Opinion TRENDS in Biotechnology Vol.22 No.8 August 2004

Comparison of network-based pathway analysis methods

Jason A. Papin[1], Joerg Stelling[2], Nathan D. Price[1], Steffen Klamt[2], Stefan Schuster[3] and Bernhard O. Palsson[1]

[1]Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive #0412, La Jolla, CA 92093-0412, USA
[2]Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstrasse 1, 39106 Magdeburg, Germany
[3]Jena University, Department of Bioinformatics, Ernst-Abbe-Platz 2, 07743 Jena, Germany

Network-based definitions of biochemical pathways have emerged in recent years. These pathway definitions insist on the balanced use of a whole network of biochemical reactions. Two such related definitions, elementary modes and extreme pathways, have generated novel hypotheses regarding biochemical network function. The relationship between these two approaches can be illustrated by comparing and contrasting the elementary modes and extreme pathways of previously published metabolic reconstructions of the human red blood cell (RBC) and the human pathogen *Helicobacter pylori*. Descriptions of network properties generated by using these two approaches in the analysis of realistic metabolic networks need careful interpretation.

As an ever-increasing number of biochemical reaction networks are being reconstructed [1–3], there is a growing need to assess the properties that emerge from these networks. Network-based pathway analysis facilitates such an assessment. Recent network-based metabolic pathway analysis has focused on two approaches, those of elementary modes [4] and extreme pathways [5]. Both of these methods use convex analysis [6], a branch of mathematics that enables the analysis of inequalities and systems of linear equations, to generate a convex set of vectors that can be used to characterize all of the steady-state flux distributions of a biochemical network. The roughly 20-year history of this field has recently been reviewed [7,8] and is summarized in Figure 1.

The subtle differences in the definitions of elementary modes and extreme pathways can lead to discrepancies in their use. For many situations the two definitions lead to identical results, but for others there are differences. These differences have led to two recent opinion articles in *Trends in Biotechnology* [9,10] in which the two approaches are contrasted. Here we extend this comparison by considering realistic biological networks, clearly articulating the differences and relationship between elementary modes and extreme pathways that potential users will have to consider.

Conceptual framework

Network-based pathway analysis

Biological networks can be represented by a stoichiometric matrix (S). The rows of S correspond to the compounds (e.g. metabolites) in a reaction network. The columns of S correspond to the reactions in a network, with elements corresponding to stoichiometric coefficients of the associated reactions. At steady state, mass balance in a network can be represented by the flux-balance equation:

$$S\mathbf{v} = 0 \quad (1)$$

where $\mathbf{v}$ is a vector whose elements correspond to fluxes through the associated reactions in S. The set of all possible solutions to Equation 1 can be described by a set of basis vectors.

Convex analysis has been used to generate a set of chemically valid basis vectors to describe the solutions to Equation 1, by allowing the application of inequality constraints on the flux values of the irreversible reactions:

$$v_i \geq 0 \quad (2)$$

where $\mathbf{v}_i$ is the flux through reaction i. A set of valid solutions to Equation 1 subject to the constraints in Equation 2 can be described as a high-dimensional cone that is located in a space where each axis corresponds to a reaction flux. Extreme currents [11], elementary modes [4] and extreme pathways [5] use convex analysis to describe this solution space. Every valid flux distribution in a reaction network can be represented as a nonnegative combination of the convex basis vectors.

Elementary modes

Elementary modes ($\mathbf{e}_i$) are a set of vectors derived from the stoichiometric matrix of a biochemical network by using convex analysis [12]. They have the following three properties.

(I) There is a unique set of elementary modes for a given network.

(II) Each elementary mode consists of the minimum number of reactions that it needs to exist as a functional unit. If any reaction in an elementary mode were removed, the whole elementary mode could not operate as a functional unit. This property has been called 'genetic independence' and 'non-decomposability' [12].

Corresponding author: Bernhard O. Palsson (palsson@ucsd.edu).
Available online 3 July 2004

 0167-7799/$ - see front matter doi:10.1016/j.tibtech.2004.06.010

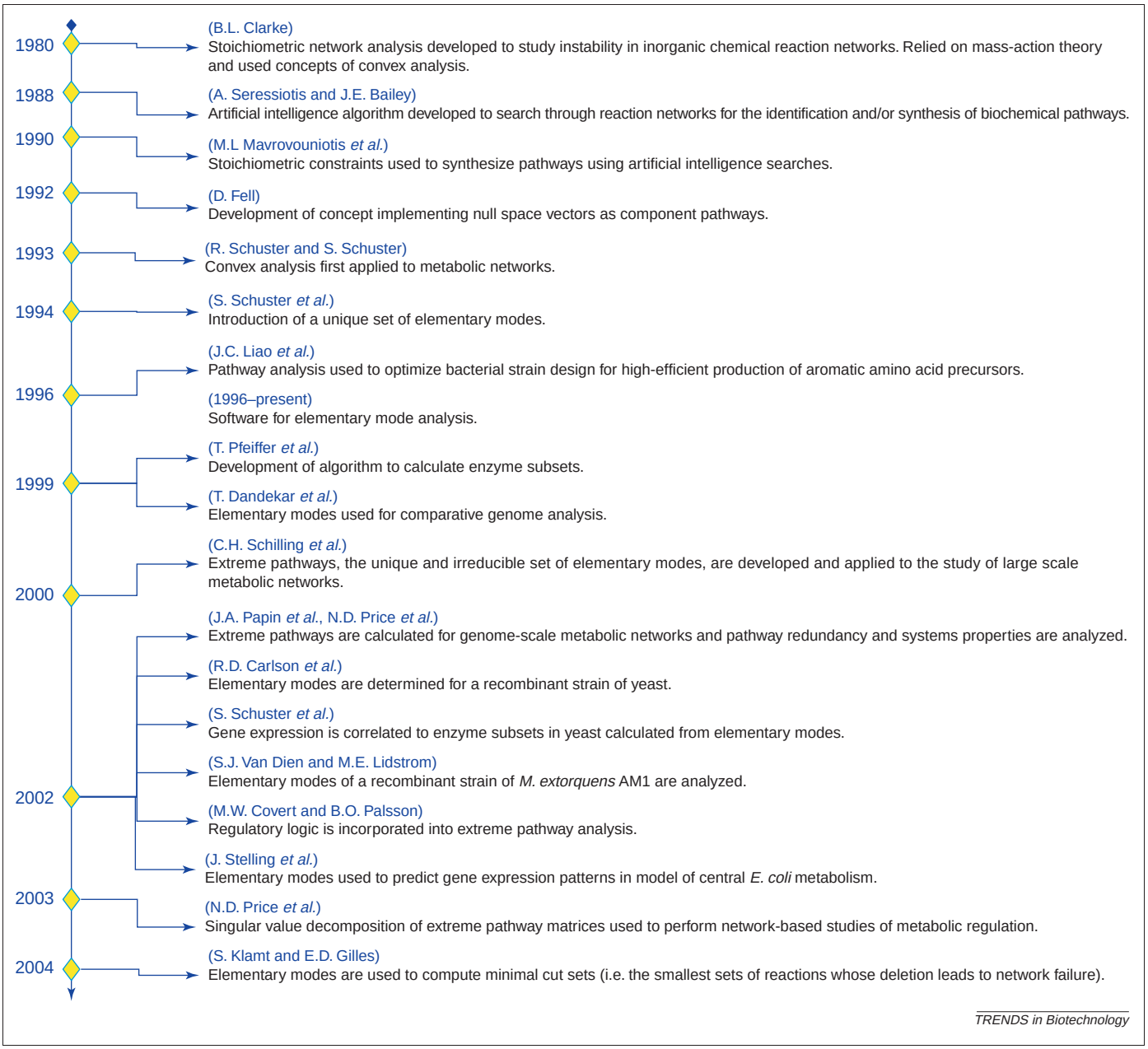

Figure 1. A brief history of the analysis of network-based pathways.

(III) The elementary modes are the set of all routes through a metabolic network consistent with property II.

Extreme pathways

Extreme pathways ($\mathbf{p}_i$) are a set of convex basis vectors derived from the stoichiometric matrix [5]. They have the following properties:

(I) There is a unique set of extreme pathways for a given network.

(II) Each extreme pathway consists of the minimum number of reactions that it needs to exist as a functional unit.

(III) The extreme pathways are the systemically independent subset of elementary modes; that is, no extreme pathway can be represented as a nonnegative linear combination of any other extreme pathways.

Comparison of elementary modes and extreme pathways

Properties I and II are shared between extreme pathways and elementary modes. The extreme pathways correspond to the edge representation of convex polyhedral cones [6]. The algorithms for elementary modes and extreme pathways treat internal reversible and irreversible reactions differently: extreme pathway analysis decouples all internal reversible reactions into two separate reactions for the forward and reverse directions, and subsequently calculates the pathways; elementary mode analysis accounts for reaction directionality through a series of rules in the corresponding calculations of the modes.

The extreme pathways and elementary modes for a simple reaction system are shown in Figure 2. Satisfying the requirement of systemic independence can result in fewer extreme pathways than elementary modes. For

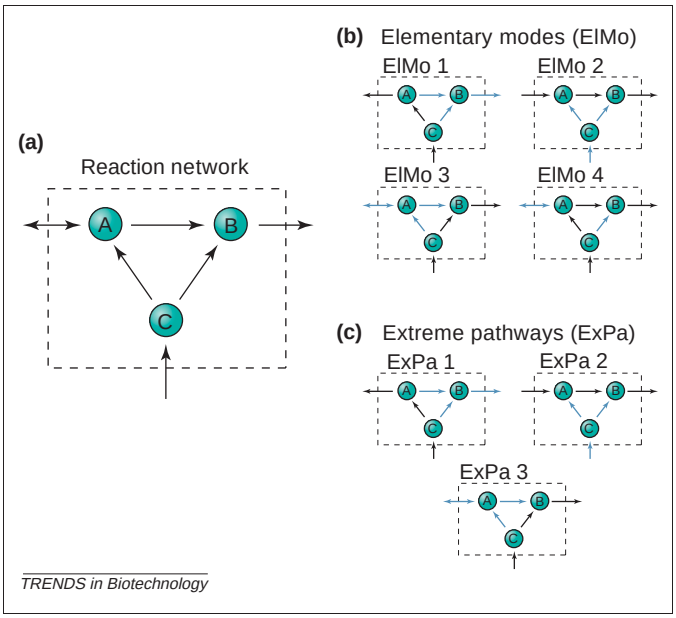

Figure 2. Simple example of a biochemical network and its extreme pathways and elementary modes. The network consists of three metabolites, three internal reactions and three exchange reactions **(a)**, and there are four elementary modes **(b)** and three extreme pathways **(c)**. The difference between the two sets of pathways revolves around the use of the reversible exchange flux for the metabolite A. The elementary mode EIMo 4 is a nonnegative linear combination of the extreme pathways ExPa 1 and ExPa 2, because the reversible exchange flux for metabolite A can be canceled out. This simple example demonstrates the principal characteristic that differentiates the two approaches to network-based pathway analysis discussed in this article.

example, linear combinations of the extreme pathways might satisfy the genetic independence requirement of the elementary modes (in Figure 2, the extreme pathways ExPa 1 and ExPa 2 can be combined to describe the elementary mode EIMo 4). Therefore, the extreme pathways are not a set of all genetically independent routes through a metabolic network; rather, they are the edges of the high-dimensional convex solution space of a biochemical network, and as such are the convex basis vectors [5,13]. The elementary modes are a superset of the extreme pathways, including additional network pathways that meet specific criteria (property III in the elementary mode section above). Thus, the number of extreme pathways is less than or equal to the number of elementary modes.

Decomposing a flux vector

All valid steady-state flux distributions (valid solutions to Equations 1 and 2) can be represented as nonnegative linear combinations of the extreme pathways or elementary modes. The relationship between a given flux vector, $\mathbf{v}$, and the set of network-based pathways, P, can be described by:

$$P\boldsymbol{\alpha} = \mathbf{v}; \quad \alpha_i \geq 0 \tag{3}$$

where P is a matrix whose columns are the network-based pathways, $\mathbf{v}$ is a given flux distribution, and $\boldsymbol{\alpha}$ is a vector of weights on the corresponding columns (extreme pathways or elementary modes) of P [14]. The vector $\mathbf{v}$ comprises internal and exchange fluxes; the vector $\boldsymbol{\alpha}$ need not be unique, and the term 'α-spectrum' refers to the range of weights on the columns of P that can be used to reconstruct a particular steady-state flux distribution, $\mathbf{v}$, in a network [14]. The range for the allowable weights on a given extreme pathway or elementary mode is not necessarily independent of the range for the allowable weights on another given extreme pathway or elementary mode in such a reconstruction.

Because the systemically dependent elementary modes are nonnegative linear combinations of the extreme pathways, Equation 3 can be used to describe the relationship between the two sets of network-based pathways. Thus, if $\mathbf{v} = \mathbf{e}_i$, then there are many possible values in $\boldsymbol{\alpha}$ that satisfy Equation 3. Conversely, if $\mathbf{v} = \mathbf{p}_i$, then only $\alpha_i = 1$ and all other values of $\boldsymbol{\alpha}$ are zero $[\alpha_j \ (j \neq i) = 0]$, owing to property III in the definition of extreme pathways.

The human red blood cell

The elementary modes have been previously calculated for a reduced RBC metabolic network [15]. Recently, a full metabolic network for the RBC has been studied with extreme pathways [14]. In the more complete network, all of the reactions are elementally balanced. There are reversible exchange fluxes for the metabolites carbon dioxide, water, ammonia, adenine, adenosine, inosine, phosphate, ATP, ADP, NADPH, NADP, NADH, NAD, hydrogen, 2,3-diphosphoglycerate (2,3-DPG), pyruvate and lactate, which reflect the possible physiological states of this network.

The reversible exchange of cofactors (e.g. ATP and ADP) does not necessarily represent physical transport of the compound between the extracellular environment and the cytoplasmic compartment but rather represents movement across a defined network boundary. The cofactors might be used in various non-metabolic functions of the cell such as cytoskeletal protein phosphorylation. Consequently, the exchange of cofactors represents a virtual exchange across the defined system boundaries.

Helicobacter pylori

A metabolic network of *H. pylori* has been reconstructed that accounts for 27% of the open reading frames with functional assignment on the genome [16]. This metabolic network represents 388 reactions and 403 metabolites, and it has been analyzed by flux balance and extreme pathway methods [16–18]. The extreme pathways have thus been computed and are available for analysis. All exchange fluxes are irreversible in this network. Minimal reaction sets, growth requirements, correlated reaction sets, pathway redundancy and other emergent properties have been analyzed for this network.

Results

Using Metatool [19] and FluxAnalyzer [20], we have calculated the elementary modes for the metabolic network models of *H. pylori* and the RBC for which the extreme pathways were previously calculated [17,21]. Thus, we can compare full sets of elementary modes and extreme pathways to illustrate the differences and similarities between the two sets. Below, the differences between these two network-based pathway definitions, as well as their relationship to each other, are discussed with

regard to mathematical and statistical characteristics and to biological interpretations.

Mathematical and statistical characteristics

For biochemical networks with only irreversible exchange fluxes, the set of elementary modes and the set of extreme pathways are equivalent. This equivalence can be readily seen in the sample system shown in Figure 2. If the metabolite A were solely an input and not allowed as an output, then ElMo 2, ElMo 3 and ElMo 4 would be the set of elementary modes and extreme pathways. The set of elementary modes and the set of extreme pathways for the *H. pylori* metabolic network are equivalent, as expected on the basis of previously advanced arguments [9].

The extreme pathways of the well-studied model of the RBC metabolic network were previously calculated and analyzed [21]. The dimension of the null space of the RBC stoichiometric matrix is 23 (i.e. there are 23 linear basis vectors). There are 55 extreme pathways for this network. We have now calculated 6180 elementary modes for this network. The presence of reversible exchange fluxes in the RBC results in non-equal sets of extreme pathways and elementary modes.

With these computational results at hand, some summary statements can be made regarding the relationship between the number of elementary modes and that of extreme pathways. Defining n as the dimension of the null space (the number of linear basis vectors), m as the number of extreme pathways, k as the number of elementary modes, and b_i as an exchange flux, then:

(I) elementary modes are a superset of extreme pathways and thus, typically, $n \leq m \leq k$;
(II) if b_i is irreversible for all i, then $m = k$;
(III) if b_i is not irreversible for all i, $m \ll k$ for some systems (e.g. the RBC).

The ratio of elementary modes to extreme pathways (k/m) could possibly be studied as a function of the number of reversible exchange fluxes and of internal network complexity and its topological characteristics. For the RBC, this ratio is $6180/55 \approx 112$.

Relationship between elementary modes and extreme pathways of RBC

The decomposition of elementary modes into extreme pathways can be studied by using Equation 3. As described above, any flux vector in the solution space for a metabolic network can be written as a nonnegative linear combination of the extreme pathways or elementary modes. This decomposition need not be unique. Because elementary modes are a superset of extreme pathways, the 6180 elementary modes for the RBC metabolic network can be described as nonnegative linear combinations of the extreme pathways. There are 39 type I extreme pathways (through pathways) that form this decomposition, because the 16 type III pathways (cyclic pathways) have zero flux by thermodynamic necessity (see Ref. [22] for a brief overview of the extreme pathway classification scheme and the thermodynamic characterization).

We calculated the number of elementary mode decompositions in which a given extreme pathway can participate (Figure 3). An extreme pathway can participate in the decomposition of up to 4353 elementary modes. The three extreme pathways that participate in the highest number of elementary mode decompositions are ExPa 12, ExPa 22 and ExPa 24: ExPa 22 involves the route from inosine uptake to adenosine secretion; ExPa 24 comprises the simple route from pyruvate uptake to lactate secretion; and ExPa 12 is a much more elaborate route involving the pentose phosphate pathway and glycolysis (Figure 4).

The three extreme pathways that participate in the lowest number of elementary mode decompositions are ExPa 5, ExPa 17 and ExPa 23: ExPa 5 is the classical route of glycolysis, which uses reactions from the 2,3-DPG shunt; ExPa 17 is another very interconnected route with reactions from glycolysis and the pentose phosphate pathway; and ExPa 23 is the reverse route of ExPa 24, describing the route from lactate uptake to pyruvate secretion (Figure 4).

As these example pathways show, there is not an obvious correlation between pathway simplicity and the number of times that a given extreme pathway can participate in the decomposition of elementary modes. This result emphasizes the complexity of such pathway analyses because a high level of participation in elementary mode decompositions is not simply an expansion around simple routes (e.g. ExPa 23 and ExPa 24), but can involve complex combinations of highly interconnected routes (e.g. ExPa 12 and ExPa 17).

We also evaluated the maximum number of extreme pathways that can be used in a given elementary mode decomposition (Figure 5). As expected, elementary modes that are equivalent to extreme pathways use only one extreme pathway in their decomposition: 38 elementary mode decompositions use one extreme pathway (note that there are 39 type I extreme pathways). The current implementations of the elementary mode algorithm [19,20] do not calculate two elementary modes for pathways comprising only reversible internal fluxes and

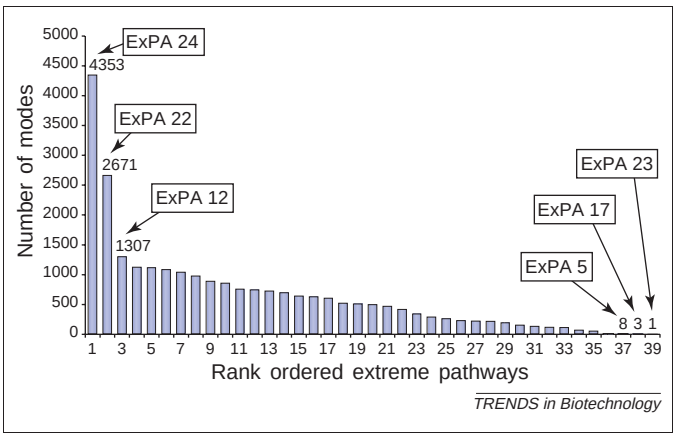

Figure 3. Number of elementary mode decompositions in which the given extreme pathway of the human red blood cell metabolic network can participate. The *y*-axis indicates the number of elementary modes for which the corresponding extreme pathway can be used in a decomposition; the *x*-axis lists the extreme pathways ordered according to the number of elementary modes in whose decomposition they can be used. Although it need not be the case for some flux vectors, the decomposition of 5515 of the elementary modes into extreme pathways is unique (not shown).The three extreme pathways that can participate in the most and the least number of elementary mode decompositions are labeled. As this figure illustrates, there is a non-obvious and complex relationship between elementary modes and extreme pathways.

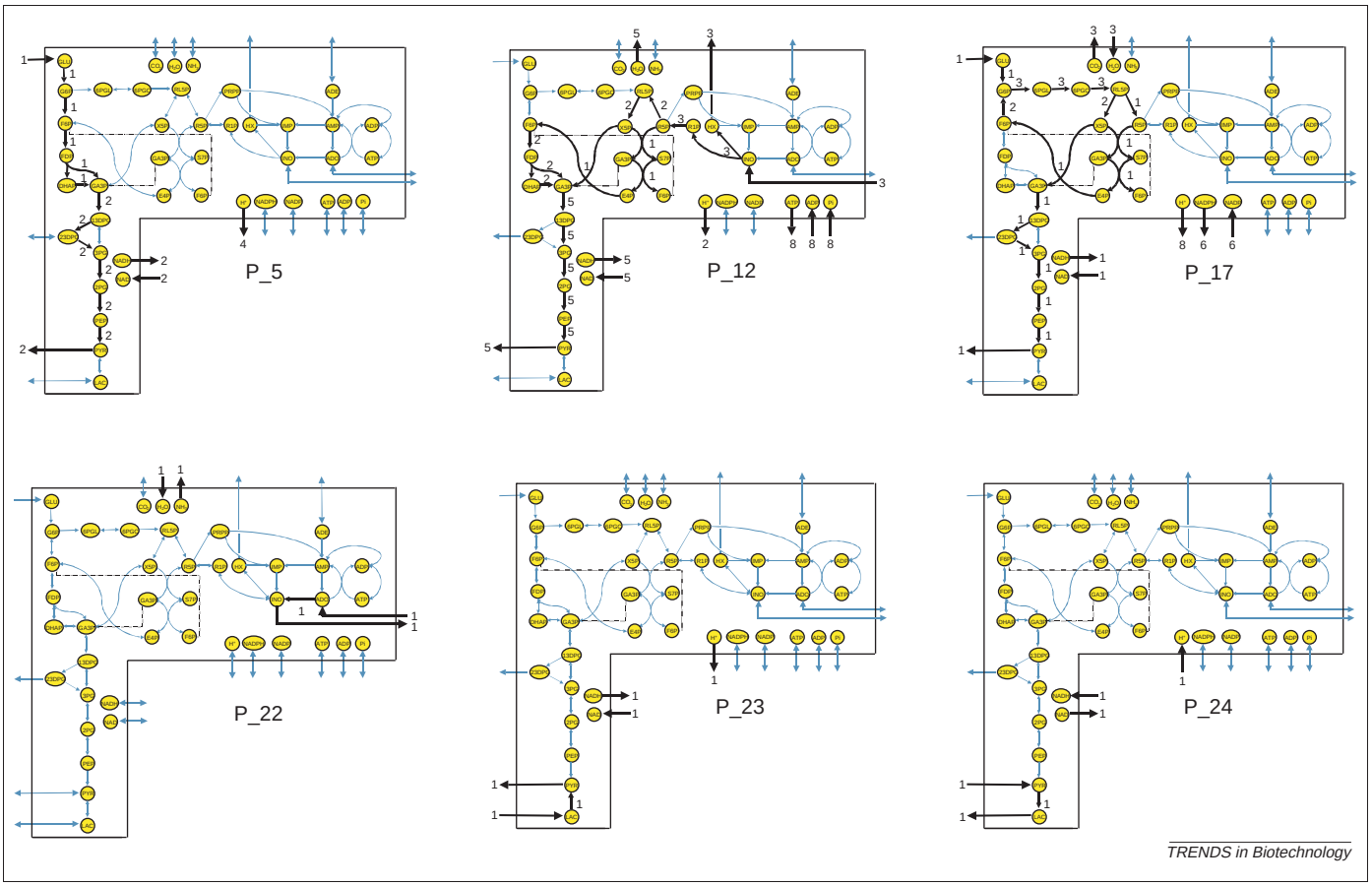

Figure 4. Maps of the six extreme pathways of the human red blood cell metabolic network that are labeled in Figure 3. These six maps correspond to the extreme pathways reported in Ref. [21]. The complex and non-obvious relationship between elementary modes and extreme pathways is demonstrated by the high number of decompositions in which pathway 12 participates (see Figure 3).

reversible exchange fluxes (e.g. only one of ExPas 23 and ExPas 24 is generated from the elementary mode program because they are the reverse of each other). We found that the average number of extreme pathways used in a given elementary mode decomposition is four, and some elementary mode decompositions require up to eight extreme pathways. These results again emphasize the combinatorial explosion that occurs as extreme pathways are combined to represent genetically independent elementary modes.

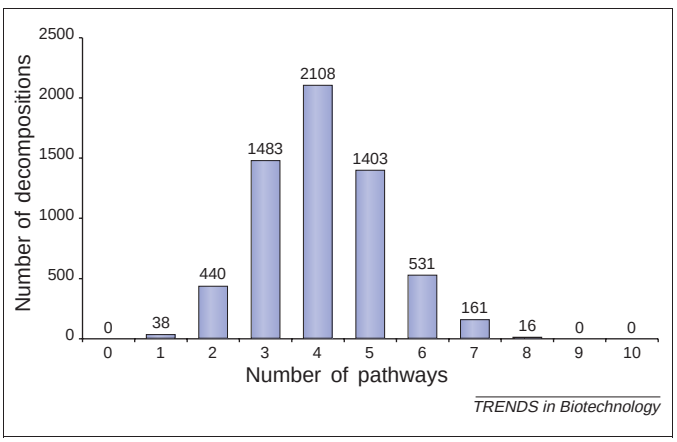

Figure 5. Maximum number of extreme pathways of the human red blood cell metabolic network that can be used in a given decomposition of an elementary mode. The *x*-axis indicates the number of extreme pathways that can be used in a given elementary mode decomposition; the *y*-axis indicates the number of elementary mode decompositions. For example, the decomposition of each of 2108 elementary modes can use up to four extreme pathways. The decomposition is not necessarily unique. For 5515 of the 6180 elementary modes, however, the decomposition into the extreme pathways is unique (not shown).

Interpretation of emergent properties

The elementary modes and extreme pathways of biochemical networks can describe important network properties [8]. They have been used to calculate product yields, to evaluate pathway redundancy, to determine correlated reaction sets, to calculate minimal reaction sets and to assess the effects of gene deletions, among other types of analysis (see Figure 1).

Given the differences between elementary modes and extreme pathways discussed above, some of these biological properties require special interpretation. For example, in the network shown in Figure 2, two elementary modes and one extreme pathway describe the synthesis of component B from component C. Consequently, both approaches could result in different descriptions of network properties, such as the number of pathways that connect inputs to outputs (related to pathway redundancy) and the yield of products from given substrates. The biological property of interest must be interpreted with reference to the method used to generate the network-based pathways.

Discussion

Elementary modes and extreme pathways are related network-based approaches that can be used to study biochemical networks. Although their development is still in its infancy, both methods have been used to study the properties of biochemical networks [8,12]. Here, we have applied the two approaches to metabolic networks of biologically meaningful size. The key results from this comparison are that (i) the elementary modes and extreme pathways are equivalent when all exchange fluxes are irreversible [9], as demonstrated for the *H. pylori* metabolic model; (ii) the number of elementary modes for the human RBC metabolic network (with reversible and irreversible exchange fluxes) is significantly greater than the number of extreme pathways; (iii) the elementary modes that are additional to the extreme pathways are non-obvious and highly complex combinations of the basis vectors (extreme pathways); and (iv) the interpretation of system properties obtained with extreme pathways requires careful application, owing to the possible cancellation of reversible exchange fluxes in such computations.

Each approach to pathway analysis has its advantages and disadvantages and, as we begin to understand these differences, these methods can be used appropriately to study the properties of biochemical networks. The extreme pathways of a network can be fewer in number than the elementary modes. They also correspond to the edge representation of cones. Being basis vectors, the extreme pathways might have to be added together to represent a particular flux state that cancels out a reversible exchange flux. Such occurrences complicate the full evaluation of network properties, such as pathway redundancy and product yield. The elementary modes do not have this shortcoming, but instead have a much larger set of vectors to account for the absence of a reversible exchange flux.

Calculating the elementary modes and extreme pathways represents an NP-hard computational problem (i.e. the running time for their computation increases at least exponentially with system size) and is thus difficult to practice on a genome scale [8]. Under such conditions, linear optimization [23] can be used to identify the particular edges of a polytope on the basis of a stated property (represented as an objective function). Typically, there can be many edges with identical properties, and mixed-integer linear programming (MILP) procedures [24] must be used to identify them. In addition, subdividing the reconstructed biochemical network into conceptually [25] and algorithmically [3] defined modules has been used to address the computational problem of calculating large sets of network-based pathways.

Network-based pathway definitions are likely to become more important with our ability to reconstruct genome-scale networks. Increasing effort is being expended in reconstructing such networks and subsequently analyzing them. We can thus anticipate more progress with network-based pathway definitions over the coming years.

Acknowledgements

We thank the Whitaker Foundation (Graduate Research Fellowship to J.P.) for financially supporting this work.

References

1 Covert, M.W. and Palsson, B.O. (2002) Transcriptional regulation in constraints-based metabolic models of Escherichia coli. *J. Biol. Chem.* 277, 28058–28064
2 Price, N.D. *et al.* (2003) Genome-scale microbial in silico models: the constraints-based approach. *Trends Biotechnol.* 21, 162–169
3 Schuster, S. *et al.* (2002) Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumoniae*. *Bioinformatics* 18, 351–361
4 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
5 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
6 Rockafellar, R.T. (1970) *Convex analysis*, Princeton University Press
7 Schilling, C.H. *et al.* (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* 15, 296–303
8 Papin, J.A. *et al.* (2003) Metabolic pathways in the post-genome era. *Trends Biochem. Sci.* 28, 250–258
9 Klamt, S. and Stelling, J. (2003) Two approaches for metabolic pathway analysis? *Trends Biotechnol.* 21, 64–69
10 Palsson, B.O. *et al.* (2003) Development of network-based pathway definitions: the need to analyze real metabolic networks. *Trends Biotechnol.* 21, 195–198
11 Clarke, B.L. (1988) Stoichiometric network analysis. *Cell Biophys.* 12, 237–253
12 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
13 Nuno, J.C. *et al.* (1997) Network organization of cell metabolism: monosaccharide interconversion. *Biochem. J.* 324, 103–111
14 Wiback, S.J. *et al.* (2003) Reconstructing metabolic flux vectors from extreme pathways: defining the α-spectrum. *J. Theor. Biol.* 224, 313–324
15 Schuster, S. *et al.* (1998) Elementary modes analysis illustrated with human red cell metabolism. In *BioThermoKinetics in the Post Genomic Era* (Larsson, C. *et al.*, eds), pp. 332–339, Chalmers
16 Schilling, C.H. *et al.* (2002) Genome-scale metabolic model of *Helicobacter pylori* 26695. *J. Bacteriol.* 184, 4582–4593
17 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
18 Papin, J.A. *et al.* (2002) Extreme pathway lengths and reaction participation in genome-scale metabolic networks. *Genome Res.* 12, 1889–1900
19 Pfeiffer, T. *et al.* (1999) METATOOL: for studying metabolic networks. *Bioinformatics* 15, 251–257
20 Klamt, S. *et al.* (2003) FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics* 19, 261–269
21 Wiback, S.J. and Palsson, B.O. (2002) Extreme pathway analysis of human red blood cell metabolism. *Biophys. J.* 83, 808–818
22 Price, N.D. *et al.* (2002) Extreme pathways and Kirchhoff's second law. *Biophys. J.* 83, 2879–2882
23 Chvatal, V. (1983) *Linear Programming*, W.H. Freeman
24 Lee, S. *et al.* (2000) Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comput. Chem. Eng.* 24, 711–716
25 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