ELSEVIER

Journal of Biotechnology 71 (1999) 207–223

JOURNAL OF Biotechnology

Analysis of the pathway structure of metabolic networks

Troy W. Simpson, Brian D. Follstad, Gregory Stephanopoulos *

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received 21 November 1997; accepted 10 November 1998

Abstract

Metabolic networks comprise a multitude of enzymatic reactions carrying out various functions related to cell growth and product formation. Although such reactions are occasionally organized into biochemical pathways, a formal procedure is desired to identify the independent pathways in a bioreaction network and the degree of engagement of each individual reaction in these pathways. We present a procedure for the identification of the independent pathways of bioreaction networks of any size and complexity. The method makes use of the steady-state internal metabolite stoichiometry matrix and defines the independent pathways through the reaction membership of its kernel matrix. Examples from the aromatic amino acid biosynthetic pathway and central carbon metabolism of cells in culture are provided to illustrate the method. Applications to the analysis of the control structure of bioreaction networks are also discussed.

Keywords: Metabolic flux analysis; Metabolic pathways

1. Introduction

Cellular metabolism is the sum total of anabolic and catabolic reactions responsible for the conversion of energy and carbon sources to metabolic energy, reducing equivalents, and carbon skeletons needed for the synthesis of cellular macromolecules and, ultimately, cell growth. Despite their complexity, metabolic networks possess certain features that make them amenable to systematic analysis. One such feature is their ability to attain steady state for all intermediate metabolites, implying an intrinsic balancing of individual reaction rates. Another feature is that the multitude of network reactions participate in carrying out certain tasks which are the outcome of the coordinated action of sets of reactions, usually called pathways. Although cellular reactions have been conveniently organized into pathways (e.g. glycolytic, anaplerotic, respiratory), such organization is somewhat arbitrary and not straightforward in large bioreaction networks with multiple input and output metabolites. Specifically, if a pathway is defined as a collection of reaction steps responsible for the conversion of a set of input metabolites to a set of output metabolites without accumulation of any intermediates, it is of interest to identify among them those pathways which are independent.

* Corresponding author. Fax: +1-617-2539695.
E-mail address: gregstep@mit.edu (G. Stephanopoulos)

0168-1656/99/$ - see front matter
PII: S0168-1656(99)00023-1

Enumeration of all independent pathways is an important step in the analysis of the structure of bioreaction networks. First, it identifies the degrees of freedom in a network that uniquely determine all pathway fluxes. For example, in the simple network of Fig. 1a there are three independent pathways whose fluxes are uniquely determined from the fluxes of the individual reactions 3, 4, and 5. The latter can be obtained from the rates of accumulation of metabolites D, E, and F, respectively. Similarly, the network depicted in Fig. 1c has four independent pathways, equal to the degrees of freedom that uniquely determine all pathway fluxes. In contrast to Fig. 1a, these four independent fluxes cannot be obtained from metabolite measurements but must be determined by other methods (Stephanopoulos, 1998). Second, each individual reaction of a network carries the flux corresponding to one or more of the independent pathways of the network. In Fig. 1a, for example, the flux of reaction 2 is the sum of the fluxes of pathways P2 and P3; hence a portion of this reaction is used by each pathway. If the independent pathways are identified, the relative importance of individual reactions in carrying pathway flux can be assessed. Put in a different way, the portion of each reaction's flux which is dedicated to a particular pathway defines the degree of engagement of each reaction by each pathway. Third, if the fluxes of the reactions involved in the independent pathway leading to a desired product are known, it is possible to devise a strategy to amplify the activities of all these reactions in such a way as to amplify the pathway flux to any arbitrary level (Kacser and Acerenza, 1993).

Finally, enumeration of independent pathways is usually accompanied with the identification of critical branch points where metabolic pathways converge or diverge. At such branch points, competition for common metabolites arises between the various pathways (e.g. metabolites B and C in Fig. 1a). These branch points are, in fact, key to the understanding of both the control and the partitioning of fluxes in metabolic networks (Vallino and Stephanopoulos, 1993; Simpson et al., 1995). Once such branch points have been located, groups of reactions can be rationally defined to facilitate the study of flux control using the framework of top-down metabolic control analysis (Brown et al., 1990; Stephanopoulos and Simpson, 1997).

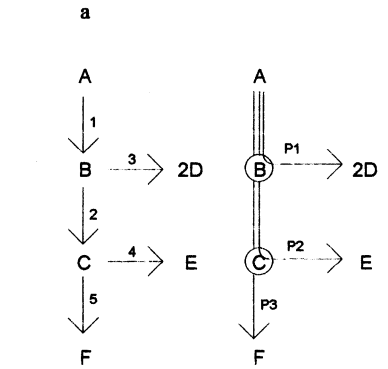
Independent pathways can be defined by inspection in simple bioreaction networks like the one depicted in Fig. 1a. For more complicated networks, however, involving interconnected reactions, feedback, and cycles, the identification of independent pathways is a demanding task requiring a systematic procedure. This can be seen in Fig. 2, which is a simplified schematic of the reactions participating in aromatic amino acid biosynthesis. The purpose of this paper is to present a method for the identification of all independent pathways in a network of any size and complexity. The method makes use of the *scheme matrix* (Reder, 1988), or internal metabolite stoichiometry matrix, and defines the independent pathways through the membership of its kernel matrix. Two case studies from aromatic amino acid biosynthesis and cell culture are presented to illustrate the approach.

2. Identification of independent pathways

An *independent pathway* is defined as the smallest set of reactions connecting a single network output with the necessary network inputs in a way that permits the levels of internal species to reach a steady state (Simpson et al., 1995). The requirement of invariant internal metabolite pools ensures that the system can attain a steady state under a constant input. For a network consisting of r reactions and m_0 independently variable internal metabolites whose pools should be able to reach a steady state, the maximum number of independent pathways P is

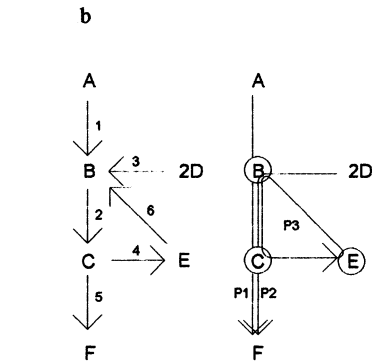
$$P = r - m_0 \tag{1}$$

For the simple examples of Fig. 1a,b, one can easily identify three independent pathways (P1, P2 and P3), as shown in the figures. For more complicated networks, such as the case studies discussed below, the independent pathways are not obvious and a systematic approach is needed for their identification. Such an approach begins by



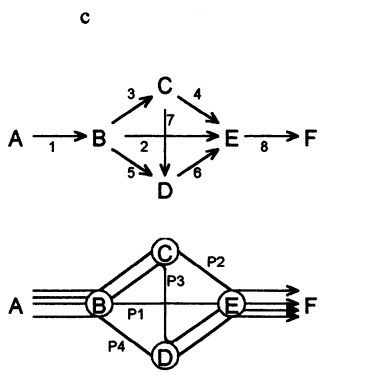
N	Reaction				
	1	2	3	4	5
B	1	-1	-1	0	0
C	0	1	0	-1	-1

K	Pathway		
Rxn.	P1	P2	P3
1	1	1	1
2	0	1	1
3	1	0	0
4	0	1	0
5	0	0	1



N	Reaction					
	1	2	3	4	5	6
B	1	-1	1	0	0	1
C	0	1	0	-1	-1	0
E	0	0	0	1	0	-1

K	Pathway		
Rxn.	P1	P2	P3
1	1	0	0
2	1	1	1
3	0	1	0
4	0	0	1
5	1	1	0
6	0	0	1



N	Reaction							
	1	2	3	4	5	6	7	8
B	1	-1	-1	0	-1	0	0	0
C	0	0	1	-1	0	0	-1	0
D	0	0	0	0	1	-1	1	0
E	0	1	0	1	0	1	0	-1

K	Pathway			
Rxn.	P1	P2	P3	P4
1	1	1	1	1
2	1	0	0	0
3	0	1	1	0
4	0	1	0	0
5	0	0	0	1
6	0	0	1	1
7	0	0	1	0
8	1	1	1	1

Fig. 1. Identification of independent pathways. Three example reaction networks are shown with their resulting independent pathways, and their SIMS and kernel matrices. (a) Network with branches between multiple outputs. (b) Network with branches between multiple inputs and a recycle pathway. (c) Network with many internal reactions.

defining the *steady-state internal metabolite stoichiometry* (SIMS) matrix, which describes the stoichiometry of the network (Reder, 1988). The SIMS matrix **N** (otherwise referred to as the *scheme matrix*), is an $m \times r$ matrix, in which m is the number of explicit steady-state metabolites in

the network, and r is the number of explicit reactions. Each element N_{ji} of this matrix is the stoichiometric coefficient α_{ji} of metabolite X_j participating in reaction i, if each reaction is written as:

$$\sum_{\text{reactants}} (-\alpha_{ji} X_j) + \sum_{\text{products}} (\alpha_{ji} X_j) = 0 \qquad (2)$$

In order to ensure accurate grouping of reactions based upon the method described below, it is crucial that the direction of each reaction described by the above equation be the same as its net flux in the actual network. Thus, since the net flux through reaction 3 in Fig. 1a is opposite that in Fig. 1b, the signs in its columns of the corresponding SIMS matrices are reversed. If the direction of the net flux of a reaction is not immediately known, that reaction can be represented by both forward and reverse reactions in the SIMS matrix. These two opposing reactions would define a cyclic independent pathway, the flux of which cannot generally be determined using simple means of flux analysis. It is therefore generally preferable to redefine the directions of the reactions, if necessary, based upon the direction of the actual fluxes of the network, once the latter have been determined.

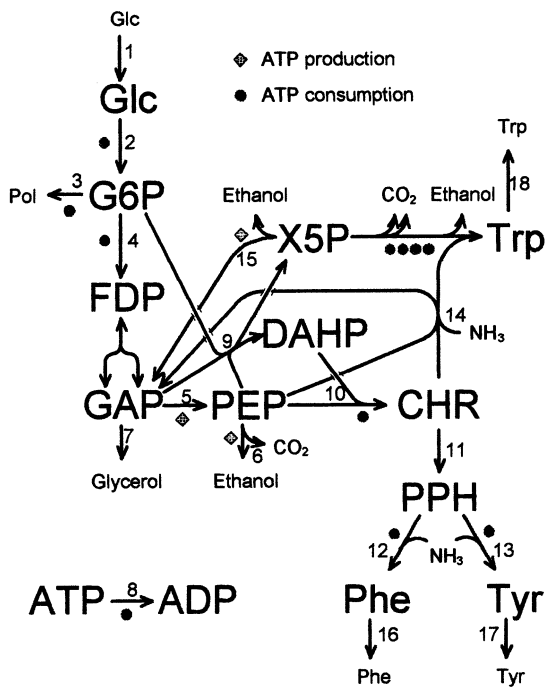


Fig. 2. Amino acid biosynthetic network. Schematic diagram of key reactions participating in the production of aromatic amino acids in *Saccharomyces cerevisiae*, including reactions involved in free-energy production. Steady-state internal metabolites are shown in large type. Reactions (numbered for reference) may comprise one or more coordinated enzymatic steps. The unnumbered reaction converting FDP into GAP is considered to be near equilibrium, with a net rate equal to that of reaction 4. CHR, chorismate; DAHP, 3-deoxy-D-*arabino*-heptulosonate-7-phosphate; FDP, fructose-1,6-diphosphate; G6P, glucose-6-phosphate; GAP, glyceraldehyde-3-phosphate; Glc, glucose; PEP, phosphoenolpyruvate; Phe, phenylalanine; PPH, prephenate; Pol, polysaccharides; Trp, tryptophan; Tyr, tyrosine.

Strictly speaking, the SIMS matrix is a subset of the full stoichiometry matrix, which is also defined by Eq. (2), but which applies to *all* metabolites in the network. Therefore, any external metabolites which serve as feeds or sinks have corresponding rows in the stoichiometry matrix, but not in the SIMS matrix. Since B and C are the only metabolites in Fig. 1a that can be maintained at steady state, m for this system is 2 (as is m_0, since the concentrations of these two metabolites vary independently). The species A, D, E, and F are therefore not included in the resulting SIMS matrix.

Two important types of reactions are easily identified in the SIMS matrix. An output step produces only external species, and is represented in the SIMS matrix by a column that includes no positive values. An input step consumes only external species, and is represented by a column with no negative values. The identification of output steps, in particular, facilitates the enumeration of independent pathways as shown below.

Once the SIMS matrix has been satisfactorily constructed, the independent pathways are obtained as the vectors of the kernel matrix of **N**. By definition, a kernel matrix **K** is any nontrivial solution to Eq. (3):

$$\mathrm{N\,K} = 0 \qquad (3)$$

Fig. 1a–c illustrate these concepts with the indicated SIMS and kernel matrices for the corresponding networks. Note that the entries of the kernel matrix vectors denote the relative rates of the reactions utilized exclusively in the corre-

sponding independent pathway. Thus, if K_{ij} are the elements of the kernel matrix **K**, K_{1j} is a representation of the relative rate of depletion of metabolite A (i.e. the relative rate of reaction 1), as utilized by pathway P$j \in \{1, 2, 3\}$ in Fig. 1a. In this case, the total rate of reaction 1 (equal to the total rate of depletion of A) would be equal to $K_{11} + K_{12} + K_{13}$. The product of the kernel matrix and the SIMS matrix in Eq. (3) simply defines the steady-state balances for the internal metabolites B and C for each of the three independent pathways (a total of 2×3 equations). Thus, the first column of the kernel matrix of Fig. 1a indicates that pathway P1 comprises reactions 1 and 3, and that the indicated stoichiometry will convert A into 2D with no net production of either B or C through the series of reactions defining pathway P1:

$$A \rightarrow B$$

$$\underline{B \rightarrow 2D}$$

$$A \rightarrow 2D$$

If the stoichiometric ratio between steps is not 1:1, entries corresponding to the appropriate ratio will instead appear in the kernel.

The exact composition of the kernel is admittedly arbitrary, since Eq. (3) will still be valid if any column is multiplied by a constant or added to any other column. In fact, the only formal constraints on the kernel are that it be of full rank with independent columns. For the purpose of determining reaction groups and branch points, though, it is essential to define the simplest kernel possible. For a network of m_0 independent steady-state metabolites, there are a total of m_0 steady-state relationships among the r reaction rates of each pathway. Consequently, there are $(r - m_0)$ degrees of freedom in the determination of the reaction rates of each pathway, or, equivalently, the columns of the kernel matrix. By choosing the elements of the kernel matrix so that each column corresponds to a fundamental vector of the base of the kernel, the structure of each independent pathway can be established. The base of the kernel usually consists of the $(r - m_0)$ rows of the kernel corresponding to steps which are each unique to a single independent pathway. We term these steps eigenreactions or characteristic reactions, because they uniquely define the stoichiometric composition and flux of each independent pathway. Such steps are frequently identifiable as output steps leading to different products, but may also encompass steps required for recycles or alternate inputs. Often, in fact, each independent pathway is found to be associated with the synthesis of a different product, in competition with the products of every other independent pathway.

In the example of Fig. 1a, it is apparent that the three degrees of freedom define a base of three reactions. Since the base of the kernel must consist of the three output steps (reactions 3, 4 and 5), each column of the kernel can be defined according to the fundamental vectors for these three steps. This is done by setting values for the three columns of the kernel matrix equal to 100, 010, and 001 for reactions 3, 4 and 5, respectively, and applying Eq. (3) to determine the remainder of the elements. If the proper base is used in a network, this procedure will yield a complete kernel with no negative elements (indicating the inclusion of reactions only in the net direction of the actual fluxes). Owing to the use of the fundamental vectors of the base of the kernel matrix, any steady state observed in a network will be a linear combination of the basic fluxes of the pathways established by the above method (i.e. the fluxes through the characteristic reactions). In addition, once the characteristic reaction for each pathway has been identified, the actual fluxes through each pathway can be found simply by multiplying each value within its column by the flux through the characteristic reaction. This fact simply reiterates the conclusion that there are as many independent fluxes in a network as independent pathways. For the example of Fig. 1a, since the base consists of steps 3–5, the characteristic reactions for pathways P1, P2 and P3 are reactions 3, 4 and 5, respectively. The fluxes through the other two reactions are simply a linear combination of these fluxes. Thus, the flux through each step can be expressed as:

$$\begin{bmatrix} J^1 \\ J^2 \\ J^3 \\ J^4 \\ J^5 \end{bmatrix} = \begin{bmatrix} K_{11} \\ K_{21} \\ 1 \\ 0 \\ 0 \end{bmatrix} J^3 + \begin{bmatrix} K_{12} \\ K_{22} \\ 0 \\ 1 \\ 0 \end{bmatrix} J^4 + \begin{bmatrix} K_{13} \\ K_{23} \\ 0 \\ 0 \\ 1 \end{bmatrix} J^5 \tag{4}$$

2.1. Construction of the kernel

In a simple enough system, the kernel matrix can be constructed by simple inspection of the network schematic, according to the following procedure:

1. Starting with the first column of the kernel, put a 1 in the row of the first column corresponding to one of the output steps, and note the metabolite(s) consumed by that reaction.
2. Determine which step(s) produce(s) the metabolite consumed by the output step identified in step 1.
3. If only one reaction step produces the identified metabolite, place in the entry of the current column corresponding to that reaction the number of times that reaction must proceed in order to produce enough of the metabolite to balance the amount consumed by the previous step; this will usually be a 1. If more than one reaction can produce the metabolite, copy the entries of the entire current column into the next column(s) and continue constructing the current column using the first choice. Once this column is completed, construct the next column(s) using the alternate choices.
4. If the identified reaction is not an input step, repeat Steps 2–3 with that reaction. If it is an input step, the column is finished, and the values in its remaining rows are 0. If the identified reaction is one that was already employed in this column, there will already be a number in the corresponding row. This may indicate the presence of a cycle, in which case the entries downstream from the identified reaction should be replaced with 0 in order to define an independent pathway consisting solely of the steps composing the cycle.
5. Repeat Steps 1–4 beginning with any other output steps.

In complex networks, reactions are likely to have multiple reactants, complicating the choice of metabolites to balance. In such cases, it is often easiest to proceed first with a metabolite which can be produced by only one particular reaction. After the resolution of the production of this metabolite, any *remaining* unbalanced metabolites from the previous step(s) must subsequently be balanced. Additional network features which may confound the determination of a kernel by inspection include networks invoking multiple inputs with multiple outputs, or the presence of coupled pathways. In such cases, it may be simpler to solve the matrix equations than to balance the reactions by hand through this procedure.

The application of these procedures is illustrated with two case studies presented in the next section.

3. Case study 1: aromatic amino acid biosynthesis

The first case study is based on a stoichiometric and kinetic model of *Saccharomyces cerevisiae* biochemistry leading to aromatic amino acid synthesis (Galazzo and Bailey, 1990, 1991). A complete description of the model and several adjustments that were introduced to it can be found in (Stephanopoulos and Simpson, 1997). Fig. 2 shows a schematic of the resulting biosynthetic network, which couples central carbon metabolism and free-energy production with amino acid biosynthesis. This network depicts the production of the three aromatic amino acids, tryptophan, phenylalanine and tyrosine, from simple precursors, with glucose as the sole carbon source. Other metabolic functions included in this model are the production of storage compounds in the form of polysaccharides and glycerol, the production and utilization of free energy, and the requirement of excess free energy production for maintenance. The resulting metabolic network comprises 13 metabolites participating in 18 reactions which form several reaction branches and recycles and exhibit significant feedback and allosteric inhibition. Table 1 provides a list of the reactions explicitly accounted in this network.

Table 1
Stoichiometry of reactions in the amino acid biosynthetic network

Reaction[a]	Reaction stoichiometry[b]
1	$Glc_{ext} \rightarrow \mathbf{Glc}$
2	$\mathbf{Glc} + \mathbf{ATP} \rightarrow \mathbf{G6P} + \mathbf{ADP}$
3	$\mathbf{G6P} + \mathbf{ATP} \rightarrow \mathbf{ADP} + Pol + 2P_i$
4	$\mathbf{G6P} + \mathbf{ATP} \rightarrow \mathbf{FDP} + \mathbf{ADP}$
FDP/GAP[c]	$\mathbf{FDP} \leftrightarrow 2\mathbf{GAP}$
5	$\mathbf{GAP} + \mathbf{ADP} + P_i \rightarrow \mathbf{PEP} + \mathbf{ATP}$
6	$\mathbf{PEP} + \mathbf{ADP} \rightarrow \mathbf{ATP} + EtOH + CO_2$
7	$\mathbf{GAP} \rightarrow Gol + P_i$
8	$\mathbf{ATP} \rightarrow \mathbf{ADP} + Pi$
9	$\mathbf{G6P} + \mathbf{GAP} + \mathbf{PEP} \rightarrow \mathbf{X5P} + \mathbf{DAHP} + P_i$
10	$\mathbf{PEP} + \mathbf{DAHP} + \mathbf{ATP} \rightarrow \mathbf{CHR} + \mathbf{ADP} + 3P_i$
11	$\mathbf{CHR} \rightarrow \mathbf{PPH}$
12	$\mathbf{PPH} + \mathbf{ATP} + NH_3 \rightarrow \mathbf{Phe} + \mathbf{ADP} + P_i$
13	$\mathbf{PPH} + \mathbf{ATP} + NH_3 \rightarrow \mathbf{Tyr} + \mathbf{ADP} + P_i$
14	$\mathbf{PEP} + \mathbf{X5P} + \mathbf{CHR} + 4\mathbf{ATP} + NH_3 \rightarrow \mathbf{GAP} + \mathbf{Trp} + 4\mathbf{ADP} + EtOH + 2CO_2 + 3P_i + PP_i$
15	$\mathbf{X5P} + \mathbf{ADP} + P_i \rightarrow \mathbf{GAP} + \mathbf{ATP} + CO_2$
16	$\mathbf{Phe} \rightarrow Phe_{ext}$
17	$\mathbf{Tyr} \rightarrow Tyr_{ext}$
18	$\mathbf{Trp} \rightarrow Trp_{ext}$
Adenosine[d]	$\mathbf{ATP} + AMP \leftrightarrow 2\mathbf{ADP}$

[a] Reaction numbers correspond to numbering scheme of Fig. 2.

[b] Metabolites in bold face are maintained at steady state. Several of the steps shown are composites of coordinated sets of individual enzymatic reactions.

[c] This reaction represents the equilibrium maintained between FDP and triose sugars, which are considered to be a single metabolite pool.

[d] This reaction represents the equilibrium maintained between ATP, ADP, and AMP.

Note that several of the reactions listed in Table 1 are the stoichiometric equivalent of sets of individual reactions lumped together to yield the overall reactions shown in the table.

Thus, following transport into the cell and phosphorylation (reactions 1 and 2), glucose-6-phosphate (G6P) can be converted to polysaccharides (Pol) by a series of reactions (lumped as reaction 3), or further phosphorylated in the glycolytic pathway to yield fructose-1,6-diphosphate (FDP) through reaction 4. FDP (or its equilibrium equivalent, glyceraldehyde phosphate, GAP), is converted into phosphoenolpyruvate (PEP) and further to ethanol (via pyruvate) with the simultaneous release of free energy in the form of ATP (reactions 5 and 6). Alternatively, GAP can be converted to glycerol (Gol) through a different pathway (reaction 7). The initiation of the shikimate pathway that ultimately leads to aromatic amino acids occurs with the synthesis of the pentose phosphate pathway intermediate xylulose-5-phosphate (X5P) and erythrose-4-phosphate (not shown), which then condenses with PEP to form 3-deoxy-D-*arabino*-heptulosonate-7-phospate (DAHP). These two steps are combined in reaction 9. Chorismate (CHR) is formed from DAHP by a series of reactions depicted as reaction 10. Chorismate, in turn, can be converted to prephenate (PPH), and finally to the products phenylalanine (Phe) and tyrosine (Tyr) through a series of reaction steps lumped in reactions 11, 12, and 13, respectively. An alternative fate of chorismate lies in the production of tryptophan (Trp) by the tryptophan biosynthetic pathway represented by reaction 14; this pathway requires the consumption of X5P in addition to chorismate, as well as the recycle of PEP into GAP. The remaining X5P is recycled to GAP by reaction 15. Amino acid export, degradation, and utilization steps are depicted in reactions 16–18.

In addition to balancing the formation and depletion of the metabolites involved in the synthesis of the carbon skeletons, a steady state in this network must also balance the production and consumption of ATP, as identified in Table 1. These ATP-producing and -consuming reactions are also marked by circles and diamonds, respectively, in Fig. 2 to underline the strong coupling that exists between the various reactions and parts of the network beyond the obvious ones resulting from direct stoichiometric balances. Reaction 8 has been added for the conversion of ATP generated in excess of what is explicitly required by the other reaction steps. This reaction can be thought of as the total of all ATP consumed by other biosynthetic reactions, for cell maintenance, or in futile cycles.

The SIMS matrix of this network, as constructed from the stoichiometry of Table 1, is given in Table 2. Note that only 12 of the 13 metabolites in this matrix are truly independent, since the rows corresponding to ATP and ADP

Table 2
SIMS matrix of amino acid biosynthetic network

Species	Reaction[a]																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Glc	1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G6P	0	1	−1	−1	0	0	0	0	−1	0	0	0	0	0	0	0	0	0
FDP	0	0	0	1	−0.5	0	−0.5	0	−0.5	0	0	0	0	0.5	0.5	0	0	0
PEP	0	0	0	0	1	−1	0	0	−1	−1	0	0	0	−1	0	0	0	0
ATP	0	−1	−1	−1	1	1	0	−1	0	−1	0	−1	−1	−4	1	0	0	0
DAHP	0	0	0	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	0
CHR	0	0	0	0	0	0	0	0	0	1	−1	0	0	−1	0	0	0	0
PPH	0	0	0	0	0	0	0	0	0	0	1	−1	−1	0	0	0	0	0
X5P	0	0	0	0	0	0	0	0	1	0	0	0	0	−1	−1	0	0	0
Phe	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	−1	0	0
Tyr	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	−1	0
Trp	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	−1
ADP	0	1	1	1	−1	−1	0	1	0	1	0	1	1	4	−1	0	0	0

[a] Reaction numbers correspond to numbering scheme of Fig. 2.

Table 3
Kernel of the SIMS matrix of amino acid biosynthetic network

Reaction[b]	Independent pathway[a]					
	P1 Trp	P2 Phe	P3 Tyr	P4 Pol	P5 Gol	P6 ADP
1	5.5	3	3	2	1	0.5
2	5.5	3	3	2	1	0.5
3	0	0	0		0	0
4	4.5	2	2	1	1	0.5
5	9	4	4	2	1	1
6	6	2	2	2	1	1
7	0	0	0	0		0
8	0	0	0	0	0	
9	1	1	1	0	0	0
10	1	1	1	0	0	0
11	0	1	1	0	0	0
12	0	1	0	0	0	0
13	0	0	1	0	0	0
14	1	0	0	0	0	0
15	0	1	1	0	0	0
16	0		0	0	0	0
17	0	0		0	0	0
18		0	0	0	0	0

[a] These independent pathways are shown schematically in Fig. 3. A characteristic reaction for each pathway is indicated by an entry within a black circle.
[b] Reaction numbers correspond to numbering scheme of Fig. 2.

mirror each other. Eq. (1), therefore, shows that a total of six independent pathways can be constructed from these 18 reactions, each corresponding to one of the six output reactions. These reactions (3, 7, 8, 16, 17 and 18) are identified as the characteristic reactions of the network, and thus form the base of the kernel. (Strictly speaking, reactions 3 and 8 are recycle steps of ATP to ADP.) At first glance, it would seem a straightforward exercise to derive the kernel matrix (Table 3) by working backwards and balancing each metabolite according to the method described in the previous section. In actuality, however, the interdependence of the energy-associated reactions introduces some additional complications.

Consider, for example, the pathway comprising reactions 1, 2 and 3, which results in the production of polysaccharides (Pol). From Fig. 2, it can be seen that reactions 1 and 2 should balance the glucose-6-phosphate (G6P) consumed by reaction 3. Hence, from the standpoint of the carbon conversion schematic of Fig. 2 alone, these three steps should complete the column for the polysaccharide pathway (pathway 4) in the kernel. However, inspection of the fourth column of the kernel shown on Table 3 reveals additional entries for this pathway. These entries correspond to the reactions that must also be included along with reactions 1, 2 and 3 to allow for the closure of the ATP (and by extension, ADP) balance. As it happens, the only way this network can provide the ATP needed by these reactions is through the glycolytic pathway ending with reaction 6. Thus, the completion of the polysaccharide pathway requires the addition of reactions 1, 2 and 4, as well as two equivalents of reactions 5 and 6. If the same procedure is followed for each of the other output steps, the result will be the kernel matrix shown in Table 3. This kernel matrix can also be found through the solution of Eq. (3), using the six output steps identified above as the base. Fig. 3 provides a schematic representation of the balanced fluxes for each of the pathways identified in the kernel.

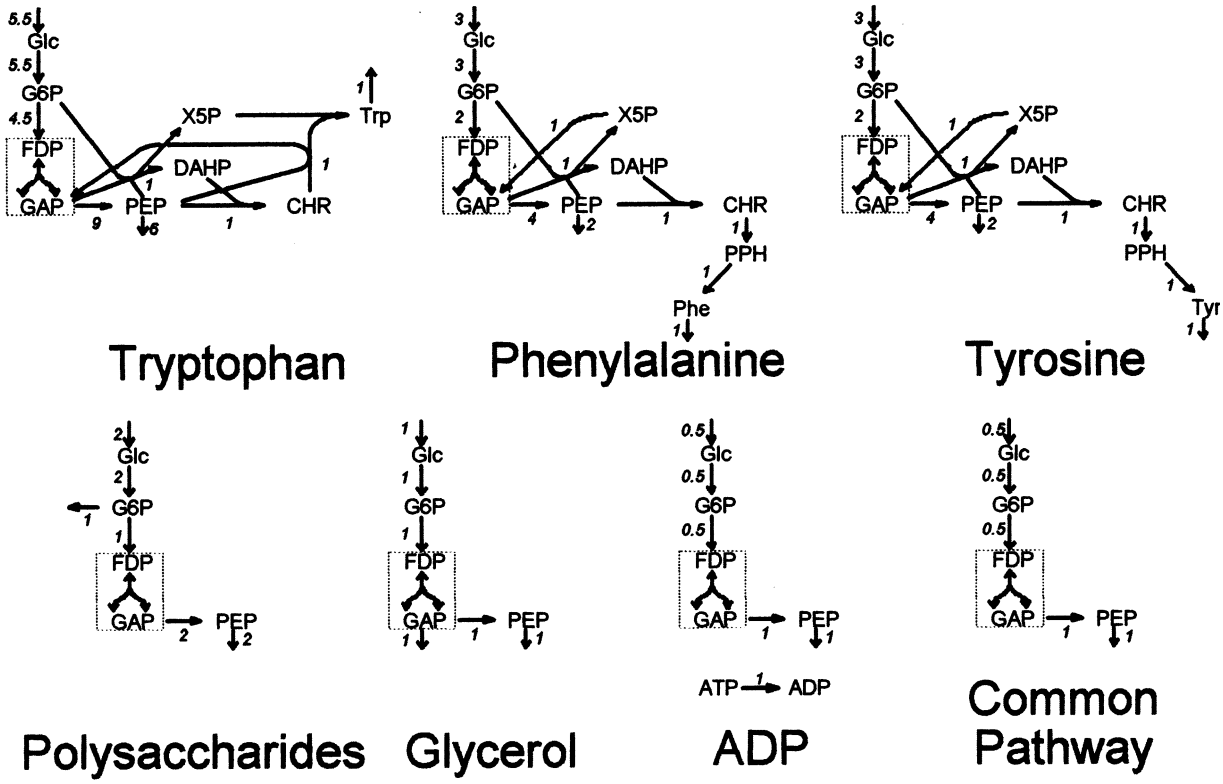


Fig. 3. Independent pathways of the aromatic amino acid biosynthetic network. There are six independent pathways associated with the six products, as shown. The entries of the kernel matrix are shown schematically as balanced fluxes through each pathway. The relative flux through the energy-producing glycolytic pathway, which is common to all six independent pathways, is also identified.

As shown in Fig. 3, the glycolytic ATP-producing pathway is used by all six independent pathways. Thus, the first branch point in this network actually occurs *downstream* of this pathway, with ATP identified as the first link metabolite. This is a critical fact which is not immediately apparent from the reaction schematic of Fig. 2, since the latter is drawn in order to depict only the carbon metabolite conversions. Nevertheless, as shown in Fig. 4a, competition for ATP exists between the biosynthetic pathways (P1–P5) and the hydrolysis of ATP to ADP for maintenance (or futile cycle ATP usage) in pathway P6. It must be emphasized that the branch points and pathways found solely from the carbon-structure map, without the inclusion of ATP as link metabolite, can in fact lead to grossly erroneous conclusions in the calculation of flux split ratios and flux control coefficients.

The branching structure of this network and its link metabolites are summarized in Fig. 4. From this diagram, showing both the carbon- and energy-associated fluxes, it should be clear that the metabolic network comprises a series of divergent branch points and that each independent pathway is associated with a secreted product formed by an output step. Experimentally, measuring the accumulation of these secreted metabolites, as well as other by-products (in this case, CO_2 and ethanol), provides the means of estimating the flux through each independent pathway using flux analysis methods (Zupke and Stephanopoulos, 1995). Furthermore, changes in their accumulation during different phases of a fermentation process or in response to induced perturbations can be used for the quantitative evaluation of the kinetic control exercised by each reaction group (Stephanopoulos and Simpson, 1997; Simpson et al., 1998).

It is noted that, without the explicit identification of the independent pathways in this network, it would be difficult to decouple the energy-production pathway from biosynthesis. Furthermore, the carbon conversion map alone could lead to the mistaken conclusion that reaction 6 competes for PEP with the amino acid-producing pathways, when this step is actually necessary for the production of energy needed for these pathways. Although chorismate and xylulose-5-phosphate may be correctly identified as link metabolites, without a steady-state pathway analysis it may not be clear that they constitute a single branch point, since they are in fact always produced and consumed together by pathways P1–P3.

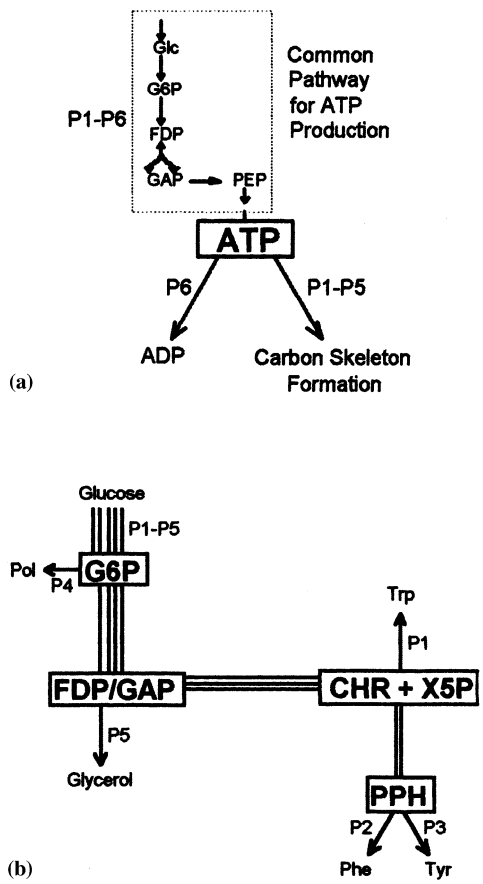


Fig. 4. Branch points of the aromatic amino acid biosynthetic network. The participation of each independent pathway in the system's five branch points is shown schematically. (a) The energy-related branch point meets the excess ATP requirements of products formed through pathways P1–P5, as well as the maintenance pathway P6. (b) Four downstream branch points separate the five product pathways P1–P5. Note that identification of the two distinct glucose fluxes that feed (a) and (b) is necessary for the analysis of reaction groups.

4. Case study 2: central carbon metabolism in cell culture

The second case study provides an interesting challenge, because it involves multiple network inputs and outputs, as well as coupled energy and redox steps. This network represents central carbon metabolism in mammalian cells cultured on simultaneous glucose and glutamine feed (Zupke and Stephanopoulos, 1995; Bonarius et al., 1996). The reactions making up this network are shown schematically in Fig. 5. The biochemical pathways represented by these reactions include glycolysis, the pentose–phosphate cycle, and the TCA cycle, as well as the production of the amino acid alanine. Apart from the reactions shown, which depict the carbon flow through the network, five additional reactions are responsible for the coupled interconversion of the pairs of energy species ATP/ADP and GTP/GDP, and the redox pairs $NADH/NAD^+$, $NADPH/NADP^+$, and FAD^+. Because each of these pairs of species are interdependent, the 29 reactions and 24 independent metabolites yield five independent pathways, as shown in Fig. 6. These pathways were derived from the kernel of the SIMS matrix of this network through the procedure described above (Tables 4 and 5).

It should be noted first that this network displays a more complex branching structure, as shown systematically in Fig. 7. In particular, pyruvate is the only carbon metabolites common to all five independent pathways. Consequently, except for the coupled energy and redox reactions, there are no reactions common to all five pathways. Thus, the key branch points are found by identifying pathways sharing a long series of reactions, and contrasting them with other pathways which do not. In so doing, it becomes clear that various pathways diverge and reconverge in several sections of the network. Pathways 1 and 2,

Table 4
SIMS matrix of central carbon metabolism network

Species	Reaction[a]																												
	Glycolysis					Pentose phosphate cycle						TCA Cycle and Related Steps									Energy and Redox					Input/Output			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
AcCoA	0	0	0	0	0	0	0	0	0	0	0	0	1	0	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ADP	1	0	1	−2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	−3	0	−2	−1	0	0	0	0
AKG	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	−1	0	1	0	0	0	0	0	0	0	0	0	0
ALA	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	−1	0	0
ATP	−1	0	−1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−1	3	0	2	1	0	0	0	0
CIT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0
E4P	0	0	0	0	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F6P	0	1	−1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FAD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−1	0	0	0	0	0	0	1	0	0	0	0	0
FADH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	−1	0	0	0	0	0
G6P	1	−1	0	0	0	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GAP	0	0	2	−1	0	0	0	1	−1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GDP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−1	0	0	0	0	0	0	0	1	0	0	0	0
GLC	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
GLN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−1	0	0	0	0	0	0	0	0	1
GLU	0	0	0	0	0	0	0	0	0	0	0	0	0	−1	0	0	0	0	−1	1	0	0	0	0	0	0	0	0	0
GTP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	−1	0	0	0	0
LAC	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−1	0
MAL	0	0	0	0	0	0	0	0	0	0	0	−1	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	0	0
NAD	0	0	0	−1	1	0	0	0	0	0	0	0	−1	0	0	−1	−1	−1	−1	0	0	1	0	0	0	0	0	0	0
NADH	0	0	0	1	−1	0	0	0	0	0	0	0	1	0	0	1	1	1	1	0	0	−1	0	0	0	0	0	0	0
NADP	0	0	0	0	0	−2	0	0	0	0	0	−1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
NADPH	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	−1	0	0	0	0	0	0
OAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−1	0	0	1	−1	0	0	0	0	0	0	0	0	0	0
PYR	0	0	0	1	−1	0	0	0	0	0	0	1	−1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R5P	0	0	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RU5P	0	0	0	0	0	1	−1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S7P	0	0	0	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X5P	0	0	0	0	0	0	1	0	−1	0	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

[a] Reaction numbers correspond to numbering scheme of Fig. 5.

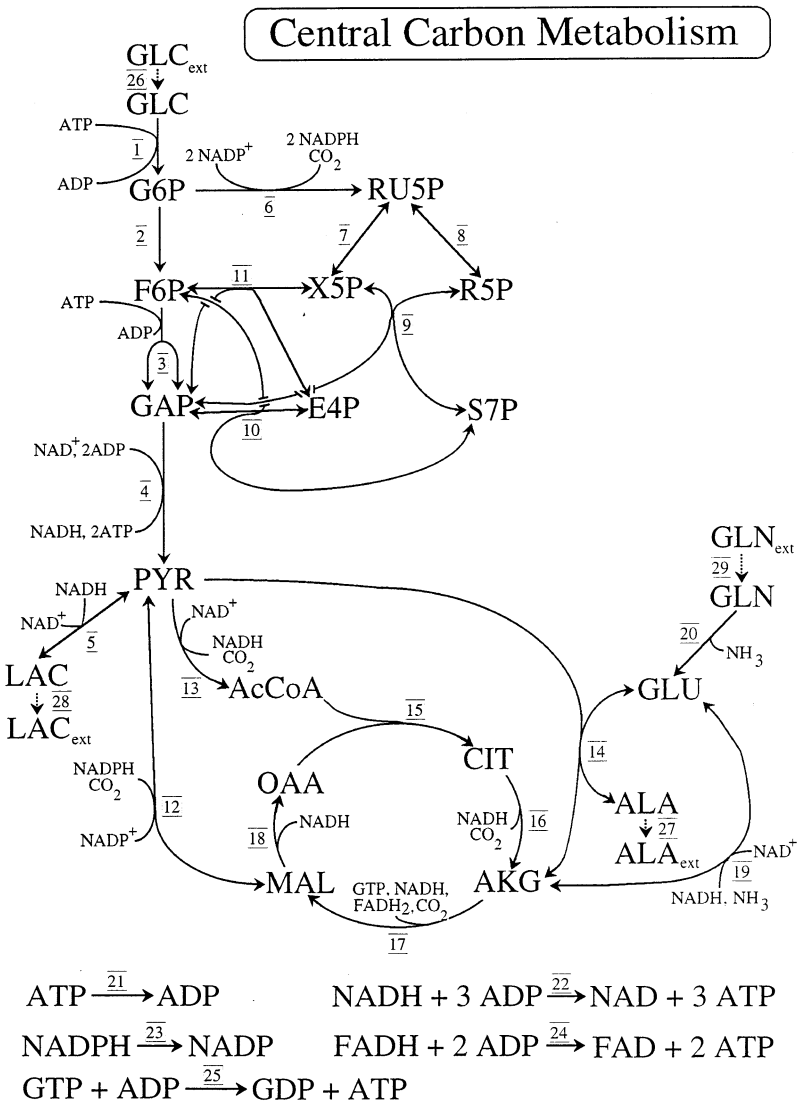


Fig. 5. Central carbon metabolism network. Schematic diagram of key reactions participating in the central carbon metabolism and related pathways in hybridoma cells. AKG, α-ketoglutarate; ALA, alanine; CIT, citrate; E4P, erythrose-4-phosphate; FDP, fructose-1,6-diphosphate; F6P, fructose-6-phosphate; G6P, glucose-6-phosphate; GAP, glyceraldehyde-3-phosphate; GLC, glucose; GLN, glutamine; GLU, glutamate; LAC, lactate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PYR, pyruvate; R5P, ribose-5-phosphate; RU5P, ribulose-5-phosphate; S7P, sedulose-7-phosphate; Trp, tryptophan; Tyr, tyrosine.

for example, diverge at G6P and reconverge with each other at GAP. Thus, G6P can be considered to be the primary branch point separating these two pathways, although the analysis of the branch point GAP may also be useful. Similarly, pathways 3–5 each convert the glutamine substrate into glutamate, but they convert glutamate into α-ketoglutarate through different mechanisms. Specifically, pathway 4 diverges from pathways 3 and 5 at the glutamate branch point and reconverges at α-ketoglutarate. Hence, the (equivalent) flux distribution at either of these points should be of significant interest. (Although pathway 4 also appears to diverge at the pyruvate branch

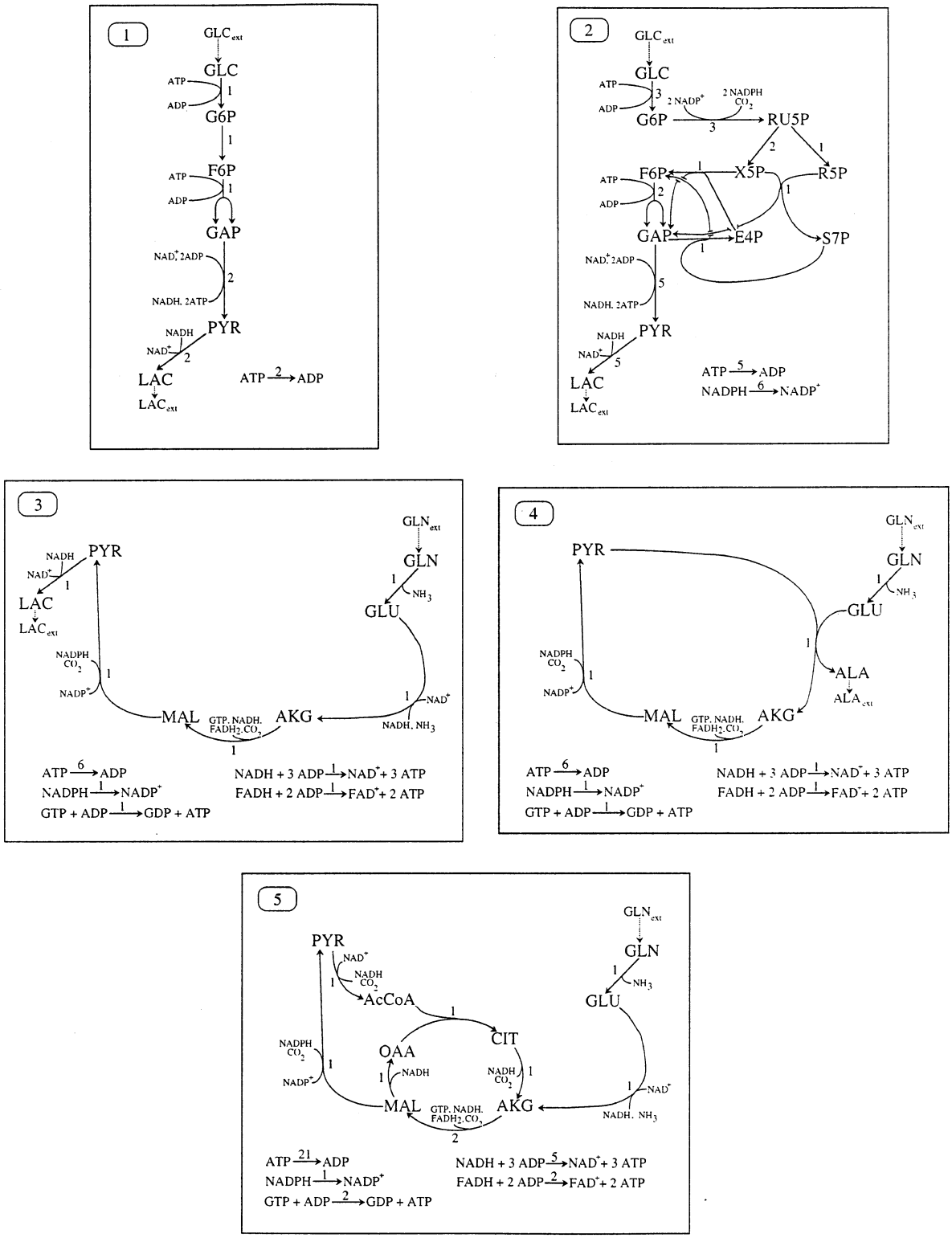


Fig. 6. Independent pathways of the central carbon metabolism network. There are five independent pathways associated with five independent sets of inputs and outputs, as shown. The entries of the kernel matrix are shown schematically as balanced fluxes through each pathway.

point, it is more simply and properly analyzed at α-ketoglutarate.)

The intricate branch point at pyruvate is of greater interest, because the pathways from the two different feed substrates converge at this point. In addition, pyruvate can either be converted into lactate (pathways 1–3) or enter the TCA cycle (pathway 5). This is actually a multiple branch point with three inputs (from glycolysis, the pentose–phosphate cycle, or gluconate) and two outputs (into lactate or the TCA cycle). There are, therefore, six unique steady-state paths through pyruvate (not counting the alternate use of pyruvate by pathway 4), only four of which are independent. (Notice that the standard route of oxidative phosphorylation of pyruvate from glycolysis is one of the two arbitrary 'missing' pathways.) Because of this branch point's unique nature, its analysis is likely be critical. In addition, it is actually *two* branch points (convergent and divergent), whose flux distributions will vary independently. Thus, there are, as expected, four branch points (one each at G6P and glutamate, and two at pyruvate) identified among these five pathways. Other metabolites that might be perceived as branch points in this network, such as malate, can be excluded from consideration on the basis of this analysis, thus facilitating subsequent experimental and analytical efforts.

Table 5
Kernel of the SIMS matrix of central carbon metabolism network

Reaction[b]	Independent pathway[a]				
	P1	P2	P3	P4	P5
1	1	3	0	0	0
2		0	0	0	0
3	1	2	0	0	0
4	2	5	0	0	0
5	2	5	1	0	0
6	0	3	0	0	0
7	0	2	0	0	0
8	0		0	0	0
9	0	1	0	0	0
10	0	1	0	0	0
11	0	1	0	0	0
12	0	0	1	1	1
13	0	0	0	0	
14	0	0	0	1	0
15	0	0	0	0	1
16	0	0	0	0	1
17	0	0	1	1	2
18	0	0	0	0	1
19	0	0		0	1
20	0	0	1	1	1
21	2	5	6	6	21
22	0	0	1	1	5
23	0	6	1	1	1
24	0	0	1	1	2
25	0	0	1	1	2
26	1	3	0	0	0
27	0	0	0		0
28	2	5	1	0	0
29	0	0	1	1	1

[a] These independent pathways are shown schematically in Fig. 6. A characteristic reaction for each pathway is indicated by an entry within a black circle. Note that the flux through pathway P3 must be determined from the flux through its characteristic reaction *minus* the flux through pathway P5.

[b] Reaction numbers correspond to numbering scheme of Fig. 5.

5. Discussion

It is important to realize that the independent steady-state pathways in a metabolic network are firmly interconnected through their constituent reactions. This, of course, implies that a network harbors several branch points at which these pathways diverge in order to form the final products. Each such branch point in fact defines a point of *competition* between two (or more) independent pathways for the metabolite at the branch point, i.e. the *link metabolite*. The identification and characterization of each branch point is, therefore, very useful. Once the independent pathways defined by the kernel matrix have been identified, it is a rather straightforward procedure to compare these pathways in order to locate the points at which they diverge. The ultimate purpose of this procedure is to locate the link metabolites in a metabolic network and to systematically define the reaction groups and their corresponding fluxes around each. In fact, nearly any branch point, no matter the complexity, can be represented as a simple branch point with three reaction groups. This is shown schematically in Fig. 8, which depict a branch point at link

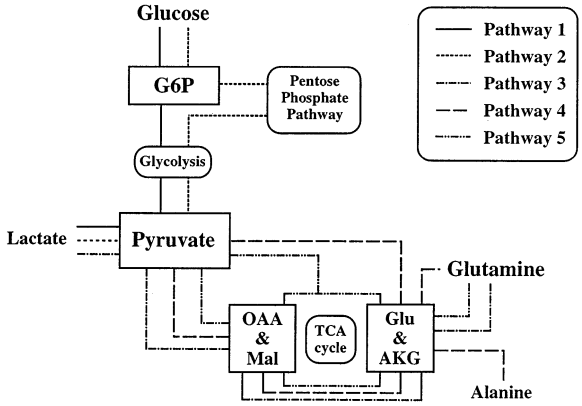


Fig. 7. Branch points of the central carbon metabolism network. The participation of each independent pathway in the system's branch points is shown schematically. Note that the pyruvate branch point is both a convergent branch point (from glycolysis or from malate) and a divergent branch point (to lactate or the TCA cycle). Although a portion of the TCA cycle is utilized in pathways 3 and 4, only pathway 5 fully invokes the TCA cycle for energy production through oxidative phosphorylation; hence, malate need not be treated as a branch point.

metabolite X_j along with the three reaction groups A, B, and C formed around it. Each such branch point is characterized by the fluxes J^A, J^B, and J^C around the branch point metabolite. The correct accounting of these fluxes (which can be found from the kernel matrix) allows each branch point to be studied qualitatively through its flux distribution and split ratios, and in particular can facilitate the implementation of top-down metabolic control analysis.

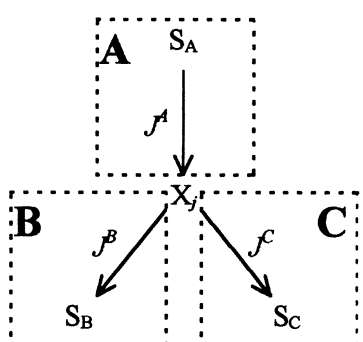


Fig. 8. General representation of a divergent branch point along with the reaction groups and metabolic fluxes defined around the branch point. Reaction groups A, B and C include all network reactions upstream and downstream of the branch point metabolite.

On the simplest level, it is extremely useful to see at a glance exactly which reactions are involved in the pathway leading to a useful product or an unwanted by-product. This allows one to begin considering options for amplifying the pathway of interest or attenuating a competing pathway. More importantly, identification of the independent steady-state pathways in a metabolic network provides an excellent basis for describing the interactions and competition between different sections of the network. An ensuing analysis of the branch points which delineate different pathways can identify the most critical link metabolites in the network. In addition, experimental study of the groups of reactions surrounding each branch point (Fig. 8) can identify groups, and subsequently individual reactions, which are critical to the formation of products of interest (Stephanopoulos and Simpson, 1997).

Finally, it should be pointed out that the conclusions that can be reached from pathway and branch point analysis are based entirely upon the stoichiometric structure of the network. At this level, it is possible to ignore the complexities introduced by reaction kinetics and, in particular, inhibition and feedback. Traditionally, the study of metabolism has been rather narrowly focused upon mechanisms of feedback and kinetics, which are undoubtedly of great importance to the study of isolated enzymes. When considering large sets of reactions, however, it becomes nearly impossible to reach salient conclusions based upon kinetics and feedback, without first considering the intrinsic structure of the reaction network. Hence, the identification of the pathway structure is fundamental to the systematic and effective study of metabolic networks.

Acknowledgements

Financial support from the National Science Foundation, Grant No. BES-9634569 and the De-

partment of Energy, Grant No. DE-FG02-94-ER14487, is gratefully acknowledged.

References

Bonarius, H.P.J., Hatzimanikatis, V., Meesters, K.P.H., de Gooijer, C.D., Schmid, G, Tramper, G., 1996. Metabolic flux analysis of hybridoma cells in different culture media using mass balances. Biotechnol. Bioeng. 50, 299–318.

Brown, C., Hafner, R.P., Brand, M.D., 1990. A 'top-down' approach to determination of control coefficients in metabolic control theory. Eur. J. Biochem. 188, 321–325.

Galazzo, J.L., Bailey, J.E., 1990. Fermentation pathway kinetics and metabolic flux control in suspended and immobilized *Saccharomyces cerevisiae*. Enzyme Microb. Technol. 12, 162–172.

Galazzo, J.L., Bailey, J.E., 1991. Errata. Enzyme Microb. Technol. 13, 363.

Kacser, H., Acerenza, L., 1993. A universal method for achieving increases in metabolite production. Eur. J. Biochem. 216, 361–367.

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

Simpson, T.W., Colón, G.E., Stephanopoulos, G., 1995. Two paradigms of metabolic engineering applied to amino acid biosynthesis. Trans. Biochem. Soc. 23, 381–387.

Simpson, T.W., Shimizu, H., Stephanopoulos, G., 1998. Experimental determination of group flux control coefficients in metabolic networks. Biotechnol. Bioeng. 58, 149–153.

Stephanopoulos, G., 1998. Metabolic fluxes and metabolic engineering. Metabol. Eng. (in press).

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

Vallino, J.J., Stephanopoulos, G., 1993. Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. Biotechnol. Bioeng. 41, 633–646.

Zupke, C., Stephanopoulos, G., 1995. Intracellular flux analysis in hybridomas using mass balances and in vitro ^{13}C NMR. Biotechnol. Bioeng. 45, 292–303.