow 2\ ATP + Asp_{ex} + FADH_2 + NADP + 4\ NADH + 2\ CO_2$ {2Eno 2Pyk 2AceEF GltA Acn Sdh Fum 2Mdh Icl Mas AspC AspCon Gdh}
8	$ATP + FADH_2 + NADPH + CO_2 + PG \rightarrow Suc_{ex} + ADP + FAD + NADP$ {Eno −SucCD −Sdh AspC AspA Gdh Ppc SucCoACon}
9	$ADP + 3\ NAD + 2\ PG \rightarrow Suc_{ex} + ATP + 3\ NADH + 2\ CO_2$ {2Eno 2Pyk 2AceEF GltA Acn −SucCD Mdh Icl Mas SucCoACon}
10	$ATP + FADH_2 + NADH + CO_2 + PG \rightarrow Suc_{ex} + ADP + FAD + NAD$ {Eno −SucCD −Sdh −Fum −Mdh Ppc SucCoACon}
11	$FADH_2 + 2\ NAD + 3\ PG \rightarrow 2\ Suc_{ex} + FAD + 2\ NADH + CO_2$ {3Eno 2Pyk 2AceEF GltA Acn −2SucCD −Sdh −Fum Icl Mas Ppc 2SucCoACon}
12	$ADP + NH_3 + NAD + 2\ PG \rightarrow Glu_{ex} + ATP + NADH + CO_2$ {2Eno Pyk AceEF GltA Acn Icd Gdh Ppc GluCon}
13	$ADP + NADP + 2\ NAD + 2\ PG \rightarrow Suc_{ex} + ATP + NADPH + 2\ NADH + 2\ CO_2$ {2Eno Pyk AceEF GltA Acn Icd SucAB Ppc SucCoACon}
14	$3\ ADP + NH_3 + FAD + 5\ NAD + 3\ PG \rightarrow Glu_{ex} + 3\ ATP + FADH_2 + 5\ NADH + 4\ CO_2$ {3Eno 3Pyk 3AceEF 2GltA 2Acn Icd Sdh Fum 2Mdh Icl Mas Gdh GluCon}
15	$2\ ADP + FAD + NADP + 3\ NAD + PG \rightarrow 2\ ATP + FADH_2 + NADPH + 3\ NADH + 3\ CO_2$ {Eno Pyk AceEF GltA Acn Icd SucAB SucCD Sdh Fum Mdh}
16	$3\ ADP + FAD + NADP + 6\ NAD + 3\ PG \rightarrow Suc_{ex} + 3\ ATP + FADH_2 + NADPH + 6\ NADH + 5\ CO_2$ {3Eno 3Pyk 3AceEF 2GltA 2Acn Icd SucAB Sdh Fum 2Mdh Icl Mas SucCoACon}

The enzyme names written in braces indicate the enzymes used in the respective mode weighted with their fractional flux (unity if no number is given). Negative values indicate that the reaction is used in the reverse sense.
Abbreviations of metabolites: PG, 2-phosphoglycerate; Ala_{ex}, Asp_{ex}, Glu_{ex}, Suc_{ex}, 'external' alanine, aspartate, glutamate and succinate (that is, formed in reactions that utilize Ala, Asp, Glu and SucCoA in the formation of other amino acids, proteins etc.).
Abbreviations of enzymes: AceEF, pyruvate dehydrogenase; Acn, aconitase; AspA, aspartase; AspC, aspartate aminotransferase; Eno, enolase; Fum, fumarase; Gdh, glutamate dehydrogenase; GltA, citrate synthase; Icd, isocitrate dehydrogenase; Icl, isocitrate lyase; Mas, malate synthase; IlvE/AvtA, branched-chain amino acid aminotransferase/valine-pyruvate aminotransferase; Mdh, malate dehydrogenase; Pck, phosphoenolpyruvate (PEP) carboxykinase; Ppc, PEP carboxylase; Pps, PEP synthetase; Pyk, pyruvate kinase; Sdh, succinate dehydrogenase; SucAB, 2-oxoglutarate dehydrogenase; SucCD, succinyl-CoA synthetase.
Consumption of alanine, aspartate, glutamate and succinyl-CoA are represented by AlaCon, AspCon, GluCon and SucCoACon, respectively.

responsible for the switch between the TCA cycle and the glyoxylate shunt[22]. Indeed, most of the elementary modes that involve Icd do not involve the glyoxylate shunt and *vice versa*, but mode 14 is an exception. If Pck is not used, then glutamate synthesis requires both the dehydrogenation of isocitrate to provide 2-oxoglutarate and the regeneration of OAA via the glyoxylate cycle. This is in agreement with results showing that, upon growth on acetate, Icd is still active in order to produce 2-oxoglutarate[6,7].

What does it all mean?

Elementary modes are idealized situations; the question is, do they occur in living cells? Usually, demands are mixed so that the flux distribution is a superposition of several modes. However, cellular metabolism can support radically different flux distributions in response to different environmental stimuli[1]. When microorganisms grow on single substrates, all pathways that use other substrates are down-regulated; even if a mixture of carbon sources is provided, catabolite repression causes the most efficient substrate to be consumed first. Using DNA microarrays, it has been demonstrated that, during the diauxic shift in yeast, 183 genes are induced and the expression of 203 genes is diminished, both by a factor of at least four[25]. Likewise, in muscle tissue, aerobic metabolism is strongly reduced after prolonged work. This is indicative of a switch between markedly different metabolic modes, which are likely to be superpositions of a limited number of elementary modes. Specific stress situations are dominated by certain pathways. For instance, during starvation, catabolic modes predominate.

Although futile cycles might play a role in thermogenesis and in sensitive regulation in eukaryotes[10,17], it is not so clearly established that they are operative in prokaryotes[6,24]. In any event, it is interesting to identify them, in order to avoid wasteful fluxes following the overexpression of enzymes involved in such cycles. In large networks, cycles that are more complex than the ones described above (Figs 1 and 2) are difficult to find by inspection. However, they can be detected by calculating the elementary modes, which include both cyclic and non-cyclic pathways.

Biotechnological applications

Optimal conversion yields

In many situations, the biosynthesis of a product is feasible by a number of different routes. It is then interesting to determine the way in which the molar yield (that is, the product:substrate ratio) is maximal for the desired product and minimal for byproducts[1,11,12,15]. The optimal flux distributions might not always be obtainable but it is nevertheless interesting to calculate the upper limits for yield and productivity, to know the best that can be expected.

After determining the elementary modes for a given reaction system, the conversion yields can be computed easily by calculating the net stoichiometries for the external substances. As any flux pattern is a superposition of elementary modes with non-negative coefficients, the optimal solution for any product:substrate ratio must always coincide with an elementary mode (Box 2).

The problem of calculating optimal yields has been tackled by methods of linear programming[7,11–13]. One of these is the 'simplex' method, which starts at a feasible solution and then moves along the edges of the admissible region in parameter space towards the point with the highest or lowest value of the cost function; alternatively, computing the elementary modes gives a number of candidate solutions, from which the best can easily by selected. Thus, the elementary-modes method gives a systematic overview of the multitude of flux distributions realizable in the system.

For example, the solutions to the problems of maximizing the ATP:glucose, NADPH:glucose, erythrose-4-phosphate:glucose and PEP:glucose ratios studied by Varma and Palsson[12] in the fueling reaction network of *E. coli* represent elementary modes, because they cannot be decomposed into simpler flux distributions. The graphical representation of these solutions[12], as well as of the solutions in a more complex system[13], has the skeleton-like form typical of elementary modes; that is, not all of the enzymes of the network are used. Often, the yields of optimal and suboptimal situations differ only slightly; for example, in the system studied by Varma and Palsson[12], the optimal ATP:glucose yield, realized by seven modes, is 18.7, and seven other elementary modes give a yield of 18.2. Thus, the elementary-modes method enables the detection of suboptimal distributions; in particular; these can sometimes be more easily realizable in biotechnological applications than the optimal solution.

Tryptophan synthesis

The synthesis of tryptophan from glucose and ammonia is of great biotechnological interest[21,26]. It has been shown previously by a variety of theoretical and experimental approaches that it is the availability of PEP that limits aromatic synthesis in bacteria such as *E. coli*, where the uptake of glucose by the phosphotransferase system uses half the PEP produced in glycolysis[21]. One mole of PEP is used in the synthesis of one mole of the common intermediate for aromatics, 3-deoxyarabinoheptulosonate-7-phosphate (DahP), but this is exacerbated in the synthesis of tryptophan, which also uses a second molecule. In addition, tryptophan synthesis requires the generation of erythrose-4-phosphate (entering the DahP molecule) and ribose 5-phosphate.

Assuming that *E. coli* metabolism is supported by the complete aerobic oxidation of glucose to CO_2, a system comprising 65 reactions can be considered. It gives rise to 26 modes, two of which are the most efficient (for one such example, see Fig. 3). They form 105 moles of tryptophan from 233 moles of glucose, a molar yield of 0.451. (As in many other cases, these very large coefficients in the balanced stoichiometric equation for the optimal yield show why it is very unlikely that the true capabilities of even small metabolic networks might be found by simple inspection.) The extra PEP and pentose needed for tryptophan synthesis cause a lower molar yield than the 0.86 derived by Liao *et al.* using elementary-modes analysis for DahP synthesis[21]. In accordance with these results, all of the modes obtained here require PEP synthetase for pyruvate recycling. Tryptophan synthesis without this enzyme can only be obtained by relaxing some constraints in the model to allow various low-efficiency routes to be used.

The maximal molar yield for the above example (Fig. 3) is slightly larger than the 0.41 found by Varma *et al.*[13] using linear programming. This small discrepancy results from slightly different stoichiometries of oxidation and transhydrogenation used in the two models. The advantage of elementary-modes analysis is in revealing whether the optimal solution is particularly special. In this case, there are six other modes with only slightly lower yields. Interestingly, several of these do not rely on glucose-6-phosphate dehydrogenase to generate the pentoses and tetroses, as for the solution

above (Fig. 3) and DahP synthesis described by Liao *et al.*[21]

Niederberger *et al.*[26] demonstrated that up-modulation of only one of the five enzymes leading from chorismate to tryptophan in yeast increased the tryptophan accumulation rate only a little and that simultaneous overexpression of all of these enzymes has a substantial, non-additive effect. The synthesis flux of an end product should be increased without compensatory effects caused by changes in intermediate concentrations when all enzymes of a particular metabolic route (identifiable by elementary-modes analysis) are overexpressed[27].

Predicting the effects of the insertion or deletion of enzymes

The method presented enables an efficient comparison of the functional capacities of slightly different metabolic networks. In medicine, it can be used as an initial guideline in evaluating the severity of enzyme deficiencies and devising a more specific treatment of such conditions, for example, by stimulating alternative enzymatic activities, which are easily overlooked without exact analysis.

The method also allows a comparison of the sets of admissible routes for a wild-type cell and of an engineered or otherwise-altered cell. This concerns manipulations leading to either an increased or decreased enzyme set. By comparing the elementary modes in the complete system with those in a deficient system, it can be shown whether or not an essential biological substance can still potentially be synthesized, via a bypass in the network system.

For the system shown in Fig. 2, the elementary modes following the blocking of one or more enzymes can easily be computed. For example, the repression of the 2-oxoglutarate-dehydrogenase complex during anaerobic growth of *E. coli* can be modelled by deleting this enzyme; this causes the elementary modes 13, 15 and 16 to drop out. What is completely harmless in this scheme is the deletion of PEP synthetase; this enzyme occurs in one mode only, which is a futile cycle. However, for the synthesis of aromatic amino acids, this enzyme helps to realize a high yield (see above). By contrast, blocking aconitase (which is, for example, severely inhibited by oxidative stress) leads to the deletion of nine modes, including all the modes that generate reducing power and all the pathways that produce glutamate. In this case, the reductive pathway producing succinyl-CoA can still operate (modes 8 and 10).

The insertion of additional enzymes normally leads to an increase in the number of modes. This can result in the synthesis of new end products or the formation of previously available substances with a higher yield. For example, by introducing the *crt* genes from *Erwinia uredovora* or *Erwinia herbicola* into *E. coli*, a pre-existing pathway can be extended so as to produce carotenoids[32]. Large increases in ethanol yields from sugars in *E. coli* were achieved by the insertion of the pyruvate-decarboxylase and alcohol-dehydrogenase-II genes from *Zymomonas mobilis*[33]. Although analysis of the newly emerging metabolic maps is relatively easy when the pathways are linear, the proposed maps often contain question marks, as in the work by Richaud *et al.*[2] They overexpressed L-cystathionine γ synthase (MetB) and disrupted L-cystathionine β lyase (MetC) in *dap* strains of *E. coli*. This induced new or rarely used pathways (which have not yet been completely characterized) to produce *meso*-lanthionine and L-*allo*-cystathionine.

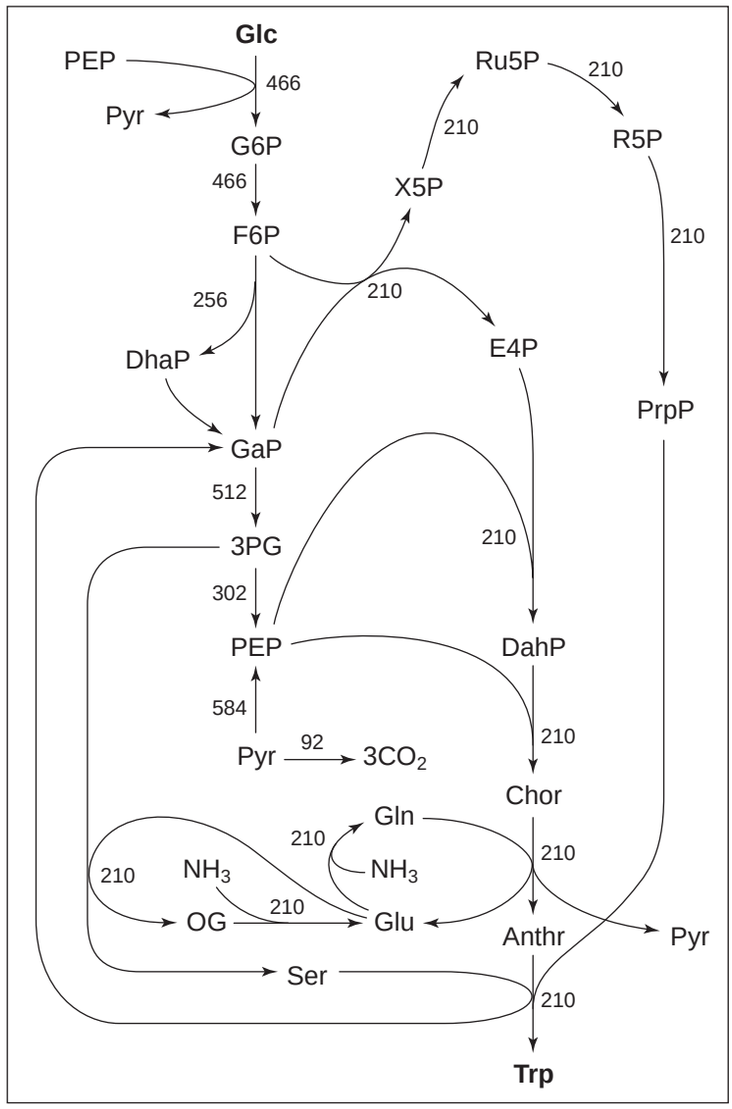

Figure 3

One of the two elementary modes with the highest tryptophan:glucose yield in a simplified reaction scheme of tryptophan synthesis in *Escherichia coli*. The numbers indicate the relative flux carried by the reactions in this mode. They have not been reduced by the factor two because there is an odd number in oxidative phosphorylation (not shown). Glucose, tryptophan, NH_3 and all cofactors (not shown) are treated as external metabolites. Abbreviations: Anthr, anthranilate; Chor, chorismate; DahP, 3-deoxy-arabinoheptulosonate-7-phosphate; DhaP, dihydroxyacetone phosphate; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; GaP, glyceraldehyde-3-phosphate; Glc, glucose; Gln, glutamine, Glu, glutamate; G6P, glucose-6-phosphate; OG, 2-oxo-glutarate; PEP, phosphoenolpyruvate; 3PG, 3-phosphoglycerate; PrpP, phosphoribosylpyrophosphate; Pyr, pyruvate; R5P, ribose-5-phosphate; Ru5P, ribulose-5-phosphate; Ser, serine; Trp, tryptophan; X5P, xylulose-5-phosphate.

Future directions

Various extensions of the elementary-modes approach are worth pursuing. It is certainly interesting to adapt the method for protein biosynthesis. The stoichiometric coefficients of amino acid utilization

in protein synthesis can be estimated from the average amino acid composition of proteins. However, these coefficients are not normally integers, nor are they sufficiently certain. The same problem occurs with the P:O ratio in oxidative phosphorylation. It has turned out that the stoichiometric coefficients can affect not only the weighting coefficients in the modes but also the number of elementary modes. Thus, it might be interesting to develop a more qualitative approach in which the exact values of stoichiometric coefficients are not needed. Another possibility is the consideration of the amino acids as externals, calculating the elementary modes and combining these using the composition ratios as weighting coefficients (which then need not be integers).

Several models take into account the increase in biomass (cell growth)[13]. Growth can be defined as a flux, draining biosynthetic precursors and cofactors with certain stoichiometric proportions. An extension of the method of elementary modes to such growth models deserves special attention in the future.

In many biochemical systems, the number of admissible routes through the system increases rapidly as the number of reactions considered increases (combinatorial explosion)[15]. This occurs, for example, in the presence of isoenzymes or longer parallel branches performing the same overall transformation. For example, adenosine monophosphate (AMP) can be degraded by dephosphorylation (by 5′-nucleotidase) and subsequent deamination (by adenosine deaminase), or the other way round (by AMP deaminase and 5′-nucleotidase). It is then sensible to combine these two branches. Another means for avoiding a combinatorial explosion is to dissect large networks into smaller ones. This can often be achieved by considering appropriate substances (e.g. cofactors) as externals.

One promising field of application is the reconstruction of bacterial metabolism, which has become feasible by the recent advances in sequence analysis[28], proteomics and the study of regulatory RNAs[29]. For example, it has been reported that several non-glycolytic mycoplasma species, such as *Mycoplasma hominis*, not only lack phosphofructokinase and aldolase but also glucose-6-phosphate dehydrogenase[30]. Pollack *et al.*[30] argued that glucose-6-phosphate would enter the pentose phosphate pathway (PPP) and eventually appear as glyceraldehyde-3-phosphate, which enters the lower glycolytic pathway. However, a pathway analysis of the combined glycolysis and PPP scheme shows that there is no elementary mode bypassing the above-mentioned enzymes, because several intermediates of the PPP (such as erythrose-4-phosphate) would not be regenerated. It can be concluded that there is a missing link in the metabolism of *M. hominis*. An example where the analysis of elementary modes does support an assertion made in the reconstruction of metabolism can be seen in mode 8 of Table 1. This corresponds to a main pathway proposed for *Haemophilus influenzae*[31] on the basis of the fact that the *fumC* gene contains a frameshift, so that the enzyme might not be functional.

The use of the concept of elementary modes in metabolic-flux balancing has been outlined by Bonarius *et al.*[8] A further potential application is in drug design, which often concentrates on the structure of the intended target. However, blocking only one enzyme activity will usually evoke a plethora of counter-reactions to achieve homoeostasis. The new metabolic routes after blocking of the drug target can be calculated by elementary-mode analysis. Resistance effects in the design of new antibiotics can be reduced by identifying the most vulnerable sites in the metabolic network where no side routes are open for escape from the effect of the antibiotic.

This method appears to be valuable for comprehending the complex architecture of cell physiology and, together with other theoretical tools such as metabolic-control theory[9,10], it can help to engineer living cells in a directed and rational way.

References

1 Stephanopoulos, G. and Vallino, J. J. (1991) *Science* 252, 1675–1681
2 Richaud, C., Mengin-Lecreulx, D., Pochet, S., Johnson, E. J., Cohen, G. N. and Marlière, P. (1993) *J. Biol. Chem.* 268, 26827–26835
3 Mendes, P. and Kell, D. (1997) *Trends Biotechnol.* 15, 6–7
4 Nielsen, J. (1998) *Biotechnol. Bioeng.* 58, 125–132
5 Maniatis, T., Fritsch, E. F. and Sambrook, J. (1989) *Molecular Cloning; A Laboratory Manual*, Cold Spring Harbor Laboratory Press
6 Holms, W. H. (1986) *Curr. Top. Cell. Regul.* 28, 69–105
7 Pramanik, J. and Keasling, J. D. (1997) *Biotechnol. Bioeng.* 56, 398–421
8 Bonarius, H. P. J., Schmid, G. and Tramper, J. (1997) *Trends Biotechnol.* 15, 308–314
9 Heinrich, R. and Schuster, S. (1996) *The Regulation of Cellular Systems*, Chapman & Hall
10 Fell, D. (1997) *Understanding the Control of Metabolism*, Portland Press
11 Fell, D. A. and Small, J. R. (1986) *Biochem. J.* 238, 781–786
12 Varma, A. and Palsson, B. O. (1993) *J. Theor. Biol.* 165, 477–502
13 Varma, A., Boesch, B. W. and Palsson, B. O. (1993) *Biotechnol. Bioeng.* 42, 59–73
14 Seressiotis, A. and Bailey, J. E. (1988) *Biotechnol. Bioeng.* 31, 587–602
15 Mavrovouniotis, M. L., Stephanopoulos, G. and Stephanopoulos, G. (1990) *Biotechnol. Bioeng.* 36, 1119–1132
16 Mavrovouniotis, M. L. (1992) *Ind. Eng. Chem. Res.* 31, 1637–1653
17 Fell, D. A. (1990) *Comments Theor. Biol.* 6, 1–14
18 Schilling, C. H. and Palsson, B. O. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 4193–4198
19 Schuster, S. and Hilgetag, C. (1994) *J. Biol. Syst.* 2, 165–182
20 Schuster, S., Hilgetag, C., Woods, J. H. and Fell, D. A. (1996) in *Computation in Cellular and Molecular Biological Systems* (Cuthbertson, R., Holcombe, M. and Paton, R., eds), pp. 151–165, World Scientific
21 Liao, J. C., Hou, S-Y. and Chao, Y-P. (1996) *Biotechnol. Bioeng.* 52, 129–140
22 Cronan, J. E., Jr and LaPorte, D. (1996) in *Escherichia coli and Salmonella. Cellular and Molecular Biology* (Vol. I) (Neidhardt, F. C., ed.), pp. 206–215, ASM Press
23 Meléndez-Hevia, E., Waddell, T. G. and Cascante, M. (1996) *J. Mol. Evol.* 43, 293–303
24 Chao, Y-P. and Liao, J. C. (1994) *J. Biol. Chem.* 269, 5122–5126
25 DeRisi, J. L., Iyer, V. R. and Brown, P. O. (1997) *Science* 278, 680–686
26 Niederberger, P., Prasad, R., Miozzari, G. and Kacser, H. (1992) *Biochem. J.* 287, 473–479
27 Kacser, H. and Acerenza, L. (1993) *Eur. J. Biochem.* 216, 361–367
28 Bork, P. and Koonin, E. V. (1998) *Nat. Genet.* 18, 313–318
29 Dandekar, T. and Sharma, K. *Regulatory RNA*, Springer-Verlag (1998)
30 Pollack, J. D., Williams, M. V. and McElhaney, R. N. (1997) *Crit. Rev. Microbiol.* 23, 269–354
31 Tatusov, R. L. *et al.* (1996) *Curr. Biol.* 6, 279–291
32 Misawa, N. and Shimada, H. (1998) *J. Biotechnol.* 59, 169–181
33 Alterthum, F. and Ingram, L. O. (1989) *Appl. Environ. Microbiol.* 55, 1943–1948