ELSEVIER

Biophysical Chemistry 99 (2002) 63–75

Biophysical Chemistry

www.elsevier.com/locate/bpc

Treatment of multifunctional enzymes in metabolic pathway analysis

Stefan Schuster*, Ionela Zevedei-Oancea

Max Delbrück Center for Molecular Medicine, Department of Bioinformatics, Robert-Rössle-Str 10, 13092 Berlin, Germany

Received 22 March 2002; received in revised form 25 April 2002; accepted 26 April 2002

Abstract

In metabolic pathway analysis, it should be considered that many enzymes operate with low specificity (e.g. nucleoside diphosphokinase, uridine kinase, transketolase, aldolase), so that various substrates and products can be converted. Here, we analyze the effect of enzymes with low substrate specificity on the elementary flux modes (pathways). We also study the benefits of two different approaches to describing multifunctional enzymes. The usual description is in terms of (overall) enzymatic reactions. At a more detailed level, the reaction steps (half-reactions, hemi-reactions) of the formation and conversion of enzyme-substrate complexes are considered. Multifunctional enzymes operate according to various mechanisms. This is illustrated here by the reaction schemes for the different enzyme mechanisms of bifunctional enzymes. For enzymes with two or more functions, it is important to consider only linearly independent functions, because otherwise cyclic elementary modes would occur which do not perform any net transformation. However, the choice of linearly independent functions is not a priori unique. We give a method for making this choice unique by considering the extreme pathways of the hemi-reactions system. A formal application of the algorithm for computing elementary flux modes (pathways) yields the result that the number of such modes sometimes depend on the level of description if some reactions are reversible and the products of the multifunctional enzymes are external metabolites or some multifunctional enzymes partly share the same metabolites. However, this problem can be solved by appropriate interpretation of the definition of elementary modes and the correct choice of independent functions of multifunctional enzymes. The analysis is illustrated by a biochemical example taken from nucleotide metabolism, comparing the two ways of description for nucleoside diphosphokinase and adenylate kinase, and by several smaller examples.

Keywords: Multifunctional enzyme; Low-specificity enzyme; Elementary flux modes; Nucleoside diphosphokinase; Half-reaction

1. Introduction

Pathway analysis has become an important tool in biochemical modeling [1–6]. It is helpful in functional genomics [7–9] and biotechnology [2,6,10,11]. A central concept in this analysis is that of elementary flux modes [2,11,12]. An elementary mode is a minimal set of enzymes that could operate at steady state. 'Minimal' means that if only the enzymes belonging to this set were operating at steady state, complete inhibition of

*Corresponding author. Tel.: +49-30-9406-3125; fax: +49-30-9406-2834.

E-mail address: stschust@mdc-berlin.de (S. Schuster).

0301-4622/02/$ - see front matter
PII: S0301-4622(02)00122-9

one of them would stop any steady-state flux in the system. If we write a flux distribution as a vector, **V**, with the components corresponding to the fluxes through the particular enzymes, then this definition implies that **V** represents an elementary mode if there is no other steady-state flux distribution, **V′**, having zero components wherever **V** does and at least one additional zero component [13].

Another definition says that an elementary mode cannot be decomposed into two flux modes involving fewer enzymes [12,14],

$$\mathbf{V} \neq \lambda_1 \mathbf{V}' + \lambda_2 \mathbf{V}'', \lambda_{1,} \lambda_2 > 0$$

It has recently been proved that these two definitions are equivalent [13]. Any flow pattern can be expressed as a non-negative linear combination of the elementary modes. In this way, we can find, for a given biochemical system, the metabolic routes that lead from a particular starting metabolite to the desired reaction product, and it becomes clear whether a particular enzyme is essential in the biotransformation. Elementary modes have been determined, for example, for sucrose metabolism in sugar cane [5] and for carbon metabolism in *Methylobacterium extorquens* [6]. Related concepts in pathway analysis were proposed [1,10,15,16]. Nuño et al. [15] and Schilling et al. [1] argued to use the set of extreme pathways (sometimes called 'convex basis'). The convex basis is a subset of the set of elementary modes with the property that no extreme pathway can be expressed as a superposition of other extreme pathways. Elementary modes and the convex basis can be computed by the C program METATOOL [17], which is available from http://www.bioinf.mdc-berlin.de/projects/metabolic/metatool/.

Many enzymes are known to be multifunctional. This means that different substrates might be bound in one or more than one active sites. We focus on those enzymes having only one active site, but with a low specificity for the substrates. An example is provided by uridine kinase. It can phosphorylate uridine, cytidine and a number of derivatives of these nucleosides (such as 5-fluorouridine) [18,19]. As phosphate donors, various nucleoside triphosphates can be used. Competitive inhibition by alternative substrates binding to the same active site need not be taken into account in computing the elementary modes because these modes refer to potential flux distributions. The magnitudes of reaction rates and the percentage of each elementary mode with which it contributes to the actual flux distribution is not the point of our interest in pathway analysis.

The description of multifunctional enzymes can be made in two different ways. The usual way is by writing overall reactions in terms of substrates and products. In the context of pathway analysis, an alternative was proposed by Nuño et al. [15]. This approach describes the enzymatic mechanism in more detail considering the steps of the formation, interconversion and decay of the enzyme–substrate complex (half-reactions, hemi-reactions). Such a detailed description has also been used in other areas of biochemical modeling [20].

Apart from the study by Nuño et al. [15], up to now, multifunctional enzymes have been treated in pathway analysis at the level of overall reactions. Of course, the different functions of such an enzyme then have to be treated as distinct reactions when computing the elementary modes (for example, by the program METATOOL). For example, the two functions of transketolase have been labeled as TktI and TktII in [2]. For the computation of pathways, this description implies that the two reactions are treated independently. However, if knockout mutations or enzyme deficiencies are studied, care has to be taken that the two reactions are deleted simultaneously from the list of reactions.

The number of reactions to be considered depends on the number of different functions of an enzyme and, moreover, on the level of description (half-reactions vs. overall reactions). The question arises whether the pathways computed by elementary mode analysis are invariant to the level of description. This will be studied in the present article.

2. The case where all multifunctional enzymes involve only irreversible steps

For uridine kinase, the reaction with the highest activity is

Urk1: ATP + uridine = ADP + UMP.

In the following, we list all reactions phosphorylating uridine that have more than approximately 50% of the activity of this reaction as well as all reactions phosphorylating cytidine that have more than approximately 50% of the activity of the reaction

Urk2: ATP + cytidine = ADP + CMP,

(which, in turn, has 35.6% of the activity of **Urk1**) (see database BRENDA, http://www.brenda.uni-koeln.de/).

Urk3: dATP + uridine = dADP + UMP;
Urk4: dATP + cytidine = dADP + CMP;
Urk5: dUTP + uridine = dUDP + UMP;
Urk6: dUTP + cytidine = dUDP + CMP;
Urk7: dCTP + uridine = dCDP + UMP;
Urk8: dCTP + cytidine = dCDP + CMP.

In every cell containing a given multifunctional enzyme, all its different functions are performed simultaneously, but to avoid redundancy, we will consider only independent functions. The other ones can be expressed as linear combinations of these. For example, **Urk4** = −**Urk1** + **Urk2** + **Urk3**. The number of linearly independent reactions is given by the rank of the stoichiometry matrix [14,21]. This can be calculated either by standard tools from linear algebra or by determining the number of metabolites minus the number of independent conservation relations. We here apply the second method. There are twelve metabolites involved in uridine kinase: ATP, dATP, ADP, dADP, UMP, CMP, dUTP, dUDP, dCTP, dCDP, uridine, cytidine, and seven conservation relations, notably for the moieties ADP, dADP, dUDP, dCDP, uridine, cytidine, and phosphate. Thus, the number of independent reactions is five (12 − 7 = 5). As in any vector space, the choice of basis vectors is not unique. Here, one could choose, for example, the following five reactions: **Urk1**, **Urk2**, **Urk3**, **Urk5**, **Urk7**. Two of these are shown in Fig. 1.

As mentioned in the Introduction, multifunctional enzymes can be described in two different ways. The formulation of a reaction in terms of elementary steps (half-reactions) depends on the reaction mechanism. For example, a bifunctional reaction, A + B → C + D and F + B → G + D, can proceed according to the different mechanisms shown in Fig. 2. There are opposing views in the literature about the mechanism of uridine kinase [18,19,22,23] (see also database BRENDA). Let us first assume that it operates according to an ordered sequential mechanism [19,22]. Clearly the phosphate donor binds first. For symmetry reasons, this implies that (d)NDP is released last (with N standing for A, U or C). Then, the half-reactions read:

ATP + Urk = Urk-ATP (i.e. the enzyme-substrate complex);
Urk-ATP + uridine = Urk-ADP + UMP;
Urk-ATP + cytidine = Urk-ADP + CMP;
Urk-ADP = Urk + ADP;
dNTP + Urk = Urk-dNTP;
Urk-dNTP + uridine = Urk-dNDP + UMP;
Urk-dNTP + cytidine = Urk-dNDP + CMP;
Urk-dNDP = Urk + dNDP.

More rigorously, one should write the formation and decay of ternary complexes such as Urk-ATP-uridine. However, for establishing the input to METATOOL, it is not necessary to write all elementary steps. Consecutive steps without branching in between can be lumped into one step because any elementary mode containing one of these steps will also include the other [17]. For example, in the ordered ping-pong mechanism indicated in Fig. 2a1, we can lump the first two steps (E + B → EB, EB → D + EP) into E + B → D + EP and the last two steps A + EP → EC, EC → C + E into A + EP → C + E, respectively, F + EP → EG, EG → E + G into F + EP → E + G.

If the enzymatic mechanism of uridine kinase is ordered ping-pong [23], the half-reactions read:

ATP + Urk = Urk-P + ADP;
dATP + Urk = Urk-P + dADP;
dUTP + Urk = Urk-P + dUDP;
dCTP + Urk = Urk-P + dCDP;
Urk-P + uridine = UMP + Urk;
Urk-P + cytidine = CMP + Urk.

We consider a system made up of uridine kinase and several other reactions of nucleotide metabolism (Fig. 1). Here, the number of elementary modes does not depend on the level of description

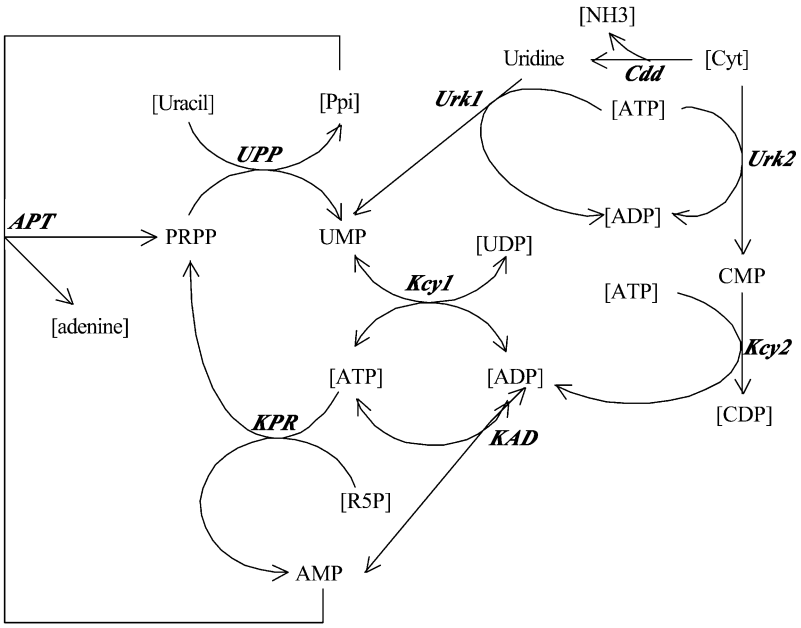

Fig. 1. Uridine kinase system. Enzymes: **APT**, adenine phosphoribosyltransferase (EC 2.4.2.7); **Cdd**, cytidine deaminase (EC 3.5.4.5); **KAD**, adenylate kinase (EC 2.7.4.3); **Kcy1**, cytidylate kinase (EC 2.7.4.14); **KPR**, ribose-phosphate pyrophosphokinase (EC 2.7.6.1); **UPP**, uracil phosphoribosyltransferase (EC 2.4.2.9); **Urk**, uridine kinase (EC 2.7.1.48). Metabolite names: ADP, adenosine diphosphate; AMP, adenosine monophosphate; ATP, adenosine triphosphate; CDP, cytidine diphosphate; CMP, cytidine monophosphate; Cyt, cytidine; NH3, ammonia; Ppi, diphosphate; PRPP, phosphoribosylpyrophosphate; R5P, ribose-5-phosphate; UDP, uridine diphosphate; UMP, uridine monophosphate. The arrow representing **KAD** corresponds to the production of two molecules of ADP. Symbols in brackets refer to external metabolites (sources and sinks).

or on the enzyme mechanism. In each case, there are four elementary flux modes, which in terms of overall reactions read: {Kcy2, Urk2}; {Kcy1, Cdd, Urk1}; {UPP, Kcy1, KAD, KPR}; {(2 UPP), (2 Kcy1), KPR, APT}. All of them are irreversible. Moreover, the convex basis here coincides with the set of elementary modes.

3. The case where some multifunctional enzymes involve reversible steps

A linear dependence among the various reactions of a multifunctional enzyme only occurs if these reactions share partly the same substances. For example, an enzyme catalyzing the reversible reactions $A \leftrightarrow B$ and $A \leftrightarrow C$ will also catalyze the reaction $B \leftrightarrow C$. In contrast, there is no linear dependence for an enzyme catalyzing the reactions $A \leftrightarrow B$, $C \leftrightarrow D$. The term 'bifunctional enzyme' should be used for an enzyme with two linearly independent functions.

A prominent example of a reversible multifunctional enzyme is transketolase (EC 2.2.1.1). Each of its monomers can catalyze the reactions [24] **Tkt1**: ribose-5P + xylulose-5P = sedoheptulose-7P + glyceraldehyde-P; **Tkt2**: erythrose-4P + xylulose-5P = fructose-6P + glyceraldehyde-P (where P stands for phosphate). Interestingly, a linear combination can be written: **Tkt3**: erythrose-4P + sedoheptulose-7P = xylulose-5P + fructose-6P (**Tkt1**, −**Tkt2**), which is also a bimolecular reaction. Usually, in the literature, only **Tkt1** and **Tkt2** are mentioned [24–26]. However, one can likewise indicate, for example, **Tkt1** and **Tkt3** as independent functions of transketolase.

We will now consider the simple hypothetical system shown in Fig. 3a. Reaction 2 is depicted in terms of hemi-reactions. Half-reaction b is

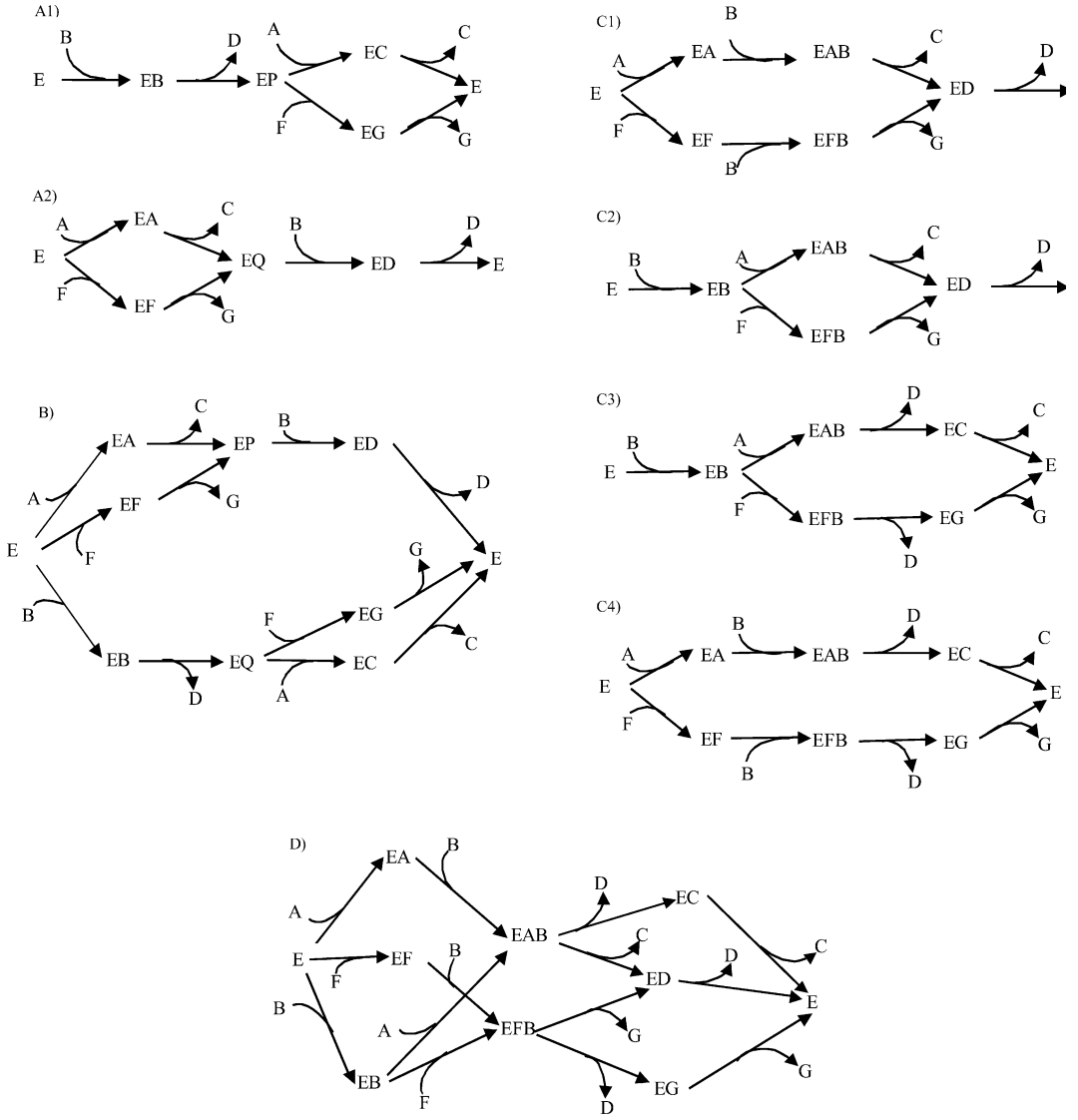

Fig. 2. Different reaction mechanisms for the bifunctional reaction A + B → C + D and F + B → G + D. (a1) Ordered ping-pong with low specificity in the lower part. (a2) Ordered ping-pong with low specificity in the upper part. (b) Random ping-pong. (c1–4) Ordered sequential with low specificity in various steps. (d) Random sequential.

reversible, while a and c are irreversible. Applying the algorithm for computing elementary modes [2,13] formally to the hemi-reactions system, the following modes are obtained:

$\{\mathbf{E}_1,\ \mathbf{a},\ -\mathbf{b}\},\ \{\mathbf{b},\ \mathbf{c},\ \mathbf{E}_3\},\ \{\mathbf{E}_1,\ \mathbf{a},\ \mathbf{c},\ \mathbf{E}_3\}.$

The last mode fulfils the conditions mentioned in the two definitions given in the Introduction provided that the components of **V** correspond to reaction steps. However, when steps **a** and **c** are operative, also step **b** can be operative because it belongs to the same enzyme, $\mathbf{E}_2$. Therefore, one can write the third mode as the sum of the first two and, accordingly, it is not elementary.

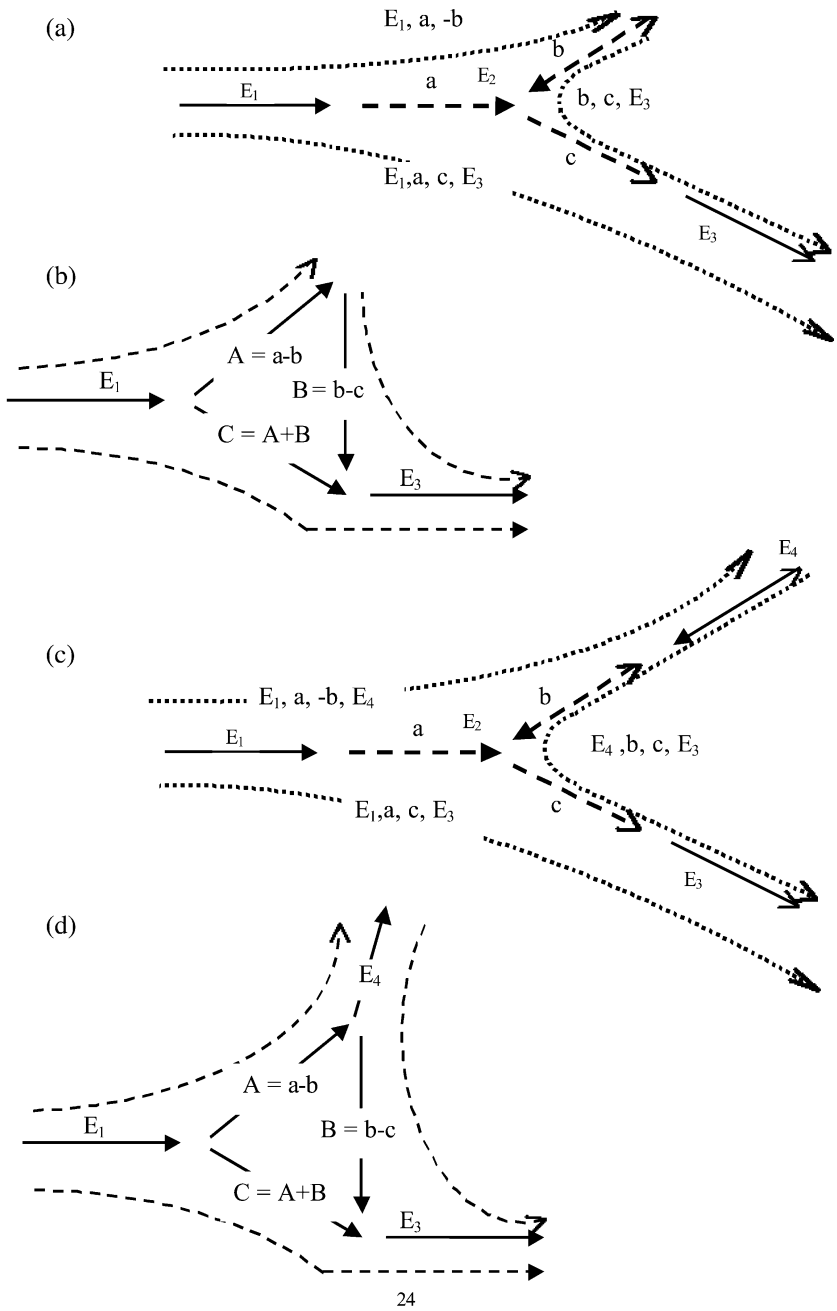

Fig. 3. Simple reaction schemes illustrating elementary modes and convex basis on different levels of description. (a) The multifunctional enzyme is represented by half-reactions (labeled by **a**, **b** and **c**). Arrows with single (double) arrow-heads refer to irreversible (reversible) reactions. The forward orientation of reversible reactions is symbolized by full arrow-heads. Elementary flux modes are depicted by dashed arrows. (b) The same system as in (a) in terms of overall reactions. (c) Half-reactions system with an additional enzyme, $\mathbf{E_4}$. (d) The same system as in (c) in terms of overall reactions.

The overall-reactions of enzyme $\mathbf{E_2}$ could be, for example:

$\mathbf{A}=\mathbf{a}-\mathbf{b}$, $\mathbf{B}=\mathbf{b}+\mathbf{c}$.

The third overall reaction, $\mathbf{C}=\mathbf{a}+\mathbf{c}$, is linearly dependent on **A** and **B**. Therefore, it should not be included on computation of pathways. The translation of the above-mentioned modes from the hemi-reactions system in terms of overall-reactions is (Fig. 3b):

$\{\mathbf{E}_1,\ \mathbf{A}\}$, $\{\mathbf{B},\ \mathbf{E}_3\}$, $\{\mathbf{E}_1,\ \mathbf{A}+\mathbf{B},\ \mathbf{E}_3\}$.

Now it can clearly be seen that the last one is, according to the definition of elementary flux modes, not elementary, because it is the sum of two modes involving fewer reactions. In the light of the verbal definition of elementary modes given in the Introduction, one may wish to write the modes in terms of enzymes rather than reactions: $\{\mathbf{E_1},\ \mathbf{E_2}\}$, $\{\mathbf{E_2},\ \mathbf{E_3}\}$, $\{\mathbf{E_1},\ \mathbf{E_2},\ \mathbf{E_3}\}$. However, the last of these is still non-elementary because it involves the former two as subsets.

We will now consider the convex basis for this example system. In the half-reactions description, the extreme pathways are $\{\mathbf{E_1},\ \mathbf{a},\ -\mathbf{b}\}$ and $\{\mathbf{b},\ \mathbf{c},\ \mathbf{E_3}\}$. The mode $\{\mathbf{E_1},\ \mathbf{a},\ \mathbf{c},\ \mathbf{E_3}\}$ is not an extreme pathway because it is the sum of these two pathways. For the given directionality of reactions, the extreme pathways are unique. However, if, for example, reactions **a** were reversible, uniqueness would be lost. We could then choose $\{\mathbf{E_1},\ \mathbf{a},\ -\mathbf{b}\}$, $\{-\mathbf{E_1},\ -\mathbf{a},\ \mathbf{b}\}$ and $\{\mathbf{b},\ \mathbf{c},\ \mathbf{E_3}\}$, or we could choose $\{\mathbf{E_1},\ \mathbf{a},\ -\mathbf{b}\}$, $\{-\mathbf{E_1},\ -\mathbf{a},\ \mathbf{b}\}$ and $\{\mathbf{E_1},\ \mathbf{a},\ \mathbf{c},\ \mathbf{E_3}\}$. In the overall description, if reaction **a** is irreversible, the convex basis reads $\{\mathbf{E_1},\ \mathbf{A}\}$, $\{\mathbf{B},\ \mathbf{E_3}\}$. If reactions $\mathbf{E_1}$ and **a** were reversible, we can take $\{\mathbf{E_1},\ \mathbf{A}\}$, $\{-\mathbf{E_1},\ -\mathbf{A}\}$, $\{\mathbf{B},\ \mathbf{E_3}\}$ or $\{\mathbf{E_1},\ \mathbf{A}\}$, $\{-\mathbf{E_1},\ -\mathbf{A}\}$, $\{\mathbf{C},\ \mathbf{E_3}\}$. Thus, in both cases, the number of extreme vectors is independent of the level of description. Moreover, for the overall reactions system, the convex basis can be obtained by translating the convex basis of the hemi-reactions system.

As mentioned above, the choice of linearly independent functions of a multifunctional enzyme is not a priori unique. In the example under study, we could take functions **A** and **C** instead of **A** and **B**. Formal application of the algorithm then yields the 'elementary modes' $\{\mathbf{E_1},\ \mathbf{A}\}$, $\{-\mathbf{A},\ \mathbf{C},\ \mathbf{E_3}\}$, $\{\mathbf{E_1},\ \mathbf{C},\ \mathbf{E_3}\}$. This result is not, however, correct because the third mode is not really elementary. It appears that this choice of linearly independent functions is not suitable. Instead, one should take the functions **A** and **B**. Importantly, the choice of these is derived from the convex basis, $\{\mathbf{E_1},\ \mathbf{a},\ -\mathbf{b}\}$ and $\{\mathbf{b},\ \mathbf{c},\ \mathbf{E_3}\}$ (with step **a** being irreversible). The convex basis and, thus, the choice of linearly independent functions are imposed on the multifunctional enzyme by the other enzymes in the system.

The above reasoning leads us to establishing a method for identifying the appropriate linearly independent functions.

1. Determine the convex basis in the half-reactions system.
2. See which combinations of half-reactions of the multifunctional enzyme occur and translate these into overall reactions.
3. Compute the elementary modes in the overall reactions system.

In this way, the elementary modes are uniquely defined and do not depend on the level of description. We recall that, on the half-reactions level, the definition of elementary modes should be applied in terms of enzymes rather than of half-reactions.

We now apply the above method to a network describing glycolysis and the pentose phosphate pathway (see [2] for the reaction equations) in which transketolase is described by half-reactions. These read **R1**: xylulose$-$5P$+$Tkt$=$glyceraldehyde$-$3P$+$TktC2; **R2**: ribose$-$5P$+$TktC2$=$sedoheptulose$-$7P$+$Tkt; **R3**: erythrose$-$4P$+$TktC2$=$fructose$-$6P$+$Tkt. This network gives a convex basis that involves six extreme pathways, three of which do not contain transketolase at all, while the others contain 2 **R1**$+$**R2**$+$**R3**. This can be simply translated into **Tkt1**$+$**Tkt2**. A translation involving **Tkt3** would be more complicated, for example, 2 **Tkt2**$+$**Tkt3**. Thus, the two functions usually given in the literature are in accordance with the above method.

It is interesting to consider the system depicted in Fig. 3c, which extends the system shown in Fig. 3a in that the substance converted by the reversible hemi-reaction **b** is also converted by an

additional enzyme, $\mathbf{E_4}$. For this system the elementary modes in terms of half-reactions read $\{\mathbf{E_1}, \mathbf{a}, -\mathbf{b}, \mathbf{E_4}\}$, $\{\mathbf{E_4}, \mathbf{b}, \mathbf{c}, \mathbf{E_3}\}$, $\{\mathbf{E_1}, \mathbf{a}, \mathbf{c}, \mathbf{E_3}\}$. The last one reads, in terms of overall reactions, $\{\mathbf{E_1}, \mathbf{A} + \mathbf{B}, \mathbf{E_3}\}$ (Fig. 3d). This mode is now elementary because it does not involve enzyme $\mathbf{E_4}$. In this situation, it is thus unnecessary to apply the above method.

4. Interconversion of nucleoside triphosphates

To illustrate the main point of this paper, we consider a biochemical system involving the enzyme nucleoside diphosphokinase (EC 2.7.4.6). This enzyme reversibly interconverts various nucleoside triphosphates as well as deoxy-nucleoside triphosphates and the corresponding diphosphates. For simplicity's sake, we here consider only the conversion of ATP, GTP, CTP and UTP. The general reaction equation reads

$$N_1TP + N_2DP = N_1DP + N_2TP \qquad (*)$$

with N_1 and N_2 denoting some nucleosides. As we here consider four different nucleosides, there are 16 different reaction equations of the type (*), of which only six are relevant for symmetry reasons and because N_1 and N_2 should be different.

Nucleoside diphosphokinase operates according to an ordered ping-pong mechanism [27] (see also database BRENDA). Therefore, in the formalism of half-reactions, the enzyme reaction can be written as

R1: ATP + NDK = ADP + NDKP
R2: GTP + NDK = GDP + NDKP
R3: UTP + NDK = UDP + NDKP
R4: CTP + NDK = CDP + NDKP.

Now we consider the reaction system shown in Fig. 4, which extends a network analyzed earlier [29]. For simplicity's sake, we do not indicate transport reactions between different organelles. The network includes, in addition to nucleoside diphosphokinase, several other enzymes of nucleotide interconversion. One of these, adenylate kinase (KAD), is multifunctional as well,

KAD1: ATP + AMP = ADP + ADP;
KAD2: GTP + AMP = GDP + ADP;
KAD3: CTP + AMP = CDP + ADP;
KAD4: UTP + AMP = UDP + ADP.

Its hemi-reactions read:

R5: ATP + KAD = ADP + KADP;
R6: GTP + KAD = GDP + KADP;
R7: UTP + KAD = UDP + KADP
R8: CTP + KAD = CDP + KADP;
R9: ADP + KAD = AMP + KADP.

The remaining enzyme reactions are:
UDP: UTP + G1P = UDP-Glc + Ppi;
CTP: CTP + CholP = CDP-CholP + Ppi;
CPT: CDP-CholP + DAG = CMP + Pchol;
KIC: CholP + ADP = Chol + ATP;
Kcy: CDP + ADP = CMP + ATP;
PGM: G6P = G1P;
IPY: PPi = 2 Pi;
UGS: UDP-Glc = Glyc + UDP.

For such a large system, the comparison of elementary modes at the two levels of description is not straightforward. Therefore, we implemented an algorithm which translates the elementary flux modes of the overall-reactions system into elementary flux modes in terms of hemi-reactions and compares this set with the set of elementary modes computed for the hemi-reactions, identifying the difference. For writing this program we used Lex (Lexical Analyzer Generator), with host language C. Our program is available upon request.

According to the method given in Section 3, we first determined the convex basis for the half-reactions system. This involves six vectors. However, three of these are irrelevant because they represent cyclic flows and perform no net transformation (e.g. **R1**, $-$**R2**, $-$**R5**, **R6**). The vectors of the convex basis contain the following combinations of hemi-reactions belonging to the same enzyme: (**R1**, $-$**R2**), (**R1**, $-$**R3**), (**R1**, $-$**R4**), ($-$**R5**, **R6**), (**R5**, $-$**R7**), (**R5**, $-$**R8**), (**R5**, $-$**R9**). In this way, we define the linearly independent functions of KAD and NDK:

NDK1 = **R1** − **R3**: ATP + UDP = ADP + UTP;
NDK2 = **R1** − **R2**: ATP + GDP = ADP + GTP;
NDK3 = **R1** − **R4**: ATP + CDP = ADP + CTP;
KAD1 = **R5** − **R9**: ATP + AMP = 2 ADP;
KAD5 = **R5** − **R8**: ATP + CDP = ADP + CTP;
KAD6 = **R5** − **R7**: ATP + UDP = ADP + UTP;
KAD7 = −**R5**+**R6**: ADP + GTP = ATP + GDP.

Out of the six reactions of NDK mentioned above, only three are linearly independent because

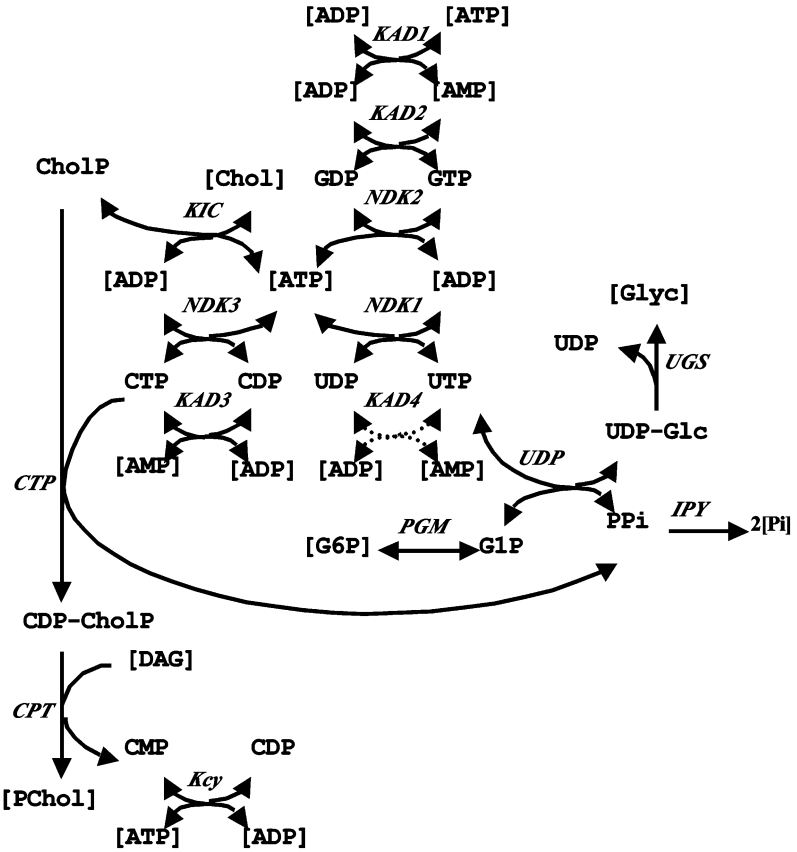

Fig. 4. Interconversion of nucleoside triphosphates. Enzymes: **NDK**, nucleoside-diphosphate kinase (EC 2.7.4.6); **UDP**, UTP-glucose-1-phosphate uridylyltransferase (EC 2.7.7.9); **PGM**, phosphoglucomutase (EC 5.4.2.2); **KIC**, choline kinase (EC 2.7.1.32); **CTP**, choline phosphate cytidylyltransferase (EC 2.7.7.15); **CPT**, diacylglycerol cholinephosphotransferase (EC 2.7.8.2); **IPY**, inorganic pyrophosphatase (EC 3.6.1.1); **UGS**, glycogen synthase (EC 2.4.1.11); **Kcy**, cytidylate kinase (EC 2.7.4.14); **KAD**, adenylate kinase (EC 2.7.4.3). Metabolite names other than those in Fig. 1: Chol, choline; CholP, cholinephosphate; CTP, cytidine triphosphate; DAG, 1,2-diacylglycerol; G1P, glucose 1-phosphate; G6P, glucose 6-phosphate; GDP, guanosine diphosphate; Glyc, glycogen; GDP, guanosine diphosphate; GTP, guanosine triphosphate; PChol, phosphatidylcholine; UDP-Glc, UDP-glucose; UTP, uridine triphosphate.

we have eight different metabolites and five conservation relations (for the ADP, GDP, CDP, UDP, and phosphate moieties). Importantly, we have obtained exactly three independent overall reactions from the convex basis. The reaction **NDK4**: GTP + CDP = GDP + CTP, for example, is the sum of the reactions **NDK3** and (−**NDK2**). Note that **NDK1**, **NDK2** and **NDK3** are all reversible. As KAD involves one additional metabolite (AMP) and implies the same number of conserved moieties (with AMP instead of ADP being a conserved moiety), the number of linearly independent functions is four. This is in agreement with the functions obtained above.

Using these functions, computation of elementary modes by the program METATOOL gives eight modes (Table 1). However, we have to cancel three of these (modes 2, 3 and 6) because they do not perform any net transformation. Cyclic modes that cannot operate due to thermodynamic reasons are not eliminated by METATOOL. To decide whether these modes can operate, one needs to know whether they are driven by a free-energy difference between external metabolites, which

Table 1
List of elementary modes for the reaction scheme of nucleotide metabolism shown in Fig. 4

No	Elementary flux modes in overall reaction system in terms of half-reactions	Elementary flux modes in terms of overall reaction	Directionality
1	**−R5 R9**	**−KAD1**	Reversible
2	**−R1 R2 R5 −R8**	**−NDK2 KAD5**	Reversible[a]
3	**−R1 R3 R5 −R7**	**−NDK1 KAD6**	Reversible[a]
4	**R1 −R3 UDP PGM IPY UGS**	**NDK1 UDP PGM IPY UGS**	Irreversible
5	**−R5 R7 UDP PGM IPY UGS**	**−KAD6 UDP PGM IPY UGS**	Irreversible
6	**R1 −R4 −R5 R6**	**NDK3 KAD7**	Reversible[a]
7	**R1 −R4 −Kcy −KIC CTP IPY CPT**	**NDK3 −Kcy −KIC CTP IPY CPT**	Irreversible
8	**R5 −R6 −Kcy −KIC CTP IPY CPT**	**−KAD7 −Kcy −KIC CTP IPY CPT**	Irreversible

[a] These modes are cyclic and perform no net transformation. Therefore, they should be deleted. They occur because several of the functions of NDK and KAD are identical.

may have been omitted in the reaction equations for the sake of simplicity. It now becomes clear that in the presence of multifunctional enzymes, elementary modes have to be checked carefully whether they realize a net transformation of external substances. Otherwise, non-elementary modes would be obtained. For example, modes 2, 3 and 6 in Table 1 include mode no. 1 as a subset, and so do modes 5 and 8. The latter two can be cancelled because they perform the same net transformation as modes 4 and 7, respectively. Also in the convex basis, irrelevant cyclic modes have to be cancelled. The number of relevant extreme pathways is three in the overall reactions system and, thus, equal to the number obtained above for the hemi-reactions system.

This example shows on a larger scale than the hypothetical example discussed in Section 3 how to deal with reversible multifunctional enzymes. Here, the reason for a special treatment of such enzymes is, however, different. There are sets of substrates of two multifunctional enzymes, which overlap. KAD and NDK both use ATP, ADP, UTP, UDP, GTP, GDP, CTP and CDP. Therefore, if we computed the elementary modes for the hemi-reactions system formally (that is, without considering the enzymes as basic units), we would obtain, for example, the mode {**R2**, **−R3**, **R5**, **−R6**, **UDP**, **PGM**, **Ppase**, **UGS**}. However, the set of hemi-reactions {**R2**, **−R3**, **R5**, **−R6**} performs the same overall transformation as the hemi-reactions **R1** and **−R3**, so that this mode is equivalent to mode 4 in Table 1.

5. Discussion

Multifunctional enzymes (in the sense of enzymes with low substrate specificity) are ubiquitous in living cells. Prominent examples are alcohol dehydrogenase, aldolase, hexokinase, transketolase, uridine kinase, cytidylate kinase, nucleoside diphosphokinase and adenylate kinase. Many other enzymes, which are commonly considered as monofunctional, actually catalyze a number of side reactions, the so-called underground metabolism [28]. To our eyes, these facts have not always been duly taken into account in the modeling of metabolic networks. Here, we have investigated how to treat multifunctional enzymes in metabolic pathway analysis.

We have shown that the various reactions that can be performed by a multifunctional enzyme catalyzing reversible bimolecular reactions are usually linearly dependent on each other. When using overall reactions in pathway analysis, it is important to consider only linearly independent reactions. Otherwise, cyclic elementary modes are obtained which do not perform any net transformation and have to be cancelled afterwards. For example, including **NDK4** into the set of reactions for the system shown in Fig. 4, we will obtain a spurious elementary mode consisting only of enzyme NDK (with the functions **NDK2**,

−**NDK3**, **NDK4**) and performing no net transformation at all.

An algorithm for choosing, in a unique way, an appropriate set of linearly independent functions has been given. We suggest that such information could be included in metabolic databases. At present, in a number of databases, e.g. ExPASy−ENZYME (http://www.expasy.ch/enzyme/) and partly in the literature, the set of possible substrates and products of a multifunctional enzyme are indicated without specifying the particular reactions, especially for enzymes with a high number of functions. ExPASy−ENZYME indicates that many nucleoside diphosphates can act as acceptors in the nucleoside diphosphokinase reaction, and many ribo- and deoxyribonucleoside triphosphates can act as donors. Obviously, the number of pathways would be much greater than the number obtained in Section 4 if all functions of this enzyme were included.

In all examples we tested so far, the number of different functions of multifunctional enzymes as derived from the convex basis of the half-reactions system is equal to the number of linearly independent functions as given by the rank of the stoichiometry matrix. It will be a subject of future studies to prove that this is true in general for multifunctional enzymes involving at least one reversible reaction.

Our main result is that the number of pathways (defined as elementary flux modes) does not depend on the level of description of multifunctional enzymes provided that the definition of elementary modes is applied correctly. This requires that enzymes are considered as the basic units in metabolism. We have to find a minimal number of enzymes rather than a minimal number of reaction steps. Note that the definition of elementary modes is based on the principle of genetic independence introduced by Seressiotis and Bailey [16]. This says that upon decomposition of a mode into two other modes, no enzymes should be used in addition that are not used by the mode in question because the other modes then would be genetically independent. Formal application of the definition on the level of hemi-reactions would imply that the principle of genetic independence would hold for hemi-reactions, which is not the case. This point is also related to the concept of activity set introduced by Nuño et al. [15]. This is the set of all enzymes operative in a given state. Note that in the presence of multifunctional enzymes, this set involves fewer items than nonzero components are contained in the vector **V** because several components refer to the same enzyme.

We have illustrated these theoretical considerations by several examples. Another example is the reaction system of monosaccharide metabolism studied in [15]. In terms of overall reactions, this system gives rise to 296 elementary flux modes, while in terms of half-reactions, 866 modes would arise if the reaction steps were considered as basic units. Note that Nuño and coworkers [15] analyzed the convex basis. This set involves eight extreme pathways in both descriptions. Also for the other examples, we have found that the number of extreme pathways is independent of the level of description even if the definition is applied to reaction steps. It can be shown that this holds for any reaction system because the extreme pathways have the property that none of them is a nonnegative linear combination of other extreme pathways. The same property holds after translating the extreme pathways of the half-reactions system into those of the overall reactions system. When a system only contains irreversible steps, the set of elementary modes coincides with the convex basis [12]. Therefore, such systems do not cause any problems with respect to a description on different levels of detail. As illustrated by an example in Section 2, the same is true as long as only monofunctional enzymes are reversible.

Both sets of fundamental pathways have their advantages: the convex basis usually involves a smaller number of pathways, but they are sometimes not uniquely determined [17] and they often do not cover all biochemically relevant routes [13]. Moreover, they may change considerably upon addition or deletion of reactions. For example, if reaction **b** in Fig. 3a is deleted, the previous extreme vectors do no longer apply and the new extreme vector {**a**, **c**} arises. The set of elementary modes is uniquely determined and involves the full set of potential basic routes in the system. Moreover, if enzymes are added to the system,

new elementary modes may arise while the previously existing modes remain unchanged [13]. Upon deletion of enzymes, some modes may disappear while none of the remaining modes will be changed. While the convex basis is always independent of the level of description, the elementary modes require a correct application of the definition in order that the same property holds for these.

In future studies, it is worth considering that different reactions of a multifunctional enzyme can proceed depending on conditions (e.g. availability of initial substrate). However, the general results of the present paper will not be affected by this specification. Moreover, the approach presented here may be of interest for the modeling of enzyme evolution because it is generally assumed that the multitude of present-day monofunctional enzymes evolved from a smaller number of enzymes with broader substrate specificity.

Acknowledgments

We would like to thank Dr R. Heinrich and Dr H.-G. Holzhütter (Berlin) for very stimulating discussions. Financial support to both authors from the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

References

[1] C.H. Schilling, D. Letscher, B.O. Palsson, Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective, J. Theor. Biol. 203 (2000) 229–248.

[2] S. Schuster, D.A. Fell, T. Dandekar, A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks, Nature Biotechnol. 18 (2000) 326–332.

[3] G.N. Stephanopoulos, A.A. Aristidou, J. Nielsen, Metabolic Engineering. Principles and Methodologies, Academic Press, San Diego, 1998.

[4] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.-L. Barabási, The large-scale organization of metabolic networks, Nature 407 (2000) 651–654.

[5] J.M. Rohwer, F.C. Botha, Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data, Biochem. J. 358 (2001) 437–445.

[6] S.J. Van Dien, M.E. Lidstrom, Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph *Methylobacterium extorquens* AM1, with application to reconstruction of C_3 and C_4 metabolism, Biotechnol. Bioeng. 78 (2002) 296–312.

[7] T. Dandekar, S. Schuster, B. Snel, M. Huynen, P. Bork, Pathway alignment: application to the comparative analysis of glycolytic enzymes, Biochem. J. 343 (1999) 115–124.

[8] C.H. Schilling, B.O. Palsson, Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis, J. Theor. Biol. 203 (2000) 249–283.

[9] B.O. Palsson, The challenges of in silico biology, Nature Biotechn. 18 (2000) 1147–1150.

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

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

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

[13] S. Schuster, C. Hilgetag, J.H. Woods, D.A. Fell, Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism. J. Math. Biol., in press.

[14] R. Heinrich, S. Schuster, The Regulation of Cellular Systems, Chapman and Hall, New York, 1996.

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

[16] A. Seressiotis, J.E. Bailey, MPS: An artificially intelligent software system for the analysis and synthesis of metabolic pathways, Biotechnol. Bioeng. 31 (1988) 587–602.

[17] T. Pfeiffer, I. Sánchez-Valdenebro, J.C. Nuño, F. Montero, S. Schuster, METATOOL: For studying metabolic networks, Bioinformatics 15 (1999) 251–257.

[18] A.S. Liacouras, E.P. Anderson, Uridine-cytidine kinase, Arch. Biochem. Biophys. 168 (1975) 66–73.

[19] A.S. Liacouras, T.Q. Garvey, F.K. Millar, E.P. Anderson, Uridine-cytidine kinase. Kinetic studies and reaction mechanism, Arch. Biochem. Biophys. 168 (1975) 74–80.

[20] R.A. Alberty, Standard apparent reduction potentials for biochemical half reactions as a function of pH and ionic strength, Arch. Biochem. Biophys. 389 (2001) 94–109.

[21] R.A. Alberty, Constraints in biochemical reactions, Biophys. Chem. 49 (1994) 251–261.

[22] R.C. Payne, N. Cheng, T.W. Traut, Uridine kinase from *Ehrlich Ascites Carcinoma*. Purification and properties of homogeneous enzyme, J. Biol. Chem. 18 (1985) 10242–10247.

[23] A. Orengo, Regulation of enzymic activity by metabolites. I. Uridine-cytidine kinase of *Novikoff ascites* rat tumor, J. Biol. Chem. 8 (1969) 2204–2209.
[24] L. Stryer, Biochemistry, Freeman, New York, 1995.
[25] E.L. Smith, R.L. Hill, I.R. Lehman, R.J. Lefkowitz, P. Handler, A. White, Principles of Biochemistry. General Aspects, McGraw-Hill, New York, 1983.
[26] M. Yudkin, R. Offord, A Guidebook to Biochemistry, Cambridge, University Press, Cambridge, 1980.
[27] M.G. Colomb, A. Cheruy, P.V. Vignais, Nucleoside diphosphokinase from beef heart cytosol. II. Characterization of the phosphorylated intermediate, Biochemistry 11 (1972) 3378–3386.
[28] R. D'Ari, J. Casadesus, Underground metabolism, Bioessays 20 (1998) 181–186.
[29] S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, T. Dandekar, Exploring the pathway structure of metabolism: Decomposition into subnetworks and application to *Mycoplasma pneumoniae*, Bioinformatics 18 (2002) 351–361.