Two approaches for metabolic pathway analysis?

Steffen Klamt and Jörg Stelling

Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, D-39106 Magdeburg, Germany

Metabolic pathway analysis is becoming increasingly important for assessing inherent network properties in (reconstructed) biochemical reaction networks. Of the two most promising concepts for pathway analysis, one relies on elementary flux modes and the other on extreme pathways. These concepts are closely related because extreme pathways are a subset of elementary modes. Here, the common features, differences and applicability of these concepts are discussed. Assessing metabolic systems by the set of extreme pathways can, in general, give misleading results owing to the exclusion of possibly important routes. However, in certain network topologies, the sets of elementary modes and extreme pathways coincide. This is quite often the case in realistic applications. In our opinion, the unification of both approaches into one common framework for metabolic pathway analysis is necessary and achievable.

Metabolic pathway analysis is the discovery and analysis of meaningful routes in metabolic networks [1–3]. Intuitively, a route (or pathway) should be a set of connected reactions; it is rather more complicated to give an exact definition of 'meaningful', which should cover physiological as well as biotechnological aspects. Two very similar mathematical concepts (Box 1) have attracted the most attention. They have been applied frequently in the study of the pathway structure in biochemical networks, with partially different intentions [2–19]. The approach published first describes pathways by elementary flux modes [2,3,20], the other by extreme pathways [21]. However, it is often difficult to identify the differences between the two concepts and to see their practical consequences. Therefore, we here compare both approaches in detail and discuss their use in biotechnology.

Elementary flux modes

The concept of elementary flux modes (EFMs) involves three basic conditions (Box 1): a pseudo steady-state condition, a feasibility condition and a non-decomposability (genetic independence) condition. Consider the fictitious network N1 shown in Table 1. We use this hypothetical network intentionally to illustrate the various aspects of pathway analysis in a very small, intuitively understandable example. Metabolites A and B can be taken up from external sources (substrates) in the medium and metabolites P and B can be excreted, of which P is a desired product. Only reactions R2 and R7 are reversible. The eight EFMs that occur in N1 are shown in Fig. 1a and Table 1 shows their stoichiometry (and the stoichiometry of the slightly modified networks discussed later). Take, for example, the first mode, EFM1. The pathway represented by this mode is built up by the reactions R1, R3, R5, R7 and R8. It describes the uptake of A, conversion to B via C and the subsequent synthesis and excretion of P. The pseudo steady-state condition is fulfilled because none of the internal metabolites is consumed in the overall stoichiometry. The pathway is feasible because only the rate of the reversible reaction R7 has a negative sign. Finally, EFM1 is non-decomposable (genetically independent): no subset of the involved reactions can hold the network in a balanced state with non-zero fluxes.

The possibility of assessing functional and structural properties in metabolic networks by EFMs has broad relevance for biotechnology and physiology (the applications described are illustrated for network N1 in Table 2). EFMs can be used to recognize operational modes, including cycles or all pathways leading from a certain metabolite to a product (or biomass) [2,3,5,6,11,13]. For metabolic engineering purposes, it is important that the set of EFMs contains all optimal (and suboptimal) routes converting a certain metabolite (substrate) to a product and, hence, allows detection of these optimal pathways [2,4]. Assessing the number of alternative routes for a given task indicates the sensitivity towards disturbances such as mutations, which enables the analysis of flexibility (redundancy, structural robustness [19]) of a network or its engineered variants.

The following three applications are also useful for suggesting regulatory rules. First, at a more detailed level, EFM analysis enables the determination of the relative importance of individual reactions for system performance under different growth regimes. This was exploited for a rough prediction of transcription level ratios in *Escherichia coli* [19]. Second, elementary mode analysis allows the identification of enzyme subsets, that is, reactions in a network that always have to operate together and which, thus, structurally need each other [22]. Typical examples are unbranched linear pathways, such as the five reactions for synthesizing tryptophan from chorismate in *E. coli*. Enzymes catalyzing reactions of the same enzyme subset can be supposed to share common regulatory circuits. Indeed, in *E. coli*, the corresponding five genes for tryptophan synthesis belong to the same operon. Third, in a similar way, we propose to search for

Corresponding author: Steffen Klamt (klamt@mpi-magdeburg.mpg.de).

 0167-7799/03/$ - see front matter PII: S0167-7799(02)00034-3

Box 1. The mathematics behind elementary flux modes (EFMs) and extreme pathways (EPs)

Metabolic networks composed of q reactions and m metabolites are usually represented by a stoichiometric matrix **N** of m rows and q columns. In such a network, each EFM and EP is defined by a vector $\boldsymbol{e}$, composed of q elements $(e_1, e_2, \ldots e_q)$ each describing the net rate of the corresponding reaction (i.e. vector $\boldsymbol{e}$ is a flux distribution). The pathway represented by $\boldsymbol{e}$ can be identified by the utilized reactions, i.e. where $e_i \neq 0$. We denote this by P($\boldsymbol{e}$), which supplies all reactions in $\boldsymbol{e}$ having a rate unequal to zero. If $\boldsymbol{e}$ is an EFM or EP then it fulfills the following three conditions [a–e]:

- (C1) *Pseudo steady-state*: $\mathbf{N}\,\boldsymbol{e} = 0$. This ensures that none of the metabolites is consumed or produced in the overall stoichiometry (metabolite balancing equation).
- (C2) *Feasibility*: rate $e_i \geq 0$ if reaction i is irreversible. This demands that only thermodynamically realizable fluxes are contained in $\boldsymbol{e}$.
- (C3) *Non-decomposability*: there is no vector $\boldsymbol{v}$ (unequal to the zero vector and to $\boldsymbol{e}$) fulfilling C1 and C2 and that P($\boldsymbol{v}$) is a proper subset of P($\boldsymbol{e}$). This is the core characteristic for EFMs and EPs and supplies the decomposition of the network into smallest units (able to hold the network in steady state). Non-decomposability is also often called 'genetic independence' because C3 implies that the enzymes in one EFM or EP are not a subset of the enzymes from another EFM or EP.

Conditions C1–C3 completely define an EFM; for an EP two more conditions have to be satisfied, which are explained below. We denote here the complete set of EFMs in a network by S_{EFM} and analogue the complete set of EPs by S_{EP}. C1 and C3 together ensure that the subnetwork spanned by the reactions participating in pathway $\boldsymbol{e}$ is connected (proof not shown). Furthermore, the basic conditions C1–C3 imply that all feasible steady-state flux distributions $\boldsymbol{v}$ can (not necessarily uniquely) be described by a non-negative superposition of all EFMs or all EPs, respectively:

- (E1) $\boldsymbol{v} = \sum_j \alpha_j \boldsymbol{e}^j$ α_j: scalar with $\alpha_j \geq 0$; $\mathbf{e}^j$ is the j-th EFM $\in S_{EFM}$ or the j-th EP $\in S_{EP}$, respectively.

Equations as in (E1) are central to a more general mathematical framework called convex analysis [f]. Indeed, both concepts were originally derived from an approach in which convex techniques have been used to assess the steady states in stoichiometric networks [g]. There, pathways have been called 'extreme currents'. A limiting restriction in these studies was that all reactions had to be irreversible.

The conditions C1–C3 do already uniquely determine the complete set of EFMs in a network (up to a scaling factor for each pathway vector). Because each EP obeys conditions C1–C3 too, the set of EPs in a network is always a (proper or non-proper) subset of the EFMs: $S_{EP} \subseteq S_{EFM}$. Two additional conditions introduced in [d] delimit the EPs from EFMs:

- ($C4_{EP}$) *Network reconfiguration*: each reaction must be classified either as exchange flux (which allows a metabolite to enter or to exit the system) or as internal reaction. Then, all reversible internal reactions must be split up into two separate, irreversible reactions (forward and backward direction). As consequence, no internal reaction can have a negative flux. Exchange fluxes can be reversible, but each metabolite can participate in only one exchange flux. The meaning of 'exchange flux' or 'exchange reaction' used in this context is different from the same term used in ^{13}C metabolic flux analysis, where it denotes an intracellular bidirectional flux.
- ($C5_{EP}$) *Systemic independence*: the set of EPs in a network (configured properly by $C4_{EP}$) is the *minimal* set of EFMs that can describe all feasible steady-state flux distributions by equation (E1). Mathematically, the EPs represent a *convex basis* in this network. The reconfiguration ($C4_{EP}$) ensures that the set of EPs is unique. (Convex bases are also defined in non-reconfigured networks but are then, in general, not unique [h]).

Whereas splitting-up reversible reactions is straightforward, declaring the exchange fluxes as required by ($C4_{EP}$) is sometimes confusing and was not precisely explained in [d]. As far as we can see, two different schemes have been applied for reconfiguring a network with respect to the exchange fluxes: (i) If in the original network each metabolite participates (exclusively) in at most one of the (pseudo) reactions crossing the system boundaries, then these reactions can be considered as the exchange fluxes (compare this with the example in [d]). Reconfiguration using this scheme is straightforward as applied in network N2 (see text and Table 1). (ii) In some applications (compare this with the central metabolic network in *Escherichia coli* studied in [i]), each external source/sink was converted into an internal one and the exchange reaction for this metabolite was defined by adding a (pseudo) reaction providing or withdrawing the metabolite from/to the environment. Only this scheme enables the consideration of more than one transporter for the same metabolite. However, when this scheme of reconfiguration is applied, the sets of meaningful EPs and EFMs are always identical.

References

a Schuster, S. *et al.* (1999) Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.* 17, 53–60

b Schuster, S. *et al.* (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* 18, 326–332

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

d Schilling, C.H. *et al.* (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* 203, 229–248

e Schuster, S. *et al.* (2002) Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol.* 45, 153–181

f Rockafellar, R.T. (1970) *Convex Analysis*, Princeton University Press

g Clarke, B.L. (1988) Stoichiometric network analysis. *Cell Biophys.* 12, 237–253

h Pfeiffer, T. *et al.* (1999) METATOOL: For studying metabolic networks. *Bioinformatics* 15, 251–257

i Schilling, C.H. *et al.* (2001) Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol. Bioeng.* 71, 286–306

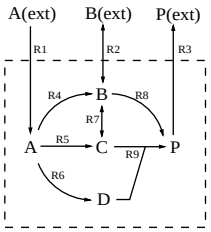
pairs of reactions that never occur together in an EFM and to call them 'excluding reaction pairs'. A very simple example is isoenzymes catalyzing an irreversible reaction, such as the three isofunctional aspartokinases in *E. coli*. Excluding reaction pairs are likely to be regulated independently, as indeed is the case for the three aspartokinases. Network N1 demonstrates that excluding reaction pairs are not limited to isoenzymes (Table 2). Note that excluding reaction pairs can, nevertheless, occur together in a steady-state flux distribution obtained by combining EFMs. Furthermore, whereas enzyme subsets can also be identified by a null space analysis of the stoichiometric matrix, excluding reaction pairs cannot and therefore require elementary-mode analysis.

The length of an EFM might also be interesting, because this stands for the demand of cellular resources for enzyme synthesis to install a pathway. Moreover, EFMs allow for calculability analysis for metabolic flux analyses [23].

The set of EFMs related to a certain network also shows extremely useful conservation properties. If a reaction is removed from the network, all EFMs not involving this reaction build up the complete set of EFMs in the new (smaller) network [3]. They are a subset of the EFMs in the

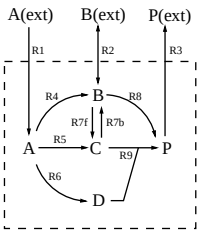
Table 1. Configurations of the example network (upper part N1 and N3; lower part N2 and N4), with corresponding elementary flux modes (EFM) and extreme pathways (EP) (see also Fig. 1)

N1 (R2 and R7 reversible) N3 (as N1 but R2 irreversible)



N1 EFMs	N3 EFMs	R1	R2	R3	R4	R5	R6	R7	R8	R9
EFM1	×	1	0	1	0	1	0	−1	1	0
EFM2	×	1	0	1	1	0	0	0	1	0
EFM3	×	2	0	1	0	1	1	0	0	1
EFM4	×	2	0	1	1	0	1	1	0	1
EFM5	×	1	1	1	0	0	1	1	0	1
EFM6		1	−1	0	1	0	0	0	0	0
EFM7		1	−1	0	0	1	0	−1	0	0
EFM8	×	0	1	1	0	0	0	0	1	0

N2 (R2 reversible, R7 split up) N4 (as N2 but R2 irreversible)



N2 EFMs	N2 EPs	N4 EFMs	N4 EPs	R1	R2	R3	R4	R5	R6	R7f	R8	R9	R7b
EFM1		×	EP1′	1	0	1	0	1	0	0	1	0	1
EFM2		×	EP2′	1	0	1	1	0	0	0	1	0	0
EFM3	EP1	×	EP3′	2	0	1	0	1	1	0	0	1	0
EFM4		×	EP4′	2	0	1	1	0	1	1	0	1	0
EFM5	EP2	×	EP5′	1	1	1	0	0	1	1	0	1	0
EFM6	EP3			1	−1	0	1	0	1	0	0	0	0
EFM7	EP4			1	−1	0	0	1	0	0	0	0	1
EFM8	EP5	×	EP6′	0	1	1	0	0	0	0	1	0	0
EFM9	EP6	×	EP7′	0	0	0	0	0	0	1	0	0	1

original network. Predicting whether a mutant missing one or several reaction(s) will be structurally able to grow, therefore, simply means searching for EFM(s) in which these reactions are not used [19]. Similarly, when a reversible reaction, Rx, is changed to be irreversible, the resulting set of feasible pathways is obtained by taking the original ones excluding those using Rx in the forbidden direction [24]. Conversely, this enables the calculation of the EFMs separately for forward and backward direction. A union of these sets then gives the complete set of EFMs for the network with the reversible reaction. In general, pathway analysis in subnetworks thus needs to consider

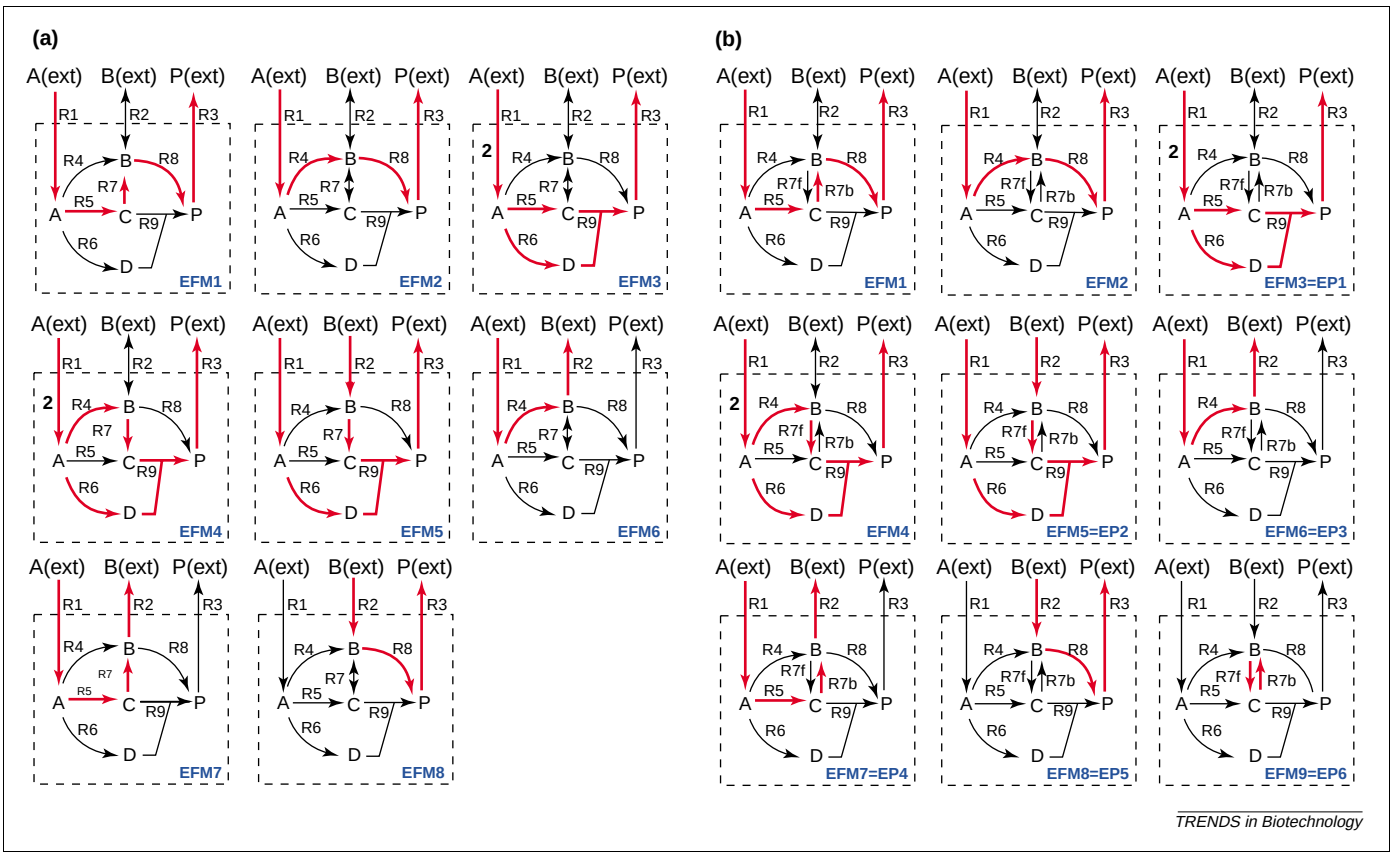


Fig. 1. Graphical representation of (a) the elementary flux modes (EFMs) in network N1 and (b) the EFMs and extreme pathways in network N2 (see also Table 1).

Table 2. Case study on the applicability of elementary flux modes and extreme pathways

Problem/example	Elementary flux modes (network N1)	Extreme pathways (network N2)
Recognition of operational modes: routes for converting exclusively A to P	Analysis yields four genetically independent routes (EFM1–EFM4)	The set of EPs does not contain all genetically independent routes. Searching for EPs leading from A to P via B, no pathway would be found (but compare with EFM1 and EFM2)
Finding all the optimal routes: optimal pathways for synthesizing P during growth on A alone	EFM1 and EFM2 are optimal because they yield one mole P per mole substrate A (i.e. R3/R1 = 1), whereas EFM3 and EFM4 are only suboptimal (R3/R1 = 0.5)[a]	One would only find the suboptimal EP1, not the optimal routes EFM1 and EFM2
Analysis of network flexibility (structural robustness, redundancy): relative robustness of exclusive growth on A or B	Four pathways convert A to P (EFM1–EFM4), whereas for B only one route (EFM8) exists. When one of the internal reactions (R4–R9) fails, for production of P from A two pathways will always 'survive'. By contrast, removing reaction R8 already impedes the production of P from B alone	Only one EP exists for producing P by substrate A alone, and one EP for synthesizing P by (only) substrate B. One might suggest that both substrates possess the same redundancy of pathways, but as shown by EFM analysis, growth on substrate A is much more flexible than on B
Relative importance of single reactions: relative importance of reaction R8	R8 is essential for producing P by substrate B, whereas for A there is no structurally 'favoured' reaction (R4–R9 all occur twice in EFM1–EFM4). However, considering the optimal modes EFM1,2, one recognizes the importance of R8 also for growth on A	Consider again biosynthesis of P from substrate A (EP1 only). Because R8 is not involved in EP1 one might think that this reaction is not important for synthesizing P from A. However, as can be easily verified, without this reaction it is impossible to obtain optimal yields (one P per A; EFM1 and EFM2)
Enzyme subsets and excluding reaction pairs: suggest regulatory structures or rules	R6 and R9 are an enzyme subset. By contrast, R6 and R9 never occur together with R8 in an EFM. Thus, (R6,R8) and (R8,R9) are excluding reaction pairs (of course, in an arbitrary composable steady-state flux distribution they might occur together)	The EPs pretend R4 and R8 to be an excluding reaction pair – but they are not (EFM2). The enzyme subsets would be correctly identified. However, one can construct simple examples where the EPs would also pretend wrong enzyme subsets (not shown)
Pathway length: shortest/longest pathway for production of P from A	The shortest pathway from A to P needs two internal reactions (EFM2), the longest requires four (EFM4)	Both the shortest (EFM2) and the longest (EFM4) pathway from A to P are not contained in the set of EPs
Removing a reaction and mutation studies: effect of deleting R7	All EFMs not involving the specific reactions build up the complete set of EFMs in the new (smaller) sub-network. If R7 is cut away, EFMs 2, 3, 6 and 8 would 'survive', hence the mutant is viable	Analyzing a subnetwork implies that the EPs must be newly computed. For example, when cutting away reaction R2 the EFM2 (not contained in the original set of EPs) would become an EP. For this reason mutation studies cannot be performed easily
Constraining reaction reversibility: effect of R7 limited to B → C	For the case of R7, all EFMs but EFM1 and EFM7 'survive', because the latter ones utilize R7 with negative rate	In general, the set of EPs must be recalculated: compare the EPs in network N2 (R2 reversible) and N4 (R2 irreversible)

[a]Unlike analyzing the EFMs, linear optimization as used for example in Flux Balance Analysis (FBA, e.g. [8,30]) finds only a single and, in many cases, a nonunique solution for pathways with optimal product yield. In terms of linear programming, the set of EFMs contains all basic feasible solution. But note, it is generally difficult to consider further constraints (such as fixing different rates to values unequal to zero) when analyzing optimality by EFMs.

only a subset of the originally calculated EFMs for the complete network.

Extreme pathways

Launching the concept of extreme pathways (EPs), two additional conditions of network reconfiguration and systemic independence were introduced to obtain a minimal set of pathways capable to describe all feasible steady-state flux distributions (Box 1) [21]. In network N2, N1 has been reconfigured as required. The reactions R1, R2 and R3 represent the exchange fluxes (scheme (i) can be applied here; Box 1), of which R2 is reversible as permitted by the concept. The reversible internal reaction R7 has had to be split up into a forward (R7f: B → C) and a backward reaction (R7b: C → B), both being irreversible. Six EPs can be calculated in this reconfigured network (Table 1, Fig. 1b). For each originally reversible internal reaction, the set of EPs always includes a 'two-cycle' of forward and backward reaction (EP6: R7f and R7b). Given that these EPs only reformulate the fact of reversibility, they have no practical meaning and are usually not considered further [8]. We can also compute the EFMs in this network, confirming that the EPs are a subset of the EFMs (Table 1, Fig. 1b). The following three EFMs in N2 are not contained in the set of EPs because they can be represented by a non-negative linear combination of EPs and hence are not systemically independent: EFM1 = EP4 + EP5, EFM2 = EP3 + EP5, EFM4 = EP2 + EP3.

Comparison of the EFMs in N1 and N2 shows that each EFM in the modified network has a 'corresponding' EFM in the original network up to the two-cycle(s). Note that possibly occurring reversible modes in the original network are considered separately for each direction. In cases where the original EFM (e.g. EFM1 in N1) involves a reversible reaction (R7), the corresponding EFM in the reconfigured network (EFM1 in N2) only assigns this flux either to the forward or backward reaction (here, to R7b because R7 in EFM1 is negative). The only difference in the

set of EFMs emerging upon reconfiguration consists in the two-cycles that result from splitting up reversible reactions. However, two-cycles are not considered as meaningful pathways. We formulate this fact explicitly as 'Property 1', which is valid for any network and can be proven by the conservation law of EFMs for restricting reversible reactions to be irreversible:

- Property 1: Reconfiguring a network by splitting up reversible reactions leads to the same set of meaningful EFMs.

The fact that the set of EPs is a subset of EFMs has often been emphasized as an advantage. Calculating EFMs is indeed computationally very demanding because of the combinatorial complexity [25,26]. However, considering the EPs instead of EFMs generally reduces not only the number of pathways but also the applicability for assessing network properties. In Table 2, several counter-examples in N2 illustrate that, in fact, none of the applications for EFMs can be handled by EPs for the general case, because the set of EPs can miss important genetically independent pathways. In N2, for example, the two optimal pathways leading from A to P would not be identified (Table 2). Thus, the claim that EPs can be used to find all optimal routes [15,16] is, in general, not true. For the same reason, structural robustness (redundancy) and the relative importance of reactions cannot be assessed properly. The information reflected by the length of EPs has only uncertain meaning. In N2 neither the shortest nor the longest pathway from A to P would be identified (Table 2). Furthermore, EPs are not suitable for calculability analysis. Even more importantly, the set of extreme pathways needs to be recalculated every time a reaction is removed or changed from reversible to irreversible.

In our eyes, these examples demonstrate that the EPs are generally much less suitable for pathway analysis, simply because they are only a subset of the EFMs. The exclusion of (perhaps important) pathways in the set of EPs is caused by demanding systemic independence, which is, however, also difficult to interpret from a physiological point of view. Consider again the genetically independent pathway from A to P via C and B (EFM1). Although it seems intuitively reasonable to consider this pathway as meaningful, it is not an EP because it is not systemically independent (EFM1 = EP4 + EP5). However, looking at the superposition of EP4 and EP5 representing pathway EFM1 suggests that B must be excreted and taken up simultaneously and, hence, that R2 is used in both directions at the same time. Because R2 occurs in EP4 and EP5, it also suggests that R2 is required for this pathway, although not involved in EFM1. Therefore, the property of systemic independence might be mathematically meaningful but, in our opinion, is not physiologically meaningful.

Many applications: both concepts coincide

It is clear that systemically dependent EFMs that are not EPs can occur when reversible exchange fluxes are contained in the network. Only these bidirectional fluxes allow a representation of an EFM by a sum of several others. However, in interesting pathway studies (relying on EPs) in genome-scale networks of *Haemophilus influenzae* [15] and *Helicobacter pylori* [16] all exchange fluxes are irreversible. Hence, none of the metabolites is considered to serve as substrate and product simultaneously. What is the consequence of when all exchange fluxes (and hence all reactions in the network) are irreversible?

To study this question, in N3 we took the original network N1 but changed R2 to be irreversible. Metabolite B is now considered as a possible substrate; the case of B being an excreted product can be treated in the same way. In N4 we reconfigured N3 in the same way as N2. We then calculated the EFMs for N3 and N4. It was possible to exploit the conservation property of EFMs: EFMs from N1 and N2 'survive' in N3 and N4, respectively, where B is not excreted (i.e. $R2 \geq 0$; indicated in Table 1 by a cross). Calculating the EPs for N4 (denoted EPx' in Table 1) reveals that they coincide exactly with the EFMs in N4. Indeed, this phenomenon always occurs when all exchange fluxes are irreversible, and so when all reactions in the reconfigured network are unidirectional. In such a case, no EFM can be represented by a non-negative sum of other EFMs and, therefore, the set of EFMs and EPs must be identical! Indeed, the coincidence of elementary modes and extreme pathways is not excluded by the conditions to which extreme pathways are subjected.

There is yet another consequence: given that each meaningful EFM in N4 has a corresponding EFM in N3 (because of 'Property 1'), each EP in N4 does too, and reconfiguring network N3 to N4 merely has the marginal effect of additionally obtaining the non-meaningful two-cycles. We summarize these findings in the generally valid 'Property 2':

- Property 2: If all exchange reactions in a network are irreversible then the sets of meaningful EFMs (both in the original and in the reconfigured network) and EPs coincide.

Hence, reconfiguring a network with only irreversible exchange reactions (such as N3) by splitting-up internal reversible reactions (as R7 into R7f and R7b in N4) is only another way to calculate the EFMs occurring in the original network. Moreover, this way is computationally more expensive because the number of reactions in the reconfigured network increases by the number of internal reversible reactions.

As mentioned above, the networks studied by EPs in [15,16] apparently comprise only irreversible exchange fluxes. Thus, the set of EPs analyzed there equals the set of EFMs. Although this fact has not previously been noticed (or mentioned), in our opinion this is an essential prerequisite for interpreting the calculated sets of EPs by the tools summarized above for the elementary modes to obtain meaningful results.

To round-off these considerations: one might argue that acetate, for example, can serve as a substrate or product, according to the environmental conditions. For an independent consideration of the cases 'substrate' or 'product', one might also split-up reversible exchange reactions (e.g. R2 in N2) into two irreversible ones as required for the internal reactions. However, because of 'Property 1' this

would have no effect on the calculated meaningful EFMs (thus, the set of EFMs in the original network already contains both cases 'substrate' and 'product' independently). Furthermore, because of 'Property 2', the EFMs would again coincide with the EPs and, moreover, in this case they are also identical to the extreme currents defined only in such networks with exclusively unidirectional reactions (Box 1) [27].

Conclusions and outlook

Comparing the two prominent approaches to pathway analysis, we feel that elementary flux modes offer great opportunities for studying functional and structural properties in metabolic networks. We assume that there are even more benefits than we have summarized here. Regarding the extreme pathways two cases can be distinguished: (1) The set of meaningful EPs is identical to the set of EFMs. As shown above, this is certainly the case if only irreversible exchange fluxes are considered, but it is not limited to such network topologies. Besides, when the exchange fluxes are defined according to scheme (ii) in Box 1 both sets are always identical. Obviously, when EPs and EFMs coincide, all the useful analysis tools for EFMs can also directly be applied to the EPs, but calculating the EPs is then computationally more expensive. (2) The set of EPs is a proper subset of the set of EFMs. In this case, the interpretation of EPs can lead to questionable results owing to missing genetically independent pathways.

To establish a unified notation for pathway analysis we suggest that the original term 'elementary flux modes' is also used for 'extreme pathways' (calculated in reconfigured networks) whenever both sets of pathways are identical. Accordingly, one should check carefully to see whether case (1) is present when a network is studied by EPs. The term 'extreme pathway' should indicate case (2), in which the set of EPs is a proper subset of EFMs and hence is not identical. However, in this case, it remains uncertain how much valuable information about the pathway structure is lost.

Pathway analysis undoubtedly has great potential for biotechnology and metabolic engineering. It helps us to gain a better understanding of the cellular metabolism and to find possible targets for manipulation. First attempts to analyze the pathway structure in complex networks, partially by subdividing them, reveal interesting results (e.g. on pathway redundancy) that are useful for comparing different networks *in silico*. A challenging task for the future is the calculation and study of the complete set of pathways at a genomic scale, and its combination with cellular regulation to obtain the whole picture [28,29].

References

1 Schilling, C.H. *et al.* (1999) Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* 15, 296–303
2 Schuster, S. *et al.* (1999) Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.* 17, 53–60
3 Schuster, S. *et al.* (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* 18, 326–332
4 Liao, J.C. *et al.* (1996) Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnol. Bioeng.* 52, 129–140
5 Nuño, J.C. *et al.* (1997) Network organization of cell metabolism: monosaccharide interconversion. *Biochem. J.* 324, 103–111
6 Rohwer, J.M. and Botha, F.C. (2001) Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data. *Biochem. J.* 358, 437–445
7 Schilling, C.H. and Palsson, B.O. (2000) Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome scale pathway analysis. *J. Theor. Biol.* 203, 249–283
8 Schilling, C.H. *et al.* (2001) Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol. Bioeng.* 71, 286–306
9 Schilling, C.H. *et al.* (2002) Genome-scale metabolic model of *Helicobacter pylori* 26695. *J. Bacteriol.* 184, 4582–4593
10 Carlson, R. *et al.* (2002) Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnol. Bioeng.* 79, 121–134
11 Van Dien, S.J. and Lidstrom, M.E. (2002) Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph Methylobacterium extorquens AM1, with application to reconstruction of C(3) and C(4) metabolism. *Biotechnol. Bioeng.* 78, 296–312
12 Schuster, S. *et al.* (2002) Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumomniae*. *Bioinformatics* 18, 351–361
13 Schuster, S. *et al.* (2002) Use of network analysis of metabolic systems in bioengineering. *Bioproc. Biosyst. Eng.* 24, 363–372
14 Schuster, S. and Klamt, S. (2002) Applying metabolic pathway analysis to make good use of methanol. *Trends Biotechnol.* 20, 322
15 Papin, J.A. *et al.* (2002) The genome scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. *J. Theor. Biol.* 215, 67–82
16 Price, N.D. *et al.* (2002) Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis. *Genome Res.* 12, 760–769
17 Wiback, S.L. and Palsson, B.O. (2002) Extreme pathway analysis of human red blood cell metabolism. *Biophys. J.* 83, 808–818
18 Foerster, J. *et al.* (2002) A functional genomics approach using metabolomics and in silico pathway analysis. *Biotechnol. Bioeng.* 79, 703–712
19 Stelling, J. *et al.* (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420, 190–193
20 Schuster, S. and Hilgetag, C. (1994) On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* 2, 165–182
21 Schilling, C.H. *et al.* (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* 203, 229–248
22 Pfeiffer, T. *et al.* (1999) METATOOL: For studying metabolic networks. *Bioinformatics* 15, 251–257
23 Klamt, S. *et al.* (2002) Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. *Biotechnol. Bioeng.* 77, 734–751
24 Schuster, S. *et al.* (2002) Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol.* 45, 153–181
25 Klamt, S. *et al.* (2002) FluxAnalyzer: exploring structure, pathways and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics* in press
26 Klamt, S. and Stelling, J. (2002) Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.* 29, 233–236
27 Clarke, B.L. (1988) Stoichiometric network analysis. *Cell Biophys.* 12, 237–253
28 Palsson, B.O. (2000) The challenges of in silico biology. *Nat. Biotechnol.* 18, 1147–1150
29 Palsson, B.O. (2002) In silico biology through 'omics'. *Nat. Biotechnol.* 20, 649–650
30 Covert, M.W. *et al.* (2001) Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* 213, 73–88