

Network organization of cell metabolism: monosaccharide interconversion

Juan C. NUÑO*†, Ignaciol SÁNCHEZ-VALDENEBRO*, Carolina PÉREZ-IRATXETA*, Enrique MELÉNDEZ-HEVIA‡ and Francisco MONTERO*§

*Departamento de Bioquímica y Biología Molecular I, Facultad de CC. Químicas, Universidad Complutense de Madrid, E-28040 Madrid, †U.D. de Matemáticas, E.T.S.I. de Montes, Universidad Politécnica de Madrid, E-28040 Madrid, and ‡Departamento de Bioquímica, Facultad de Biología, Universidad de La Laguna, E-38206-Tenerife, Spain

The structural properties of carbohydrate metabolism are being studied. The present contribution focuses mainly on those processes involving the transfer of carbon fragments among sugars. It is shown how enzymic activities fix the way the system self-organizes stoichiometrically at the steady state. It is proven that there exists a specific correspondence between the set of all possible enzymic activities, the activity set, and the set of stoichiometrically compatible flux distributions through the pathway. On the one hand, there are enzymic activities that do not allow a stoichiometrically feasible coupling at the steady state of the reactions involved in the conversion. On the other

hand, there are enzymic activities that are related to one or more flux distributions at the steady state (i.e. with one or several rate vectors respectively). For this latter group, it can be demonstrated that the structure of the system depends on other non-structural factors, such as boundary constraints and the kinetic parameters. As a consequence, it is suggested that this kind of metabolic process must be viewed as a complex reaction network instead of a sequential number of steps. Some implications of these derivations are illustrated for the particular conversion $\text{CO}_2 \rightarrow \text{C}_3$. General remarks are also discussed within the framework of network models of cell metabolism.

INTRODUCTION

Hexoses (C_6) and pentoses (C_5), free or combined, are the most abundant carbohydrates in Nature. They play specific metabolic roles: hexoses are the energy and main structural sugars, whereas pentoses are informative material (although these may also form a few structural polysaccharides, such as xylose and arabinose derivatives). Thus the sugar conversion $\text{C}_6 \rightleftharpoons \text{C}_5$ means the conversion (interchange) of energy-information material. Although in present cells sugar transformation is coupled to different functions, e.g. the synthesis of fatty acids, in the evolution of metabolism its first function (the reason to be created) was surely the synthesis of ribose to make nucleic acids, from the pre-existing glucose. Whereas prebiotic synthesis of glucose has been experimentally demonstrated by the formose reaction [1,2], the synthesis of ribose has been barely recognized under such conditions [3,4]. Therefore, the origin of ribose might have been not from prebiotic chemistry, but as a consequence of a primitive metabolism.

In addition to hexoses and pentoses, carbohydrate metabolism also involves sugars with three (C_3), four (C_4), seven (C_7) and eight (C_8) carbons. In particular, C_3 and C_4 have important metabolic roles, such as in the interconnection among glycolysis, the Calvin cycle and the pentose phosphate cycle (C_3), or in the biosynthesis of aromatic amino acids (C_4); C_7 and C_8 sugars exist only as intermediates of particular sugar metabolisms, although actually the latter are rarely to be found in present-day cells [5]. Sugars with more carbons do not exist in Nature. In any case, all of them are metabolites just for passing through, and do not accumulate in living material.

In all cases described in the literature, monosaccharide interconversion implies the presence of the same kind of enzymes (although not all of them are always involved). The enzymes catalyse the transfer of carbon fragments with, in most cases, similar mechanisms. These enzymes are known as transketolase (TK), transaldolase (TA) and aldolase (AL). TK and TA catalyse

those reactions implying the transfer of carbon fragments (of two carbons with TK and of three carbons with TA) from one sugar to another. AL catalyses the condensation or decondensation of two sugars, a triose always being involved. It is important to recall that these transfers are carried out through intermediary complexes formed by the enzymes linked to fragments of two or three carbons. The previously studied metabolic designs differ in the specificity of the enzymes in acting on particular sugars as either donors or acceptors of the carbon fragments. In addition, in all these routes some intermediates act more than once as either donor or acceptor of several reactions. For instance, in the classical pentose phosphate cycle (F-type) the fructose 6-*P* can be formed as a product of TK- and TA-catalysed reactions, and the glyceraldehyde 3-*P* can be produced in the reactions carried out by TK acting on xylulose 5-*P* and those in which TA acts on fructose 6-*P*.

As a consequence of these facts, sugar interconversion has a special level of complexity, which is, in its fundamental aspects, different from that found in linear or branched pathways. This complexity is reflected in both the structural and the dynamic properties of the system. The former focuses on sequence, relationships among the intermediates and stoichiometric flux distribution at steady state of the pathway, compatible with specific enzymic activities. The latter focuses on the global behaviour of the system (whole flux through the route, transition time, etc.) according to boundary conditions, such as constant chemical affinity or constant input flux. The study of any metabolic pathway should consider these two aspects.

Figure 1 shows a general view of the relationships between these sugars, including interaction with other pathways. Thus any particular design, such as the Calvin cycle or the F- and L-types [6] of the pentose phosphate pathway can be derived from the general scheme shown in Figure 1 by selecting the appropriate set of enzymes and defining the boundary conditions. A good example of fitting this general scheme to a particular function is the enzymic regulation and the inhibition by light of TA activity

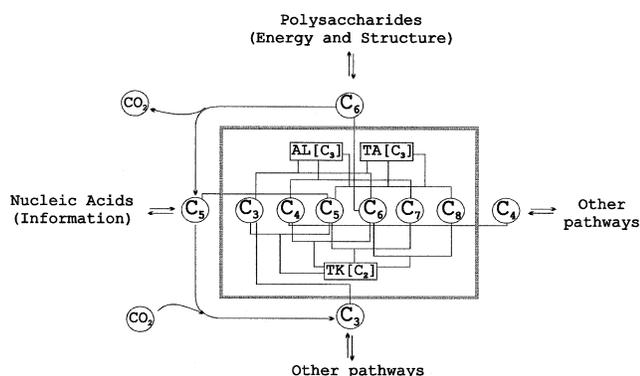


Figure 1 Schematic representation of monosaccharides interconversion as a complex reaction network

Nodes (CO_2 , sugars and enzymic complexes) are linked by lines (reversible reactions) and arrows (irreversible reactions). Only reactions of types (b) (inside the box) and (a) (outside the box) are shown. The connection with external sources and other metabolic pathways by diffusive processes is also depicted. As a consequence of this network structure, the system presents multiple ways of self-organizing according to the external couplings and the final product the cell requires (see the text).

[7], as occurs in chloroplast [8,9]. Regarding all above-mentioned cases and other examples (see, for instance, [10]), cellular metabolism can be considered as a network of enzymic reactions connected through intermediate metabolites, whose co-ordinate activity strongly depends on the stoichiometric and external constraints. As a consequence of this structural complexity there can emerge a functional versatility. For instance, the same enzymes allow the conversion of both C_5 sugars into C_6 sugars (non-oxidative phase of the pentose phosphate cycle) and C_3 sugars into C_5 sugars (non-reductive phase of the Calvin cycle).

The aim of this paper is to analyse the structural complexity of the metabolic sugar network shown in Figure 1 by exploring the couplings between the enzymic reactions involved in the system under several external constraints. By means of techniques of stoichiometric network analysis [11–14] is found the set of flux distributions that are stoichiometrically compatible with all the enzymic activities involved in the conversion. The specific relationship among enzymic activities and flux distributions through the pathway is also demonstrated. As an example of application of this analysis, the production of trioses from CO_2 is studied.

ANALYSIS

Kinetic model

Enzymic activities and mechanism

The reactions involved in this metabolic system can be classified into four different kinds: (a) Reactions that produce a global variation in the number of carbons of the system through mechanisms of carboxylation and decarboxylation respectively. These reactions are catalysed by specific enzymes, such as Rubisco for carboxylation and phosphogluconate dehydrogenase for decarboxylation. (b) Reactions that change the number of carbons of the sugars by transferring a given number of carbons between two sugars, but without a global variation of carbons. These are catalysed by TK, TA and AL. (c) Redox reactions, necessary since the net gain or loss of carbon is made through C_1 units as CO_2 , as stated in (a) above; therefore sugar metabolism must involve this kind of reaction for the conversion (CH_2O) \rightleftharpoons

CO_2 . The most characteristic enzymes involved in this step are 6-*P*-glucose dehydrogenase for oxidation and glyceraldehyde-3-phosphate dehydrogenase for reduction. (d) Auxiliary reactions that account for some rearrangements in molecular structure. They can be needed to allow subsequent reactions. For instance, addition, removal and transfer of phosphate groups by means of the action of kinases, phosphatases and phosphotransferases respectively, to be used as leaving groups for substitution reactions, or to give a degree of polarity to sugars; rearrangements of the stereochemical properties of the sugars (epimerases); change of the function aldo–keto (isomerases); etc.

It must be noted that, whereas reactions (b) and (d) are localized inside the box shown in Figure 1, reactions (a) and (c) occur outside it. This classification has been made regarding the chemical reactions, not the enzymes, since the enzymes can be considered as tools to accelerate some chemical process that previously existed [15]. For example, in the oxidative phase of the pentose phosphate cycle there is no enzyme that specifically catalyses the decarboxylation reaction, as it can occur spontaneously on the 3-oxo-6-phosphogluconate.

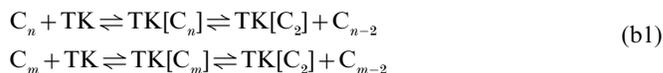
These reactions can be arranged in several ways, producing metabolic designs according to the couplings of the pathway, the external sources and the product that the cell needs (ribose, erythrose, NADPH, etc.). In addition, this general system can achieve a particular function by means of a different organization of the reactions. The pentose phosphate cycle is usually considered to be the most characteristic route where such a feature occurs (see, for instance, [5,16]), but this behaviour is not exclusive to this pathway. In fact, it occurs in every pathway to a greater or lesser degree. It is easy to see that the structural features of this network will mainly depend on reactions (a) and (b), that is to say, those that imply carbon transfer. Type (a) reactions can be assumed to be a single step and irreversible. Obviously, this simplification does not affect the structural properties of the network, although it might influence its dynamics. As will be seen later, irreversibility has important consequences on the flux distribution through the pathway, and therefore on the network structure.

As stated above, three different enzymes take part in reactions of type (b): TK, TA and AL. Usually, reactions in which TK is involved are schematically written as the single step:



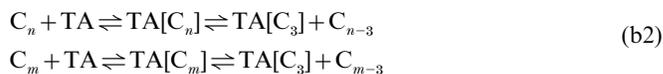
where n and m are numbers of carbons.

However, this must be interpreted as the coupling of two hemireactions:

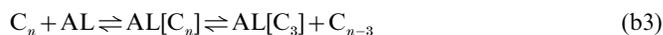


both working in opposite directions. It is worth pointing out that considering these hemireactions will have important consequences for the final structure of the network, since, as will be seen later, the stoichiometric coupling of these hemireactions can occur only under specific conditions.

A similar reasoning applies to TA. The two hemireactions carried out by this enzyme are:



The enzymic mechanism of AL has a subtle difference with respect to TK and TA; similarly to them, it reacts according to the scheme.



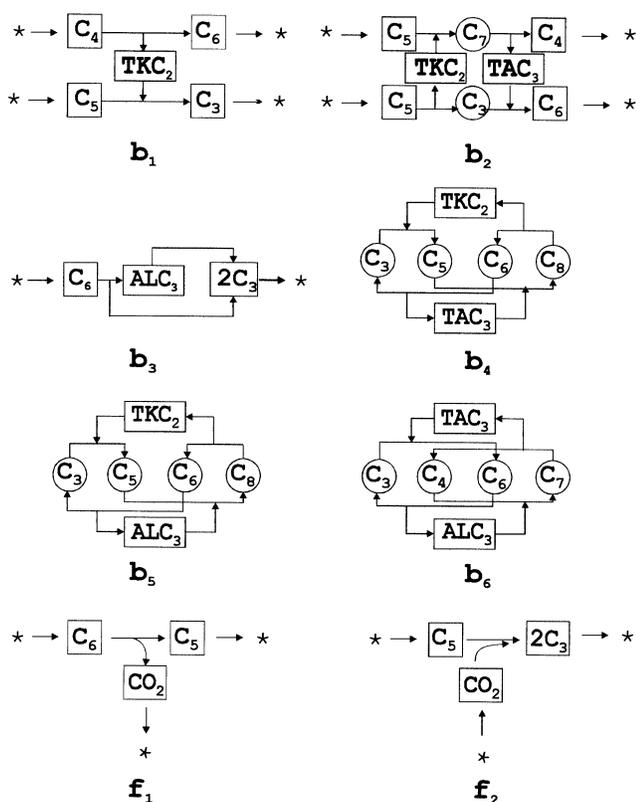


Figure 2 Schematic design of the eight fundamental flux distributions chosen as the basis

The eight fundamental flux distributions that form the simple generator set deduced in the text are schematically drawn here. Each of them represents an autonomous way of transforming sugars subject to the boundary conditions [reactions (1–5)]. Whereas schemes \mathbf{b}_1 , \mathbf{b}_2 , \mathbf{b}_3 , \mathbf{f}_1 and \mathbf{f}_2 can give rise to a net transformation of one or several sugars (boundary set), \mathbf{b}_4 , \mathbf{b}_5 and \mathbf{b}_6 are futile cycles that do not account for a net transformation (internal set). As can be seen, irreversibility appears in the boundary set through reaction schemes \mathbf{f}_1 and \mathbf{f}_2 . It is worth pointing out the role of the enzymic complexes as common pools in the reaction schemes (see the text). Stars indicate the system environment.

but unlike those enzymes, the enzymic complex $\text{AL}[\text{C}_3]$ can itself react to recover the enzyme plus one triose



The coupling of the two last reactions can produce the condensation of a triose (C_3) and a C_{n-3} sugar ($n = 6, 7$ or 8), or the opposite case. Moreover, the coupling of two hemireactions such as reaction (b3) can give rise to the transfer of a fragment of three carbons from a sugar to another, in a similar way to that of TA. Nevertheless, a fundamental difference between both mechanisms is that, whereas TA acts on a donor monophosphorylated sugar, AL reacts with a biphosphorylated sugar. It might reasonably be assumed that the enzyme TK can act on C_5 , C_6 , C_7 and C_8 and TA and AL on C_6 , C_7 and C_8 , so in this paper all these activities are considered.

Due to the particular enzymic mechanisms of the reactions, three common pools of enzymic complexes, $\text{TK}[\text{C}_2]$, $\text{TA}[\text{C}_3]$ and $\text{AL}[\text{C}_3]$ appear. Sharing these enzymic complexes will have important consequences for the flux distribution through the network, as will be seen later. On the other hand, although the existence of the complexes $\text{TK}[\text{C}_n]$, $\text{TA}[\text{C}_n]$ and $\text{AL}[\text{C}_n]$ will play a decisive role in the dynamic aspects of the pathway, their influence on the structural features of the network is null (in fact,

formally they can be obtained as linear combinations of the rest of intermediates). Therefore, these enzymic complexes are not considered in the kinetic formulation of the problem.

Before entering the formal analysis of the problem, let us consider the perspective under which monosaccharide interconversion must be viewed: a network formed by a number of nodes (the intermediates) related by lines (reactions), as shown in Figure 1.

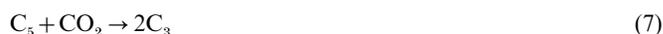
Description of the model

As can be seen in Figure 1, two kinds of processes must be taken into account in order to model the monosaccharide transformations. On the one hand are the transport processes that relate the compounds involved in the route to the surroundings. These processes can be schematically written as follows:

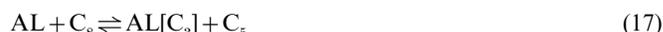


Here * denotes the system environment. On the other hand are reactions that are carried out by particular enzymes and that occur in a specific place within the cell. These reactions are of the types (a) and (b) already described above, and can be summarized in the following steps.

(a) Carboxylation and decarboxylation reactions:



(b) Reactions of carbon transfer:



Within this group, the following specific $\text{AL}-\text{C}_3$ reaction must be taken into account:



This setup addresses monosaccharide interconversion in the most general way, in the sense that it assumes that all reactions can occur simultaneously (which implies the presence of enzymes catalysing all these reactions). Nevertheless, it is well known that in present-day cells there is a particular distribution of enzyme activities giving specific reaction schemes. This concrete relationship will be investigated in detail below.

The system is related to the surroundings by means of the boundary processes (1–5) that take into account the (out/in)put of carbon compounds. In particular, the model considers that only sugars with 3, 4, 5 and 6 carbons can flow out of or into the system. In addition, the model considers that reactions (6) and (7) are irreversible, whereas the rest of reactions are reversible.

The i th reaction is governed by net velocities v_i ($i = 1, 2, \dots, 18$), where $v_i = v_{i+} - v_{i-}$, v_{i+} and v_{i-} being the forward and reverse velocities, always non-negative. Obviously, irreversible reactions (6) and (7) verify that $v_{6-} = v_{7-} = 0$. In general, they depend on the species concentration, kinetic constants and the enzyme concentrations. However, as will be seen later, in particular situations the relationship between these reaction rates is independent of these parameters, i.e. the flux distribution through the network depends exclusively on its relationship with the surroundings [i.e. reactions (1–5)]. In addition, mass conservation imposes the following restriction on the velocities under steady-state conditions:

$$v_1 + 3v_2 + 4v_3 + 5v_4 + 6v_5 = 0 \quad (19)$$

where v_1 , v_2 , v_3 , v_4 and v_5 are the reaction rates of reactions (1–5) respectively. As a consequence, the global stoichiometry of the system is not fixed.

Stoichiometric analysis

As with any set of chemical reactions, monosaccharide interconversion has to work under stationary conditions according to stoichiometric constraints. Each kind of sugar transformation under a given set of particular kinetic parameters and boundary constraints has a global stoichiometry that imposes the flux distribution through the net. These structural features, obviously time independent, must be reflected in the stationary behaviour of the system. Then, the organization of the flow through this network must be found from the study of the steady-state solutions of the system. In general, these solutions must depend on both the relationship of the system with the surroundings and the network architecture. Under stationary conditions, all stoichiometrically compatible network organizations imply a particular link among hemireactions [eqns (1–18)]. Asking for the ways of coupling reactions (1–18) that are stoichiometrically correct is equivalent to searching the set of reaction rates v_i that allow the system to be in the stationary regime. Notice that under steady-state conditions the flux arriving at each node must be necessarily equal to that leaving the node. In this scheme, a particular solution can be represented by a flux distribution through the network. The degree of coupling between two hemireactions can be obtained by comparing the flow through the corresponding routes.

It is simple to find those velocity distributions that satisfy the global stoichiometry of the system at the steady state. In fact, as shown in the Appendix, the problem can be solved algorithmically from the stoichiometric matrix. It is interesting to note that all the flux distributions stoichiometrically compatible with the system constraints can be calculated as linear combinations of eight particular and independent flux distributions. In other words, all the possible ways of getting the interconversion among monosaccharides can be achieved by linear combination of eight different routes. As is discussed in the Appendix, the choice of these eight independent flux distributions is arbitrary. Among all physical criteria, here we have used that of simplicity. Simplicity is used here to mean minimal number of reactions needed to form a stoichiometrically feasible pathway. In Figure 2, these fundamental flux distributions are schematically depicted, and are named as \mathbf{b}_1 , \mathbf{b}_2 , \mathbf{b}_3 , \mathbf{b}_4 , \mathbf{b}_5 , \mathbf{b}_6 , \mathbf{f}_1 and \mathbf{f}_2 . It follows that an infinite number of solutions appears, all the linear combinations of these fundamental routes, forming the solution set, S .

This solution set can be divided into two subsets, S_1 and S_R , each of them corresponding to specific metabolic conversions, irreversible and reversible respectively. S_1 necessarily includes the fundamental routes \mathbf{f}_1 and \mathbf{f}_2 , and, in fact, is related to the

oxidative/reductive branch of those pathways, implying decarboxylation and carboxylation reactions respectively. On the other hand, the reversible part S_R corresponds to the non-oxidative/non-reductive phase of the routes [implying type (b) reactions]. In addition, it can be remarked that a higher capacity of response of the system to external changes can be obtained with a larger number of fundamental routes involved in the basis. As will be discussed below, this fact allows metabolism to be in tune with the surroundings, and therefore becomes an actual evolutive necessity.

Another classification of S , according to the network relationship with the environment, is worthy of mention. Two fundamental kinds of simplest pathway form the basis of S chosen above. One is formed by those pathways that are able to produce a net conversion from one or several substrates into one or more final products. They are obtained by linear combinations of \mathbf{b}_1 , \mathbf{b}_2 , \mathbf{b}_3 , \mathbf{f}_1 and \mathbf{f}_2 . The other set contains internal pathways with a null net transformation (those that are obtained by linear combinations of \mathbf{b}_4 , \mathbf{b}_5 and \mathbf{b}_6). The first set will be referred to as the boundary set, and the second the internal set. As a direct consequence of the system definition, irreversibility appears only in the boundary reactions, i.e. the fixation or release of CO_2 . However, it must be noted that the global stoichiometry imposes specific couplings with carboxylation and decarboxylation reactions, as is reflected in routes \mathbf{f}_1 and \mathbf{f}_2 . In the same way, reversible pathways can be found in both the boundary and internal sets. It must be remarked that when the system is isolated from the surroundings, i.e. the velocities of the first five reactions are zero (as in routes \mathbf{b}_4 , \mathbf{b}_5 and \mathbf{b}_6), the number of carbons present in the system is conserved (that is the unique structural conservation relationship that appears in the system [11]).

As has been shown, all the stoichiometrically compatible organizations of monosaccharide interconversion can be obtained from a finite number of elementary routes. Moreover, it is possible to classify these solutions according to their character (reversible and irreversible) and the external relationship. But the fundamental question about the correspondence between the enzymic activities and these flux organizations still remains unanswered. The next two sections are devoted to discussing this problem.

The activity set

The ideas developed in the previous section have an immediate translation to the real problems stated above in the study of cell metabolism. After detecting the existence of a particular (in/out)put flux of intermediates, metabolic analysis tries to understand the pathway organization and functioning. A useful strategy is to measure the particular reaction velocities and from that to deduce the enzyme arrangement to form a pathway.

The presence or not of a particular enzyme (activity) can be easily codified by means of the activity vector, \mathbf{a} , whose i th coordinate is 1 if the enzyme has activity (or exists) for reaction i and 0 if this activity is not carried out by it (or it does not exist). Since reactions (1–5) need no enzymes, the activity vector has 12 components corresponding to reactions (6–17) (see Figure 3). The first and second components of this vector mean the activity of enzymes involved in decarboxylation (6) and carboxylation (7) respectively. The following four reactions refer to reactions (8–11) (TKs), and the next three to reactions (12–14) (TAs). The last three represent the catalytic action of ALs [reactions (15–17)]. The set of all the activity vectors is called an activity set and is denoted by A . It is a finite set formed by 2^{12} elements.

The particular feature of the system (enzymic hemireactions

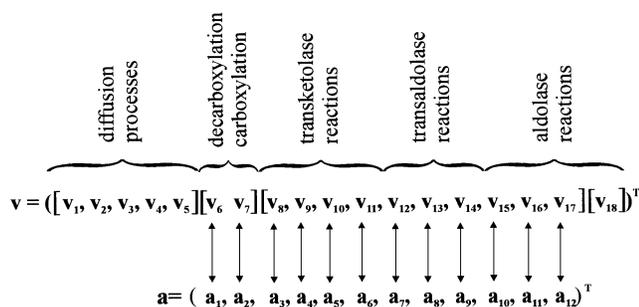


Figure 3 Rate and activity vector correspondence

Since transport processes (v_1, \dots, v_5) need no enzymes, these reactions are not reflected in the activity vectors. The rates of the decarboxylation and carboxylation (v_6 and v_7) are related to the first and second components of the activity vector. TK reaction rates (v_8, \dots, v_{11}) correspond to the next four elements of the activity vector (a_3, a_4, a_5 and a_6). The next three positions in the rate vector (TA reactions) are related to the next three positions in the activity vector (a_7, a_8 and a_9). Finally, the AL reaction velocities (v_{12}, \dots, v_{17}) are related to the last three positions in the activity vector (a_{10}, a_{11} and a_{12}). Notice that reaction (18) is implicit in the other AL reactions, and therefore needs no codification in the activity vector. T means vector transposition.

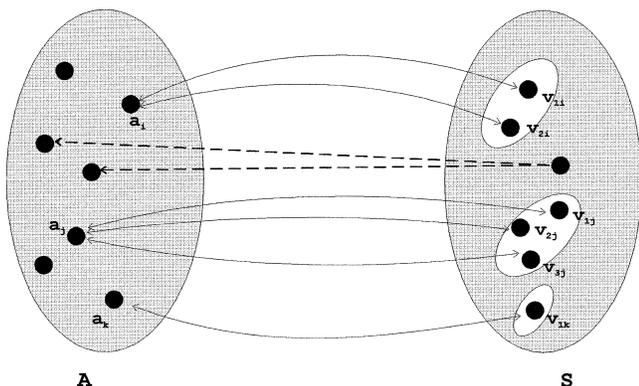


Figure 4 Relationship between the activity set (A) and the solution space (S)

Among the 2^{12} activity vectors, there are some that are connected with no element of S , others that have a one-to-one relationship, and another group that is related to an infinite subset of S . In addition, it can be easily shown that a rate vector cannot be related to more than one activity vector (dashed lines).

that must be coupled to progress, under the mechanisms described in the 'Kinetic model' section) imposes serious restrictions on the elements of A that are stoichiometrically feasible at the steady state. In consequence, only a subset of the activity set is related to flux distributions compatible with the stoichiometry. For instance, it can be easily shown that the activity vector $\mathbf{a} = (0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0)$, which represents the existence of enzymic activity of TK for C_5 , TA for C_6 and AL for C_6 , has no relationship to any flux distribution of S , since enzymic activities cannot be coupled. On the other hand, some activity vectors can be related to several flux distributions in S . That is the case of $\mathbf{a} = (0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1)$ that corresponds to those flux distributions that are linear combinations of the two basic routes \mathbf{b}_2 and \mathbf{b}_5 . In this case, associated with this activity vector are multiple ways of distributing the flow through the pathway. Moreover, it is easy to prove that any flux distributions of S are related to only one activity vector. This correspondence is schematically drawn in Figure 4.

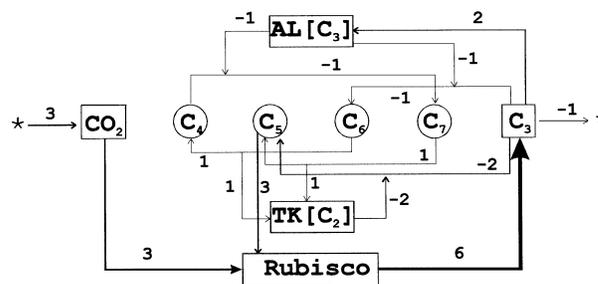


Figure 5 Reaction network scheme represented by the flux distribution $\mathbf{u} = ([3, -1, 0, 0, 0] [0, 3] [-2, 1, 1, 0, 0, 0, 0, -1, -1, 0] [2])^T$ (where T is vector transposition)

The direction of the flow through the route is indicated by arrows. The rate flux through each reaction is codified by the width of lines, and the particular values are given by the numbers over the lines. Notice that this flux distribution fits the global stoichiometry, i.e. $3\text{CO}_2 \rightarrow \text{C}_3$. This network organization represents the well-known Calvin cycle. Stars indicate the system environment.

Following this reasoning, the goal now is to find the correspondence between velocity distributions and activity vectors, i.e. between S and A . Finding this relationship provides a way to know the multiple ways of constructing the pathway compatible with the stoichiometry (different enzyme mechanism and location within the path). Thus this study allows discussion of the possibilities of establishing an enzymic network. In order to clarify well all these conclusions, the next section is devoted to analysing a well-known pathway: the production of C_3 sugars by CO_2 fixation involving monosaccharide interconversion, i.e. discarding both the Arnon cycle and the direct synthesis of acetate, as occurs in some bacteria.

Producing C_3 sugars by CO_2 fixation

Previous sections have been focused on studying in a general sense the flux organization of monosaccharide interconversion. No additional assumptions about either the kinetic features or the relationships of the system with the environment have been made there. Therefore the conclusions derived above are totally valid for any particular situation obtained after imposing any hypothesis on the general model.

Among all the rate distributions with physical meaning (S) there are particular solutions that correspond to situations of monosaccharide metabolism in present-day cells. That is the case for the fixation of CO_2 to acquire monosaccharides for plants. This conversion cycle is one of the main ways of providing energy and matter for living beings. Within photosynthesis this interconversion is linked to the pentose phosphate cycle, which supplies ribulose 5-phosphate needed to feed the pathway. As was deduced in the previous section, this transformation can, in principle, be performed by different network organizations (one of them being the Calvin cycle). In this section, we analyse the possible structures that are in agreement with the global stoichiometry of this pathway, i.e. $3\text{CO}_2 \rightarrow \text{C}_3$.

In a similar way as stated in the previous section, it is not difficult to find the basic routes from which all the possible flux distributions can be obtained. It can be proven (see the Appendix) that the general solutions to this problem are linear combinations of the basic pathway \mathbf{u} , necessary to get a net conversion (see Figure 5), and three additional routes \mathbf{b}_4 , \mathbf{b}_5 and \mathbf{b}_6 (Figure 2). As can be seen, this route \mathbf{u} is a particular combination of fundamental routes of the general problem, i.e. $\mathbf{u} =$

number of intermediates (nodes) and enzymes (connections) involved in the network. On the other, although the number of nodes and connections is not very high, here complexity appears because both intermediates and enzymes are shared and used more than once in the network. The enzymic unspecificity is not exclusive of this interconversion, although it can be considered as paradigmatic. This feature was already recognized in the first studies of the pentose phosphate cycle: whereas in the classic design of the non-oxidative phase of the pentose phosphate cycle, AL can act only on 6-carbon sugars, in the non-reductive phase of the Calvin cycle this activity is extended to 7-carbon sugars, and in the L-type described by Williams [6,17], AL catalyses also those reactions involving 8-carbon sugars. In addition, according to the principle of activity [7], whenever an enzyme has activity on a certain substrate it must act on it, increasing even more the complexity of the problem. Thus in a certain sense, the L-cycle must include those reactions involved in the non-reductive phase of the Calvin cycle (though working in the opposite direction), since the enzymes have the corresponding activities of that pathway.

The mechanism of enzymic reactions brings another important characteristic to the system. Since all these processes occur by transfer of carbon fragments among sugars by coupling of two hemireactions, a hemireaction can be coupled with more than one hemireaction. For instance, in the L-type the hemireactions (8–11) are coupled with each other and working simultaneously. Then, in a strict sense this pathway cannot be represented as a sequence of steps. A more appropriate representation is given by a metabolic network, as shown in Figure 1.

With every flux distribution of a pathway can be associated, both theoretically and experimentally, a unique activity vector. However, a more interesting question is to find those rate vectors stoichiometrically compatible at the steady state with a given activity vector. As has been derived in this work, the activity set can be divided into three subsets. The first one is formed by those activity vectors that do not allow stoichiometrically correct rate distributions. As argued above, enzymic reactions progress due to the coupling among hemireactions. So, it can be the case that the absence of a particular enzyme prevents the system from self-organizing stoichiometrically. A second subset is formed by those activity sets that are related to a unique flux distribution. For these activity vectors the flux distribution through the pathway at the steady state is independent of the system parameters. Finally, a third set contains those activity vectors related to more than one rate distribution. The members of this subset are associated with flux distributions through the network that depend on both the external constraints and the system parameters.

The L-type mentioned in the 'Producing C₃ sugars by CO₂ fixation' section belongs to the third subset. This pathway implies that TK can act on C₅, C₆, C₇ and C₈ sugars and AL on C₆, C₇ and C₈ sugars. In that section, it was shown that the corresponding activity vector $\mathbf{a} = (0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1)$ is related to all the rate vectors obtained as linear combinations of the Calvin route and the internal cycle \mathbf{b}_5 . The value of the flux through \mathbf{b}_5 depends on the external conditions under which the pathway is working. Figure 7 shows the results obtained by numerical integration of the kinetic equations of a system with a constant input flux of CO₂ and an output of C₃. As can be seen, depending on the value of this stationary flow, different flux distributions appear at the steady state, i.e. different flux distributions are stoichiometrically compatible. From an experimental point of view this fact has important consequences: a researcher could detect different routes depending on the regime in which the cell is working. In other words, any kind of external regulation

of the net interconversion of monosaccharides in the cell shouldn't forget this dynamic behaviour. A deeper study of this effect, together with the dynamic dependences with the rest of the system parameters, will be reported in a forthcoming paper.

From an evolutionary viewpoint, this structural degeneration has an important implication. In a previous work, the flux of a metabolic pathway was taken as the main target to be optimized during evolution [18]. It was demonstrated that in linear pathways this magnitude depends on the number of steps of the metabolic route: the shorter the route the higher the flux through it becomes. In addition, Meléndez-Hevia [7] described both the F pentose phosphate pathway and the Calvin cycle as the routes with the lowest number of steps to get the corresponding interconversion, and therefore they were considered as the simplest ways to perform such a purpose. However, the stoichiometrically compatible flux distributions of the monosaccharide interconversion cannot be related to sequential pathways. Therefore, structural simplicity is better associated, for a fixed number of metabolites, with the number of connections of the reaction network, i.e. simplest pathways have the lowest number of connections. Formally, this simplicity is reflected in the number of null entries in the flux distributions of the network. In these cases, the kinetic features of the route such as the flux distributions, the transition time, etc. must be related to other system characteristics. A detailed study of these kinetic properties is currently under study and will be reported shortly.

Finally, notice that those cases in which the flux distribution is degenerated with respect to the enzymic activity can offer a broader versatility to the metabolic network. To respond adequately to changes in the environment must be an important challenge for biological systems. From this view, those activity vectors that allow the system to self-organize stoichiometrically according to external fluctuations will have an evolutionary advantage. Then, this versatility should be considered as a fitness function to be optimized in the design of metabolic pathways.

This work has been supported in part by grant BIQ96-0895 from Comisión Interministerial de Ciencia y Tecnología (Spain) and PB94-0593 from Dirección General de Investigación Científica y Técnica (Spain). I.S.-V. is a recipient of a fellowship from Universidad Complutense de Madrid (Spain). We also thank Dr. David Fell for his fruitful suggestions and help with the English.

REFERENCES

- 1 Decker, P., Scheweier, H. and Pohlmann, R. (1982) *J. Chromatogr.* **225**, 281–291
- 2 Mason, S. F. (1992) in *Chemical Evolution*, pp. 231–259, Oxford University Press, Oxford
- 3 Maden, B. E. H. (1995) *Trends Biol. Sci.* **20**, 337–341
- 4 Miller, S. L. (1992) in *Major Events in the History of Life* (Schopf, J. W., ed.), pp. 1–28, Jones & Bartlett, Boston
- 5 Wood, T. (1985) *The pentose phosphate pathway*, Academic Press, Orlando
- 6 Williams, J. F. (1980) *Trends Biol. Sci.* **5**, 315–320
- 7 Meléndez-Hevia, E. (1990) *Biomed. Biochim. Acta* **49**, 903–916
- 8 ap Rees, T. (1985) in *Encyclopedia of Plant Physiology* (Douce, R. and Day, D. A., eds.), pp. 391–417, Springer-Verlag, Berlin
- 9 Dennin, D. T. and Miernyk, J. A. (1982) *Annu. Rev. Plant Physiol.* **33**, 27–50
- 10 Fell, D. A. and Small, J. R. (1986) *Biochem. J.* **238**, 781–786
- 11 Reder, C. (1988) *J. Theor. Biol.* **135**, 175–201
- 12 Clarke, B. L. (1992) *J. Chem. Phys.* **97**, 4066–4071
- 13 Masuda, M. (1993) *J. Chem. Phys.* **99**, 4873–4874
- 14 Schuster, S. and Hilgetag, C. (1994) *J. Biol. Syst.* **2**, 165–182
- 15 Meléndez-Hevia, E., Waddell, T. G. and Cascante, M. (1996) *J. Mol. Evol.* **43**, 293–303
- 16 Stryer, L. (1995) *Biochemistry*, 4th edn., pp. 563–565, W. H. Freeman, New York
- 17 Williams, J. F., Arora, K. K. and Longenecker, J. P. (1987) *Int. J. Biochem.* **19**, 749–817
- 18 Meléndez-Hevia, E., Waddell, T. G. and Montero, F. (1994) *J. Theor. Biol.* **152**, 1–22

APPENDIX

Mathematical formulation

All the results presented in the present paper can be rigorously proven using concepts of linear algebra. In this section, the main steps of these derivations are described.

Stoichiometric matrix

Let us denote by $x_1 = [\text{CO}_2]$ and $x_i = [\text{C}_{1+i}]$ ($i = 2, 3, \dots, 7$) the concentrations of CO_2 and $(i+1)$ -carbon sugars respectively, and by $x_8 = [\text{TK}[\text{C}_2]]$, $x_9 = [\text{TA}[\text{C}_3]]$ and $x_{10} = [\text{AL}[\text{C}_3]]$ the concentrations of complex intermediates involved in the reactions. The time evolution of these variables for the full design (assuming that enzymes possess all the possible activities) is governed by the following dynamic system:

$$\frac{d}{dt} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \end{pmatrix} = \mathbf{N} \times \begin{pmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_{16} \\ v_{17} \\ v_{18} \end{pmatrix} \quad (\text{A1})$$

where the reaction rates v_i correspond to the general reaction scheme stated in the ‘Stoichiometric analysis’ section. \mathbf{N} is a 10×18 matrix, the so-called stoichiometric matrix. Its rows are directly related to the metabolites, whereas the columns represent the reactions in which each of them is taking part. The expression of this matrix can be easily derived from the reaction scheme (1–18) (as usual a velocity is positive if the product is placed in the right-hand side of the reaction):

$$\mathbf{N} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & -1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & -1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \quad (\text{A2})$$

A more compact formulation using matrix notation can be given straightforwardly; let \mathbf{x} and \mathbf{v} denote the concentration and velocity vectors respectively. Thus equation (A1) reads:

$$\frac{d\mathbf{x}}{dt} = \mathbf{N}\mathbf{v}(\mathbf{x}) \quad (\text{A3})$$

It is well known that under a stationary regime the velocities of the reactions must reach a constant value. The steady-state velocity vector \mathbf{v} can be determined by solving the homogeneous linear system:

$$\mathbf{N}\mathbf{v}(\mathbf{x}) = 0 \quad (\text{A4})$$

Solution space

The derivation of all the stationary solutions in terms of the metabolic concentrations, i.e. to solve the system (A4), is a very hard task. However, given a particular setup, all possible rearrangements of reactions (1–18) must be found among the set of rate vectors that are solutions of eqn. (A4). This solution set is a linear space and is usually referred as the kernel of \mathbf{N} , $\ker(\mathbf{N})$. Notice that, since the rank of \mathbf{N} is less than the number of reactions, $\ker(\mathbf{N})$ is always non-trivial. Then, from a mathematical viewpoint all the rate vectors that are solutions of eqn. (A4) could be obtained as linear combinations of a finite number of vectors, the basis of $\ker(\mathbf{N})$. Therefore, formally the problem is solved when a basis of $\ker(\mathbf{N})$ is found. In this sense, it can be easily proven that the minimum number of vectors needed to generate this space is eight, i.e. the dimension of $\ker(\mathbf{N})$ is eight.

Owing to irreversibility of reactions (6) and (7), only a subset, S , of the whole solution space, $\ker(\mathbf{N})$, has physical meaning (all those negative linear combinations that involve non-null values of the 6th and 7th coordinates of \mathbf{v} must be discarded according to the physical meaning of the model). The immediate consequence of this fact is that set S loses the vectorial structure of $\ker(\mathbf{N})$, although it conserves those properties derived from its convex character. Even in this case, it is possible to express any vector of S as a linear combination of vector generators, i.e.:

$$S = \left\{ \mathbf{v} \mid \mathbf{v} = \sum_{k=1}^m \eta_k \mathbf{f}_k + \sum_{j=1}^r \lambda_j \mathbf{b}_j, \eta_k \geq 0, k = 1, \dots, m; \lambda_j \in \mathfrak{R}, j = 1, \dots, r \right\} \quad (\text{A5})$$

with $m+r=8$. Here, \mathbf{f}_k denote independent rate vectors that contain non-null irreversible reactions, i.e. that have $v_6 \geq 0$ or $v_7 \geq 0$ (not both simultaneously equal to zero). On the other hand, \mathbf{b}_j are independent vectors (and independent of \mathbf{f}_1 and \mathbf{f}_2) that do not contain any irreversible coordinates, i.e. $v_6 = v_7 = 0$.

It must be remarked that any generator set of S contains a subset of the irreversible vectors, and therefore the whole set S is essentially irreversible.

However, S contains a subset $S_{\mathbf{R}}$, expanded by vectors \mathbf{b}_j ($j = 1, 2, \dots, r$) that forms a linear subspace of $\ker(\mathbf{N})$. It can be proven straightforwardly that the dimension of $S_{\mathbf{R}}$ is six. It is worth recalling that each \mathbf{b}_j is the contribution of the forward \mathbf{b}_{j+} and the reverse \mathbf{b}_{j-} rate vectors, i.e. $\mathbf{b}_j = \mathbf{b}_{j+} - \mathbf{b}_{j-}$. Obviously, the vectorial character of $S_{\mathbf{R}}$ is lost when considering these partial velocities, since only non-negative linear combinations of those vectors are allowed.

Therefore, to expand the rest of S , two more independent vectors, \mathbf{f}_1 and \mathbf{f}_2 , are needed [that is, $m = 2$ in formula (A5)]. These two vectors generate a convex set $S_1 = (\mathbf{v} \in S | v = \sum_j \eta_j \mathbf{f}_j, \eta_j \geq 0, j = 1, 2)$. Moreover, $S_1 \cap S_R = \emptyset$. Then, S can be considered as the sum (direct) of the two sets S_R and S_1 , i.e. $S = S_1 \oplus S_R$.

A simplest generator set

The search for a maximal generator set (a basis as named when dealing with vector spaces) is the main goal that will guide the next derivations. Among all criteria for choosing a complete set of eight independent vectors, that of simplicity seems to be the most adequate to express the structural features of the system. Mathematically, simplicity is translated into the maximum number of null components that a rate vector can have. Therefore, looking for simplicity requires a full knowledge of the vector structure of S , i.e. knowing the distribution of vectors that contain a particular number of null entries.

Let Z be the number of zeros of a reaction rate vector. To get a simple generator set of S , for each Z we will take the maximum number of independent vectors, G_Z , from the upper limit of zeros down. After a cumbersome combinatorial search, it can be shown that if S is the physical meaningful set, then if:

$$Z = 17, 16, 15, 13 \text{ then } G_Z = \emptyset$$

$$Z = 14 \text{ then } G_{14} = 6$$

$$Z = 12 \text{ then } G_{12} = 1$$

Similar results can be derived for combinations with a smaller number of zeros. Since the dimension of S is eight, to complete the generator set of S an additional vector is needed. This can be chosen from among those rate vectors with $Z = 11$, i.e. 11 null components, that are linearly independent of the other seven vectors already selected. This procedure can be summarized by giving explicitly the following basis for S_R , $B = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4, \mathbf{b}_5, \mathbf{b}_6)$, with (as usual, \top means vector transposition):

$$\mathbf{b}_1 = ([0, -1, 1, 1, -1][0, 0][1, -1, 0, 0, 0, 0, 0, 0][0])^\top$$

$$\mathbf{b}_2 = ([0, 0, -1, 2, -1][0, 0][1, 0, -1, 0, -1, 1, 0, 0, 0][0])^\top$$

$$\mathbf{b}_3 = ([0, -2, 0, 0, 1][0, 0][0, 0, 0, 0, 0, 0, 1, 0, 0][-1])^\top$$

$$\mathbf{b}_4 = ([0, 0, 0, 0, 0][0, 0][1, 0, 0, -1, -1, 0, 1, 0, 0][0])^\top$$

$$\mathbf{b}_5 = ([0, 0, 0, 0, 0][0, 0][1, 0, 0, -1, 0, 0, 0, -1, 0, 1][0])^\top$$

$$\mathbf{b}_6 = ([0, 0, 0, 0, 0][0, 0][0, 0, 0, 0, 1, -1, 0, -1, 1, 0][0])^\top$$

Similarly, the irreversible set of S_1 can be generated by the following linearly independent vector set, $F = \{\mathbf{f}_1, \mathbf{f}_2\}$, necessarily differing at positions 6 and 7, with:

$$\mathbf{f}_1 = ([-1, 0, 0, -1, 1][1, 0][0, 0, 0, 0, 0, 0, 0, 0, 0][0])^\top$$

$$\mathbf{f}_2 = ([1, -2, 0, 1, 0][0, 1][0, 0, 0, 0, 0, 0, 0, 0, 0][0])^\top$$

In order to facilitate the reading of the vectors, they appear partitioned in three groups of components, each of them related to one of the set of reactions involved in the system: diffusion processes (1–5), carboxylation and decarboxylation reactions (6) and (7), reactions of carbon transfer (8–17) and the specific AL reaction (18) (see Figure 3).

Therefore, the set:

$$G = (\mathbf{f}_1, \mathbf{f}_2, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4, \mathbf{b}_5, \mathbf{b}_6)$$

is one of the simplest multiple sets that generates S . Figure 2 shows each of the fundamental vectors of G .

A simple example: $\text{CO}_2 \rightarrow \text{C}_3$

As a particular application of these results, let us analyse the production of trioses from CO_2 . By definition, the only input and output are CO_2 and C_3 , i.e. $v_1 > 0$, $v_2 < 0$ and $v_3 = v_4 = v_5 = 0$. Then, it can be easily proven that the rate vectors compatible with the stoichiometry must belong to a subset of S , U (containing no terms in positions 3, 4 and 5), expanded by the simplest generator set:

$$G_U = (\mathbf{u}, \mathbf{b}_4, \mathbf{b}_5, \mathbf{b}_6) \quad (\text{A6})$$

where $\mathbf{u} = 3\mathbf{f}_2 - \mathbf{b}_1 - \mathbf{b}_2 - 2\mathbf{b}_3 - \mathbf{b}_6$, that is,

$$\mathbf{u} = ([3, -1, 0, 0, 0][0, 3][-2, 1, 1, 0, 0, 0, 0, -1, -1, 0][2])^\top$$

Figure 5 shows schematically the flux distribution related to \mathbf{u} (the meanings of \mathbf{b}_4 , \mathbf{b}_5 and \mathbf{b}_6 are given in Figure 2). Notice that the vectors of U are linear combinations of a boundary vector (\mathbf{u}) and three internal vectors ($\mathbf{b}_4, \mathbf{b}_5, \mathbf{b}_6$) that are superimposed on the main stream of the pathway.