BIOINFORMATICS

Vol. 15 no. 3 1999
Pages 251–257

METATOOL: for studying metabolic networks

T. Pfeiffer[1,2], I. Sánchez-Valdenebro[2], J.C. Nuño[2,3], F. Montero[2] and S. Schuster[4]

[1]Humboldt University Berlin, Institute of Biology, Invalidenstraße 42, D-10115 Berlin, Germany, [2]Universidad Complutense de Madrid, Departamento de Bioquímica y Biología Molecular I, Grupo de Biofísica, Ciudad Universitaria, E-28040 Madrid, Spain, [3]Universidad Politecnica de Madrid, Departamento de Matemática Aplicada a los Recursos Naturales, ETSI de Montes, E-28040 Madrid, Spain and [4]Max Delbrück Center for Molecular Medicine, Department of Bioinformatics, D-13092 Berlin-Buch, Germany

Received on October 6, 1998; revised on November 30, 1998; accepted on December 22, 1998

Abstract

Motivation: *To reconstruct metabolic pathways from biochemical and/or genome sequence data, the stoichiometric and thermodynamic feasibility of the pathways has to be tested. This is achieved by characterizing the admissible region of flux distributions in steady state. This region is spanned by what can be called a convex basis. The concept of 'elementary flux modes' provides a mathematical tool to define all metabolic routes that are feasible in a given metabolic network. In addition, we define 'enzyme subsets' to be groups of enzymes that operate together in fixed flux proportions in all steady states of the system.*

Results: *Algorithms for computing the convex basis and elementary modes developed earlier are briefly reviewed. A newly developed algorithm for detecting all enzyme subsets in a given network is presented. All of these algorithms have been implemented in a novel computer program named METATOOL, whose features are outlined here. The algorithms are illustrated by an example taken from sugar metabolism.*

Availability: *METATOOL is available from ftp://bmsdarwin.brookes.ac.uk/pub/software/ibmpc/metatool*

Supplementary information: *http://www.biologie.hu-berlin.de/biophysics/Theory/tpfeiffer/metatool.html*

Contact: *Tpfeiffer@bp.biologie.hu-berlin.de; Schuster@bp.biologie.hu-berlin.de*

Introduction

The sequencing of complete genomes results in an important increase in information that can and should be translated into knowledge about metabolic pathways. However, the function of only about half the open reading frames (ORFs) has been unambiguously identified so far (cf. Oliver, 1996). In cases where methods for function prediction based on sequence comparison (Tatusov *et al*., 1996; Huynen *et al*., 1997; Bork *et al*., 1998) turn out not to give sufficiently complete or certain results, it is promising to go beyond the purely genetic approach and consider the gene products in their interplay with each other, e.g. within metabolic networks. Whereas efficient tools for comparing gene sequences are used extensively nowadays, tools to test the viability of the deduced metabolic network are not so common. Various approaches have been proposed (Fell, 1990; Mavrovouniotis *et al*., 1990; Schuster and Hilgetag, 1994; Alberty, 1996; Nuño *et al*., 1997; Stephanopoulos and Simpson, 1997), all of them centred around the properties of metabolic networks at steady state imposed by their stoichiometric structure. In particular, the approaches of Mavrovouniotis *et al*. (1990) and Schuster and Hilgetag (1994) involve algorithms for detecting all the simplest stoichiometrically feasible pathways, and have been successfully applied to various systems of carbohydrate and amino acid metabolism (Liao *et al*., 1996; Schuster *et al*., 1999).

Any set of biochemical transformations can be described mathematically by a system of ordinary differential equations (see, for instance, Reder, 1988):

$$\frac{d\mathbf{X}}{dt} = \mathbf{N}\mathbf{v}(\mathbf{X}) \qquad (1.1)$$

where $\mathbf{N}$, $\mathbf{v}$ and $\mathbf{X}$ denote the stoichiometry matrix, the vector of reaction rates and the vector of concentrations of 'internal' metabolites (i.e. metabolites with variable concentrations), respectively. In contrast, metabolites are referred to as external if their concentrations are buffered. At stationary states, equation (1.1) simplifies to:

$$\mathbf{N}\,\mathbf{v} = 0 \qquad (1.2)$$

We decompose the flux vector into two subvectors, $\mathbf{v}_{rev}$ and $\mathbf{v}_{irr}$, which include the fluxes of the reversible and irreversible reactions, respectively. This implies:

$$\mathbf{v}_{irr} \geq 0 \qquad (1.3)$$

Note that it is not necessary to split the reversible reactions into forward and reverse steps.

To decide whether a given group of enzymes (which might have been detected by function assignment of a set of ORFs) forms a functionally coherent set in metabolism [cf. the concept of activity set proposed by Nuño *et al*. (1997)], it should be tested whether the flux vectors realizable by these enzymes can obey relationships (1.2) and (1.3). To this end, it is necessary to know the structure of the region of admissible flux vectors. Although this region involves an infinite number of vectors, it can be shown that it is spanned by a finite number of generating vectors (Rockafellar, 1970; Nozicka *et al*., 1974). In applications to chemistry and biochemistry, these generating vectors can be interpreted as meaningful transformation routes (Clarke, 1980; Schuster and Schuster, 1993; Nuño *et al*., 1997).

The generating vectors can be chosen so as to have a simplicity property in that as many components as possible are zero. This is of interest because the set of identified enzyme genes is incomplete in most cases, so that it is sensible to test whether at least one functional metabolic route is realized by this set. This leads in a natural way to the concept of elementary mode (Schuster and Hilgetag, 1994; Schuster *et al*., 1996), which is defined as a minimal set of enzymes that could operate at steady state. The enzymes are weighted with the fractional flux they carry. In many situations, more elementary modes exist than are needed to construct all admissible flux distributions. Therefore, some of them can be taken as a generator set of the whole admissible region. This will be discussed in more detail in the present paper.

The network structure frequently implies that the interplay of enzymes is so tight that several enzymes always operate together in fixed flux proportions. To describe this phenomenon, we introduce and illustrate here the concept of enzyme subsets.

Even in metabolic networks of moderate complexity, it is difficult or even impossible to find the basis vectors, elementary modes and enzyme subsets by inspection of the reaction scheme. Thus, automated methods are needed. Here, we present a novel computer program, named METATOOL, performing these methods.

The structure of the region of admissible fluxes

The complete set of vectors **v** satisfying equation (1.2) defines a region called the null-space of **N** (cf. Groetsch and King, 1988). Its dimension is given by the difference of the number of reactions and the rank of **N**. The null-space can be described mathematically by a matrix, **K**, whose columns are linearly independent vectors spanning this subspace (i.e. they form a basis). Accordingly, this matrix fulfils the equation:

$$\mathbf{N}\,\mathbf{K} = 0 \qquad (2.1)$$

Owing to inequality (1.3), the mathematical problem is beyond the scope of linear algebra. In convex analysis, it is shown that the region of all vectors **v** satisfying relationships (1.2) and (1.3) is a convex polyhedral cone, ***F*** (Rockafellar, 1970). This cone can be conceived of as spanned by generating vectors. Two types of such vectors can be distinguished: vectors $\mathbf{f}_k$ ('irreversible' vectors) for which the negative, $-\mathbf{f}_k$, is not contained in ***F***, and vectors $\mathbf{b}_i$ ('reversible' vectors) for which the negative, $-\mathbf{b}_i$, is situated in ***F***. Thus, we can write:

$$\mathbf{F} = \left\{ \mathbf{v} \in \mathbb{R}^r : \mathbf{v} = \sum_k \lambda_k \mathbf{f}_k + \sum_i \beta_i \mathbf{b}_i,\ \lambda_k, \beta_i \in \mathbb{R}, \lambda_k \geq 0 \right\} \qquad (2.2)$$

where r denotes the number of reactions in the system.

If all reactions are irreversible and, in some cases, even if some reactions are reversible, no vectors $\mathbf{b}_i$ occur. Then, the cone ***F*** is pointed, i.e. there is no pair of vectors that are opposite to each other, so that they would make an angle of 180°. This implies that the set of generating vectors is uniquely determined up to multiplication by scalars (cf. Rockafellar, 1970). If not all of the reactions are irreversible, vectors $\mathbf{b}_i$ may occur (see Figure 1B). These are situated on straight lines, planes or hyperplanes confining ***F***. In this case, these vectors, as well as the vectors $\mathbf{f}_k$, are not always uniquely determined. As the coefficients λ_k in equation (2.2) are restricted to be non-negative, the cone ***F*** can be considered as a convex combination of the vectors $\mathbf{f}_k$, $\mathbf{b}_i$ and $-\mathbf{b}_i$. Accordingly, we will call the set of these vectors a convex basis of ***F***. Their number (which may be interpreted as the algebraic dimension) may be greater than the geometric dimension of the cone.

An algorithm for constructing a convex basis has been given by Nozicka *et al*. (1974). It is based on the Gauss–Jordan method (cf. Groetsch and King, 1988). A number of consecutive tableaux are computed by row operations, starting with (**I** denotes the identity matrix):

$$\mathbf{T}^{(0)} = \begin{pmatrix} \mathbf{N}_{\text{rev}}^{\text{T}} & \\ \mathbf{N}_{\text{irr}}^{\text{T}} & \mathbf{I} \end{pmatrix} \qquad (2.3)$$

where $\mathbf{N}_{\text{rev}}^{\text{T}}$ and $\mathbf{N}_{\text{irr}}^{\text{T}}$ are the transposed submatrices of the stoichiometric matrix corresponding to reversible and irreversible reactions, respectively, after a renumbering of reactions. In each step of the algorithm, one column of the left-hand side of the tableau (originating from the transposed stoichiometry matrix) is considered. All rows containing a zero in that position are transferred to the next tableau. Then, the rows of $\mathbf{N}_{\text{rev}}^{\text{T}}$ are combined with each other and with the 'irreversible' rows according to Gauss. If no combinations of 'reversible' rows are possible because they all contain zeros in the respective positions, 'irreversible' rows are combined with each other. For these, however, only non-negative linear combination is allowed in order to meet the irreversibility constraint (1.3). Moreover, the pair of rows to be combined

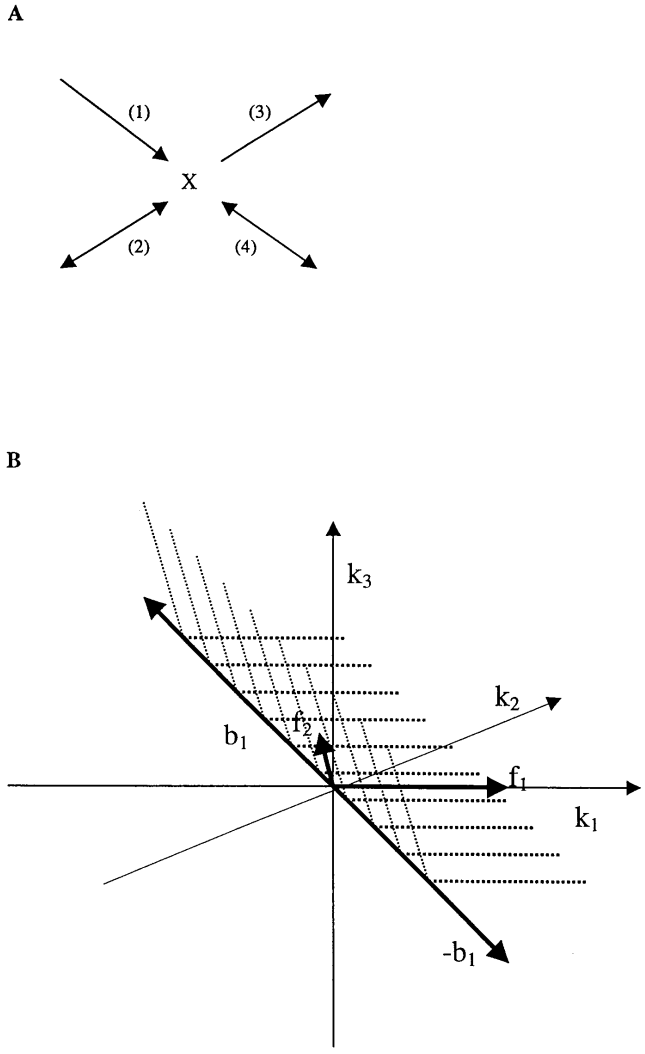

Fig. 1. (**A**) Simple reaction system. Reactions (1) and (3) are irreversible, and reactions (2) and (4) are reversible. X stands for a metabolite. The stoichiometry matrix of this system reads $\mathbf{N} = (1\ 1\ -1\ -1)^T$. (**B**) Convex flux cone belonging to the system shown in (A). The null-space spanned by the three vectors $\mathbf{k}_1 = (1\ -1\ 0\ 0)^T$, $\mathbf{k}_2 = (1\ 0\ 1\ 0)^T$ and $\mathbf{k}_3 = (1\ 0\ 0\ 1)^T$ is taken as the co-ordinate system, which is depicted as an orthogonal system for simplicity's sake. $\mathbf{b}_1 = (0\ 1\ 0\ 1)^T$ is a 'reversible' generating vector; $\mathbf{f}_1 = (1\ -1\ 0\ 0)^T$ and $\mathbf{f}_2 = (0\ 1\ 1\ 0)^T$ are 'irreversible' generating vectors. The cone is here non-pointed and has the shape of a wedge.

must meet a simplicity constraint expressed in terms of the index sets of positions where the respective row contains zeros, $S(\cdot)$:

$$S\left(\mathbf{t}_i^{(j)}\right) \cap S\left(\mathbf{t}_k^{(j)}\right) \not\subseteq S\left(\mathbf{t}_m^{(j)}\right) \text{ for all } m \neq i, k \qquad (2.4)$$

where $\mathbf{t}_i^{(j)}$ is the ith row of the left-hand side part of the jth tableau. Formally, $S(\cdot)$ is defined as:

$$S(\mathbf{v}) = \{i : v_i = 0\} \qquad (2.5)$$

Elementary flux modes

By definition, the generating vectors (edges) of a pointed convex polyhedral cone cannot be decomposed into two other vectors belonging to the cone:

$$\mathbf{f}_k \neq \lambda_1 \mathbf{v}^{(1)} + \lambda_2 \mathbf{v}^{(2)},\ \lambda_1, \lambda_2 \geq 0 \qquad (3.1)$$

If all reactions are irreversible, we have $\mathbf{v}^{(1)}, \mathbf{v}^{(2)} \geq 0$. Therefore, it can be easily shown that the generating vectors have a simplicity property in that they encompass as many zeros as possible. 'Simplicity' is meant to imply that when only the enzymes belonging to this set are operating, then complete inhibition of one of these enzymes impedes any steady-state flux in the system. To phrase this in a mathematical way, we can use the index set defined in equation (2.5). A generating vector, $\mathbf{f}_k$, of a pointed convex cone cannot be decomposed into vectors $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$ [cf. relationship (3.1)] with:

$$S(\mathbf{f}_k) \subset S(\mathbf{v}^{(1)}) \text{ and } S(\mathbf{f}_k) \subset S(\mathbf{v}^{(2)}) \qquad (3.2)$$

This property is very interesting from the biochemical point of view because the generating vectors then correspond to simple metabolic pathways. A flux vector $\mathbf{v} = (v_1, v_2, \ldots, v_r)$ with some of its components being zero can be interpreted as a set of enzymes and, hence, as a route through the system, weighted with the fluxes carried by the enzymes. In other words, generating vectors are stoichiometrically feasible transformations, from a given substrate to a product (which could be the same compound as the substrate if the pathway is cyclic), that cannot be decomposed into simpler ones (see, for instance, the 'Example' section). By adapting this property to systems involving both irreversible and reversible reactions, and taking into account the scaling indeterminacy of generating vectors, the concept of elementary flux modes has been defined (Schuster and Hilgetag, 1994).

An elementary flux mode can be defined as a class of flux vectors with a representative $\mathbf{v}^*$:

$$M = \{\mathbf{v} : \mathbf{v} = \mu\mathbf{v}^*,\ \mu \in \mathbf{IR},\ \mu \geq 0\} \qquad (3.3)$$

such that $\mathbf{v}^*$ fulfils relationships (1.2) and (1.3), and cannot be decomposed into vectors $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$ [cf. relationship (3.1)] which contain zero components wherever $\mathbf{v}^*$ does and in at least one additional position each [cf. relationship (3.2)].

The elementary modes of a biochemical reaction network are uniquely determined. Moreover, they span the admissible region by convex combination. This can be proved as follows. Assume that there is a vector $\mathbf{v}$ fulfilling the steady-state condition (1.2) and sign restriction (1.3) that does not represent an elementary mode itself and cannot be decomposed into such modes. Vector $\mathbf{v}$ then would not fulfil relationship (3.1). If $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$ represent elementary modes, the assumption of the indirect proof is falsified. If not, $\mathbf{v}^{(1)}$ or $\mathbf{v}^{(2)}$, or both, could be further decomposed, and so on. Since

this consecutive decomposition implies a permanent increase in the number of zero components, it would eventually end with a set of vectors that cannot be further decomposed, i.e. with elementary modes.

The above result implies that a convex basis can be formed by appropriately chosen elementary modes. In general, it can be shown that the number of elementary modes is never less than the number of vectors necessary to span the cone. Consider, for example, the system shown in Figure 1A. The vectors (0 1 0 1), (1 –1 0 0), (0 1 1 0), (1 0 1 0) and (1 0 0 1) represent elementary modes. However, only three of them, e.g. (0 1 0 1) as a 'reversible' vector and (1 –1 0 0) and (0 1 1 0) as 'irreversible' vectors, are sufficient to form a convex basis (Figure 1B).

To compute all elementary modes, one must modify the algorithm outlined in the previous section in that all pairwise combinations of rows are considered, so that all the simplest modes are computed irrespective of whether they are dependent on each other with respect to convex combination. Moreover, a test condition similar to equation (2.4) has to be applied to all row combinations. A detailed presentation of the algorithm was given earlier (Schuster *et al*., 1996).

Enzyme subsets

When studying gene expression or metabolic regulation, it is of interest to detect groups of enzymes that, in all steady states of the system, operate together in fixed flux proportions. We will call these groups 'enzyme subsets'. An illustrative example is given by a branched pathway structure such as that occurring in the synthesis of many amino acids. For example, the synthesis pathways of threonine, lysine and methionine all start from aspartate, with the branch leading to lysine bifurcating at aspartate semialdehyde, and the branches leading to threonine and methionine bifurcating at homoserine (cf. Lehninger, 1982). The enzymes in any one branch carry the same steady-state flux. Thus, it is likely that they are expressed simultaneously.

In Metabolic Control Analysis, the concept of a monofunctional unit was previously introduced (Kholodenko *et al*., 1995; Rohwer *et al*., 1996). This refers to a group of enzymes that control the metabolic processes outside of the unit in a coherent way so that control coefficients can be assigned not only to the single enzymes, but to the unit as a whole. Thus, this unit can in fact be considered as a 'super-enzyme'. It was shown that a group of enzymes can be considered as a monofunctional unit if it fulfils three conditions (Rohwer *et al*., 1996):

(i) there is only one independent flux going through the unit;
(ii) there are no conservation relationships linking metabolites inside of the unit with metabolites outside;
(iii) there are no allosteric or other regulatory effects of metabolites inside of the unit on reactions outside.

Condition (i) is essentially the definition of the enzyme subsets. More exactly, a group of enzymes is an enzyme subset if for any two members E_i and E_k of the set, the following two conditions are fulfilled: in all flux vectors **v** satisfying the steady-state condition (1.2), the ratio v_i/v_k has the same non-zero value, and the orientations of the irreversible reactions involved do not contradict each other. From the properties of the null-space of **N**, it follows immediately that this condition can alternatively be written as follows: in all column vectors $\mathbf{k}^{(j)}$ of **K**, the ratio $\mathbf{k_i}^{(j)}/\mathbf{k_k}^{(j)}$ has the same non-zero value. This leads us to an algorithm for detecting the enzyme subsets.

1. Detect all row vectors of **K** that are null vectors.
2. Normalize each of the remaining row vectors of **K** by dividing by its greatest common divisor.
3. Compare any normalized row vector with any other. If they are the same and there are no contradictions in the directionalities of irreversible reactions, the corresponding reactions belong to the same subset. The quotient of the normalization factors gives the flux ratio.

The row vectors detected in step 1 correspond to strictly detailed balanced reactions (Schuster and Schuster, 1991), i.e. their net velocities are zero in any steady state of the system. In step 3, it is not actually necessary to compare all rows with each other. Instead, one can arrange this comparison in a similar way as in the Gaussian elimination method so that only $n(n-1)/2$ comparison steps have to be made.

Implementation

In order to carry out the methods of pathway analysis outlined above, a computer program named METATOOL was recently developed by T.Pfeiffer in co-operation with the other co-authors of the present paper. The program was written in C (compiled with GNU gcc) and runs on UNIX and MS-DOS platforms. It is, together with a help file and an example file, freely available over the Internet from ftp://bmsdarwin.brookes.ac.uk/pub/software/ibmpc/metatool. The program parses ASCII files including the list of reversible enzymes, irreversible enzymes, internal metabolites, external metabolites and reaction equations. To cope with the changing number of rows of the tableau in the course of the algorithm, pointer variables are used. The output consists of an ASCII file including:

- the numbers of reactions and internal and external metabolites;
- the stoichiometry matrix;
- the null-space matrix;
- the enzyme subsets given both in the form of a matrix and as a list of enzyme names, including the informa-

tion about whether these subsets correspond to a reversible or irreversible transformation;
- the overall stoichiometries of the enzyme subsets;
- the stoichiometry matrix of the reduced reaction system (with the enzyme subsets taken as combined reactions);
- a convex basis both in the form of a matrix and as a list of enzyme names, including the information about reversibility/irreversibility of the basis vectors;
- the overall stoichiometries (in terms of the external metabolites) corresponding to the vectors forming this convex basis;
- the elementary modes given both in the form of a matrix and as a list of enzyme names, including the information about reversibility/irreversibility;
- the overall stoichiometries of the elementary modes in terms of the external metabolites.

The routine involved in METATOOL for computing the elementary modes has also been included in the recent version of the simulation package GEPASI (Mendes, 1997).

The computation time is shortened enormously by computing the elementary modes and convex basis from a reduced reaction system after taking advantage of a model reduction using the enzyme subsets. Nevertheless, the modes and basis are given, in the output, in terms of the original, non-reduced system. In all the simulations we have carried out with systems comprising ~50 metabolites and 50 reactions, for example, running time is <5 s on a UNIX machine. However, this time can increase considerably if the structure of the system is so complex (e.g. a highly entangled network or one with many isoenzymes) that a combinatorial explosion of modes occurs.

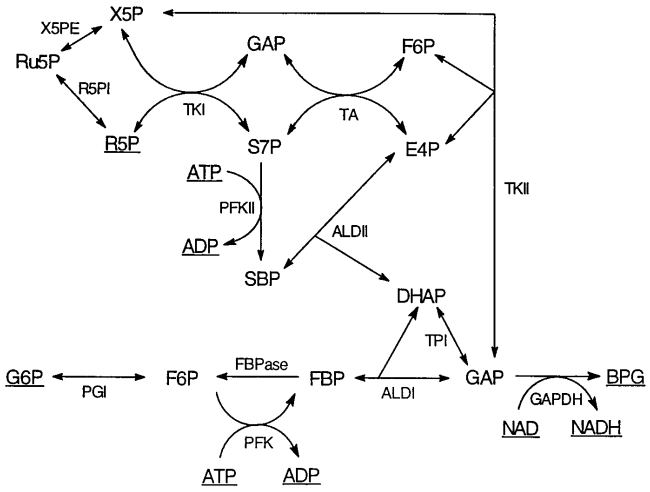

Fig. 2. Reaction scheme including the upper part of glycolysis and the non-oxidative pentose phosphate pathway. Abbreviations of enzymes: ALDI, function of aldolase acting on FBP; ALDII, function of aldolase acting on SBP; FBPase, fructose 1,6-bisphosphatase; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; PFKI, function of phosphofructokinase acting on F6P; PFKII, function of phosphofructokinase acting on S7P; PGI, phosphoglucoisomerase; R5PI, ribose 5-phosphate isomerase; TA, transaldolase; TKI, transketolase I; TKII, transketolase II; TPI, triose-phosphate isomerase; X5PE, xylulose 5-phosphate epimerase. Abbreviations of metabolites: BPG, 1,3-bisphosphoglycerate; DHAP, dihydroxyacetone phosphate; E4P, erythrose 4-phosphate; FBP, fructose 1,6-bisphosphate; F6P, fructose 6-phosphate; GAP, glyceraldehyde 3-phosphate; G6P, glucose 6-phosphate; R5P, ribose 5-phosphate; Ru5P, ribose 5-phosphate; SBP, sedoheptulose bisphosphate; S7P, sedoheptulose 7-phosphate; X5P, xylulose 5-phosphate. Underlined metabolites are considered external.

Example

Consider a reaction system involving several reactions of hexose and pentose metabolism as shown in Figure 2. The oxidative part of the pentose phosphate pathway (PPP) is considered to be absent, as is the case in several microorganisms such as *Methanococcus jannaschii* (Selkov *et al.*, 1997). We have taken into account that phosphofructokinase not only acts on fructose-6-phosphate, but also, with a lower affinity, on sedoheptulose-7-phosphate (cf. the ENZYME database, http://www.expasy.ch/sprot/enzyme.html). Likewise, the activities of aldolase in splitting both fructose-bisphosphate and sedoheptulose-bisphosphate are considered.

The following results are obtained straightforwardly by means of METATOOL. The dimension of the null-space is four. The enzyme activities of ribose 5-phosphate isomerase (R5PI), xylulose 5-phosphate epimerase (X5PE), transketolase I and II (TKI and TKII) form an enzyme subset with the flux proportions –2:2:1:1. The fact that transaldolase, which is situated in the scheme between TKI and TKII, does not belong to this subset, can hardly be seen by inspection. A further subset is {aldolase II, phosphofructokinase II}, and all other enzymes make up subsets with one member each.

The elementary modes of the system are given in Table 1. Mode 1 is the well-known fructose 1,6-bisphosphate cycle. Mode 2 represents another futile cycle, which is more complicated and hardly detectable by inspection. Mode 11 corresponds to the upper part of glycolysis. Interestingly, a similar transformation is realized by mode 6, using the alternative functions of phosphofructokinase and aldolase (PFKII and ALDII). The function of the non-oxidative part of the PPP is often stated as the transformation of six molecules of pentoses into five molecules of hexoses (Meléndez-Hevia *et al.*, 1994). This function is performed by modes 8 and 9, with the latter implying the hydrolysis of 2 ATP. The reverse transformation, accompanied by the hydrolysis of 1 ATP, is realized by modes 4 and 7. The remaining modes represent different conversions of ribose 5-phosphate into 1,3-bisphosphoglycerate (BPG) and glucose 6-phosphate or into BPG only. The multiplicity of modes with the same overall stoi-

chiometry results from the fact that ALDII together with PFKII performs exactly the same transformation as ALDI, TA and PFKI. In other words, the two glycolytic enzyme functions ALDI and PFKI can be bypassed via ALDII, PFKII and TA (with the latter used in the reverse direction).

As mentioned in the previous section 'Elementary flux modes', elementary modes are not always convexly independent. In the system considered here, a convex basis is formed by four vectors. Accordingly, algebraic and geometric dimensions coincide for this example. METATOOL computes a convex basis consisting of the elementary modes 3, 4, 7 and 8. All other modes can be obtained by convex combination of these. For example, the futile cycle represented by mode 1 is the sum of modes 7 and 8, and glycolysis is the sum of mode 7 and two times mode 3.

In several studies, the enzymes aldolase, transaldolase and transketolase were modelled on a more detailed level, by considering the elementary steps of enzyme–substrate binding and of dissociation of the enzyme–product complex (Kuchel *et al*., 1990; Nuño *et al*., 1997). Here, we have modelled these enzymes as overall reactions, but, alternatively, we have also applied that more detailed approach and found that the number of generating vectors, subsets and elementary modes does not change. For other systems, in which the elementary reactions are linearly dependent, however, these numbers do change.

Discussion

Characterizing the relationships between the genotype and metabolic phenotype requires efficient theoretical tools. One of these tools is the structural analysis of metabolic pathways, which manages without knowledge of kinetic parameters. Stoichiometric and thermodynamic constraints confine the region of admissible steady-state fluxes to be a convex polyhedral cone. The structure of this region can be characterized by vectors forming a convex basis. All flux distributions possible at steady state are superpositions of these basis vectors.

Another important concept in pathway analysis is that of elementary flux modes. These have a clear biochemical interpretation in that they correspond to simple (non-decomposable) metabolic pathways or cycles, as we have demonstrated here by a simple example. This fact can be used to define what a metabolic pathway means. Moreover, it may be speculated that some correspondence between elementary modes and different physiological functions of a biochemical network exists. The number of elementary modes in a system is a measure of the richness of its different metabolic functions. Moreover, the calculation of elementary modes is a means of determining optimal conversion yields in biotechnology (Liao *et al*., 1996; Schuster *et al*., 1999).

Table 1. Elementary modes of the reaction system shown in Figure 2. The enzymes are weighted with their fractional flux (unity if no number is given). Negative values indicate that the reaction is used in the reverse direction. All modes are irreversible. The meaning of the abbreviations is as in Figure 2

Mode	Overall stoichiometry	Enzymes involved
1	ATP = ADP	PFKI FBPase
2	ATP = ADP	–ALDI ALDII –TA PFKII FBPase
3	NAD + 3 R5P = NADH + BPG + 2 G6P	–2 PGI –2 R5PI 2 X5PE TKI TKII TA GAPDH
4	ATP + 5 G6P = ADP + 6 R5P	5 PGI ALDII TPI 4 R5PI –4 X5PE –2 TKI –2 TKII –3 TA PFKII
5	3 NAD + ATP + 3 R5P = 3 NADH + ADP + 3 BPG + G6P	–PGI ALDII TPI –2 R5PI 2 X5PE TKI TKII PFKII 3 GAPDH
6	2 NAD + ATP + G6P = 2 NADH + ADP + 2 BPG	PGI ALDII TPI –TA PFKII 2 GAPDH
7	ATP + 5 G6P = ADP + 6 R5P	5 PGI ALDI TPI 4 R5PI –4 X5PE –2 TKI –2 TKII –2 TA PFKI
8	6 R5P = 5 G6P	–5 PGI –ALDI –TPI –4 R5PI 4 X5PE 2 TKI 2 TKII 2 TA FBPase
9	2 ATP + 6 R5P = 2 ADP + 5 G6P	–5 PGI –3 ALDI 2 ALDII –TPI –4 R5PI 4 X5PE 2 TKI 2 TKII 2 PFKII 3 FBPase
10	NAD + ATP + 3 R5P = NADH + ADP + BPG + 2 G6P	–2 PGI –ALDI ALDII –2 R5PI 2 X5PE TKI TKII PFKII FBPase GAPDH
11	2 NAD + ATP + G6P = 2 NADH + ADP + 2 BPG	PGI ALDI TPI PFKI 2 GAPDH
12	5 NAD + 2 ATP + 3 R5P = 5 NADH + 2 ADP + 5 BPG	2 ALDII 2 TPI –2 R5PI 2 X5PE TKI TKII –TA PFKII 5 GAPDH
13	5 NAD + 2 ATP + 3 R5P = 5 NADH + 2 ADP + 5 BPG	2 ALDI 2 TPI –2 R5PI 2 X5PE TKI TKII TA 2 PFKI 5 GAPDH
14	5 NAD + 2 ATP + 3 R5P = 5 NADH + 2 ADP + 5 BPG	ALDI ALDII 2 TPI –2 R5PI 2 X5PE TKI TKII PFKI PFKII 5 GAPDH

In this paper, we have introduced the concept of enzyme subsets, by defining them as groups of enzymes that operate together in fixed flux proportions, in all steady states of the system. It can be assumed that the enzymes belonging to such a subset are expressed from the genome coherently and form functional units in terms of metabolic regulation.

The algorithms related to the above-mentioned concepts were implemented in a computer program METATOOL. It is aimed at complementing earlier software packages for simulation of biochemical systems [e.g. MetaModel by Cornish-Bowden and Hofmeyr (1991); GEPASI 3 by Mendes (1997)] by focusing solely on structural (topological) aspects of these systems. For the future, it is planned to extend the program so that reaction equations can be extracted automatically from databases available in the WWW.

Acknowledgements

This work has been supported by the Spanish government through projects PB95-0406 and BIO96-0895, and by the Deutsche Forschungsgemeinschaft.

References

Alberty,R.A. (1996) Calculation of biochemical net reactions and pathways using matrix methods. *Biophys. J.*, **71**, 507–515.

Bork,P., Dandekar,T., Diaz-Lazcoz,Y., Eisenhaber,F., Huynen,M. and Yuan,Y. (1998) Predicting function: From genes to genomes and back. *J. Mol. Biol.*, **283**, 707–725.

Clarke,B.L. (1980) Stability of complex reaction networks. In Prigogine,I. and Rice,S.A. (eds), *Advances in Chemical Physics*, Vol. 43. Wiley, New York, pp. 1–216.

Cornish-Bowden,A. and Hofmeyr,J.-H.S. (1991) MetaModel: a program for modelling and control analysis of metabolic pathways on the IBM PC and compatibles. *Comput. Applic. Biosci.*, **7**, 89–93.

Fell,D.A. (1990) Substrate cycles: theoretical aspects of their role in metabolism. *Comm. Theor. Biol.*, **6**, 1–14.

Groetsch,C.W. and King,J.T. (1988) *Matrix Methods and Applications*. Prentice-Hall, Englewood Cliffs, NJ.

Huynen,M.A., Diaz-Lazcoz,Y. and Bork,P. (1997) Differential genome display. *Trends Genet.*, **13**, 389–390.

Kholodenko,B.N., Schuster,S., Rohwer,J.M., Cascante,M. and Westerhoff,H.V. (1995) Composite control of cell function: Metabolic pathways behaving as single control units. *FEBS Lett.*, **368**, 1–4.

Kuchel,P.W., Berthon,H.A., Bubb,W.A., Bulliman,B.T. and Collins,J.G. (1990) Computer simulation of the pentose-phosphate pathway and associated metabolism used in conjunction with NMR experimental data from human erythrocytes. *Biomed. Biochim. Acta*, **49**, 757–770.

Lehninger,A.L. (1982) *Principles of Biochemistry*. Worth, New York.

Liao,J.C., Hou,S.-Y. and Chao,Y.-P. (1996) Pathway analysis, engineering and physiological considerations for redirecting central metabolism. *Biotechnol. Bioeng.*, **52**, 129–140

Mavrovouniotis,M.L., Stephanopoulos,G. and Stephanopoulos,G. (1990) Computer-aided synthesis of biochemical pathways. *Biotechnol. Bioeng.*, **36**, 1119–1132.

Meléndez-Hevia,E., Waddell,T.G. and Montero,F. (1994) Optimization of metabolism: the evolution of metabolic pathways toward simplicity through the game of the pentose phosphate cycle. *J. Theor. Biol.*, **166**, 201–220.

Mendes,P. (1997) Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3. *Trends Biochem. Sci.*, **22**, 361–363.

Nozicka,F., Guddat,J., Hollatz,H. and Bank,B. (1974) *Theorie der linearen parametrischen Optimierung*. Akademie-Verlag, Berlin.

Nuño,J.C., Sánchez-Valdenebro,I., Pérez-Iratxeta,C., Meléndez-Hevia,E. and Montero,F. (1997) Network organization of cell metabolism: monosaccharide interconversion. *Biochem. J.*, **324**, 103–111.

Oliver,S.G. (1996) From DNA sequence to biological function. *Nature*, **379**, 597–600.

Reder,C. (1988) Metabolic control theory: a structural approach. *J. Theor. Biol.*, **135**, 175–201.

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

Rohwer,J., Schuster,S. and Westerhoff,H.V. (1996) How to recognize monofunctional units in a metabolic system. *J. Theor. Biol.*, **179**, 213–228.

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

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

Schuster,R. and Schuster,S. (1993) Refined algorithm and computer program for calculating all non-negative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed. *Comput. Applic. Biosci.*, **9**, 79–85.

Schuster,S., Hilgetag,C., Woods,J.H. and Fell,D.A. (1996) Elementary modes of functioning in biochemical networks. In Cuthbertson,R., Holcombe,M. and Paton,R. (eds), *Computation in Cellular and Molecular Biological Systems*. World Scientific, Singapore, pp. 151–165.

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

Selkov,E., Maltsev,N., Olsen,G.J., Overbeek,R. and Whitman,W.B. (1997) A reconstruction of the metabolism of *Methanococcus jannaschii* from sequence data. *Gene*, **197**, GC11–GC26.

Stephanopoulos,G. and Simpson,T.W. (1997) Flux amplification in complex metabolic networks. *Chem. Eng. Sci.*, **52**, 2607–2627.

Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S., Borodowsky,M., Rudd,K.E. and Koonin,E.V. (1996) Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*. *Curr. Biol.*, **6**, 279–291.