

The underlying pathway structure of biochemical reaction networks

(null space/stoichiometric matrix/bioinformatics/metabolic pathways/genome annotation)

CHRISTOPHE H. SCHILLING AND BERNHARD O. PALSSON*

Department of Bioengineering, University of California, San Diego La Jolla, CA 92093-0412

Communicated by Edwin N. Lightfoot Jr., University of Wisconsin-Madison, Madison, WI, January 23, 1998 (received for review November 25, 1997)

ABSTRACT Bioinformatics is yielding extensive, and in some cases complete, genetic and biochemical information about individual cell types and cellular processes, providing the composition of living cells and the molecular structure of its components. These components together perform integrated cellular functions that now need to be analyzed. In particular, the functional definition of biochemical pathways and their role in the context of the whole cell is lacking. In this study, we show how the mass balance constraints that govern the function of biochemical reaction networks lead to the translation of this problem into the realm of linear algebra. The functional capabilities of biochemical reaction networks, and thus the choices that cells can make, are reflected in the null space of their stoichiometric matrix. The null space is spanned by a finite number of basis vectors. We present an algorithm for the synthesis of a set of basis vectors for spanning the null space of the stoichiometric matrix, in which these basis vectors represent the underlying biochemical pathways that are fundamental to the corresponding biochemical reaction network. In other words, all possible flux distributions achievable by a defined set of biochemical reactions are represented by a linear combination of these basis pathways. These basis pathways thus represent the underlying pathway structure of the defined biochemical reaction network. This development is significant from a fundamental and conceptual standpoint because it yields a holistic definition of biochemical pathways in contrast to definitions that have arisen from the historical development of our knowledge about biochemical processes. Additionally, this new conceptual framework will be important in defining, characterizing, and studying biochemical pathways from the rapidly growing information on cellular function.

Biology is going through a period of fundamental change. The entire genome sequence for an increasing number of organisms is becoming available. Currently, the full DNA sequence for 14 unicellular organisms are at hand, and it is expected that the full DNA sequences will become available for well known multicellular organisms within a few years (1). The identification of ORFs and their assignments are proceeding at a rapid pace, and 50 to 80% of the ORFs have been assigned in the fully sequenced microbial genomes (2). Once complete, these efforts will result in a complete “spare part catalogue” of the components found in a multitude of living cells.

Given the long history of metabolic research, the assignment of metabolic genes to ORFs has been particularly successful. About 91% of the known metabolic enzymes found in *Escherichia coli* K-12 had ORF assignments in the initial publication of its DNA sequence (3). Furthermore, a large fraction of the ORFs assigned in the first sequenced and annotated genomes shows that metabolic genes make up a large fraction of the genes found in the microbial genotype (4). In addition, metabolism has general features that are found in most organisms. Comparison of seven

sequenced genomes from five major phylogenetic lineages has led to the definition of 720 clusters of orthologous groups of genes based on patterns of sequence similarities (5). These clusters have been grouped into 15 functional groups, nine of which are metabolic in nature. With all of this genomic information now present, there is a demand for creative approaches to deal with these data sets (6, 7). What is now needed is a systemic definition of the underlying metabolic pathways that allow for the analysis of the overall function of individual metabolic genotypes.

Flux-balance analysis can be used to analyze, interpret, and predict the capabilities and function of metabolic genotypes (8). This analysis method is based on mass balances on all the metabolites found in a given metabolic system (9–11). The analysis relies solely on the stoichiometry of the metabolic reactions present and a handful of strain-specific parameters. This analysis method has been shown to be able to predict quantitatively the behavior of *E. coli* under a variety of growth conditions (12, 13).

The metabolic capabilities of the defined metabolic genotype correspond to the null space of the stoichiometric matrix. Particular solutions found within this space represent expressed metabolic phenotypes (10). The null space can be spanned by a set of basis vectors (14, 15), a combination of which can represent all the metabolic phenotypes found within the metabolic genotype. Here, we show how these basis vectors, originating from the fundamentals of linear algebra, are related to metabolic pathways intrinsic to stoichiometric matrices. Thus, all the metabolic phenotypes can be represented by a combination of these systemically defined metabolic pathways. This development not only redefines the notion of a metabolic pathway but also forms a fundamental basis for the development of analysis methods in microbial genomics.

Translating Biochemical Networks into the Realm of Linear Algebra

The interconnectivity of metabolites within a network of biochemical reactions is given by reaction equations defining the stoichiometric conversion of substrates into products for every reaction. Enzymatic reactions as well as the transport of metabolites across system boundaries constitute various fluxes, which serve to dissipate and generate metabolites. In following the law of conservation of mass, material balances describing the activity of a particular reactant through each reaction can be written in the form of homogeneous linear equations, that in matrix notation is:

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{0} \quad [1]$$

The stoichiometric matrix \mathbf{S} is an $m \times n$ matrix where m corresponds to the number of metabolites and n is the number of reactions or fluxes taking place within the network. The S_{mn} element of the stoichiometric matrix corresponds to the stoichiometric coefficient of the reactant m in the reaction denoted by n , and v_n is the flux through this metabolic reaction. The vector

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/954193-6\$2.00/0
PNAS is available online at <http://www.pnas.org>.

Abbreviation: 2,3-DPG, 2,3-bisphosphoglycerate.

*To whom reprint requests should be addressed; e-mail: palsson@ucsd.edu.

\mathbf{v} then refers to the relative activity of each flux and is $1 \times n$ (contained in \mathbb{R}^n). Through the matrix \mathbf{S} , the biochemical system is cast into a mathematical context, permitting the application of the methods of linear algebra. The stoichiometric matrix remains constant at all times because only the values comprising \mathbf{v} change to reflect different flux distribution patterns. Furthermore, the metabolic genotype of an organism directly leads to the definition of \mathbf{S} (8).

We are seeking to define the architecture of the network represented by \mathbf{S} in functional biochemical terms. Namely, we are interested in the inherent structure and connectivity of the metabolites within a metabolic network. Mathematically, this question is concerned with the vector space containing all of the possible solutions of the system of linear equations represented by Eq. 1. This vector space is termed the null space of the matrix \mathbf{S} , represented as NulS . Fundamentally, we are interested in determining the characteristics of this space and interpreting it from an overall metabolic perspective.

The dimension of the null space (the nullity of \mathbf{S}) depends on the number of free variables in the original set of linear equations, which is referred to as the rank of \mathbf{S} , given by the relation termed the Rank Theorem:

$$\dim \text{NulS} + \text{rankS} = n \quad [2]$$

The null space contains all of the solutions and hence vectors (\mathbf{v}) that satisfy Eq. 1. The characteristics of the solution vector \mathbf{v} is of great interest because it describes the relative distribution of fluxes throughout the entire metabolic system. A particular solution of interest to Eq. 1 can be determined either through the use of linear programming, given a suitable objective function (10), or by reducing the number of unknown fluxes to allow the system to be exactly determined (16–18). The first approach is akin to selecting a point within the null space that optimizes an objective (such as cellular growth), whereas the second is an attempt to reduce the dimensions of the null space to the zero subspace (a point). Both of these approaches aim to find a single point or solution contained within the null space. However, to address more general issues, such as metabolic capacity and regulation in regards to the entire network, a thorough analysis of the entire null space is in order.

To explicitly define the null space, it is necessary to generate a set of vectors that can be used to span the vector space (i.e., a spanning set). The most “efficient” way to span a vector space is through the use of a minimum number of linearly independent vectors that together form a basis. The minimum number of vectors necessary to form a basis is equal to the dimensions of the vector space ($\dim \text{NulS}$). As in any problem involving vector spaces in any dimension, the key is to select a basis that clearly describes all of the points contained within the space in a meaningful way.[‡] With this in mind, we will attempt to develop a rational approach for the selection of basis vectors, so as to span the null space in a manner that is theoretically and biochemically meaningful.

System Descriptions and Conventions

For any given biochemical reaction network, there are reactions involving the interconversion of chemical species along with transport processes, which serve to replenish or drain the relative amounts of certain metabolites. A system boundary can be drawn around all of the chemical reactions occurring within a system.

[‡]Whereas large dimensional spaces (more than three dimensions) become abstract, recall that in three dimension we commonly use the x , y , and z coordinate system as our basis; yet, for certain problems, this system becomes an inefficient set of basis vectors, and therefore, we transform the basis. Examples of this transformation would include simple basis rotations and more complex transformations such as spherical or cylindrical bases (e.g., see ref. 19) for solutions of various transport phenomena problems. In each case, the basis vectors all span the entire three dimensional space.

These reactions and their relative activity will be referred to as internal fluxes whose relative activity is indicated by v_1 – v_i where i is the number of internal fluxes. Only those fluxes corresponding to the transport of a metabolite are permitted to cross the system boundary. These fluxes constitute sources and sinks on the system and will be referred to as exchange fluxes denoted as b_1 – b_j where j is the total number of exchange fluxes. All of the exchange fluxes are defined as positive if the activity is draining a metabolite or leaving the system.

An example of a chemical reaction scheme is illustrated in Fig. 1A. The system contains five metabolites (A–E) undergoing the reactions indicated by each line with arrows indicating the direction of the reaction. Note that the reversible reaction occurring between metabolite B and D is represented as two opposing reactions. All of the metabolites, with the exception of B, contain either a source or a sink thus creating four exchange fluxes. The biochemical reaction network in Fig. 1A is then converted to conform to the conventions described above and is represented in Fig. 1B where all internal fluxes and exchange fluxes are labeled along with the system boundaries.

Starting from a series of material balances around each metabolite a stoichiometric matrix \mathbf{S} can be generated for the above reaction scheme which mathematically represents all 11 fluxes and the metabolites involved in each flux (Fig. 2). In this example, the stoichiometric matrix is of dimension 5×11 , here structured so that the first seven columns represent the internal fluxes and the remaining four constitute the exchange fluxes. In constructing the stoichiometric matrix in this way, the vector \mathbf{v} is composed first of entries v_1 – v_7 followed by b_1 – b_4 . With the structure and connectivity of the network now translated into mathematical form, we can begin to develop our approach to spanning the null space.

Spanning the Null Space

Beginning with Eq. 2, we determined the dimensions of the null space. In the example given, the rank of \mathbf{S} is five and thus the dimension of the null space is six. Therefore, to explicitly define the null space of this system, we need a set of six basis vectors $\mathcal{B} = \{\mathbf{b}_1$ – $\mathbf{b}_6\}$ in \mathbb{R}^{11} that span NulS . A set of basis vectors can be determined by expressing the general solution to Eq. 1 in para-

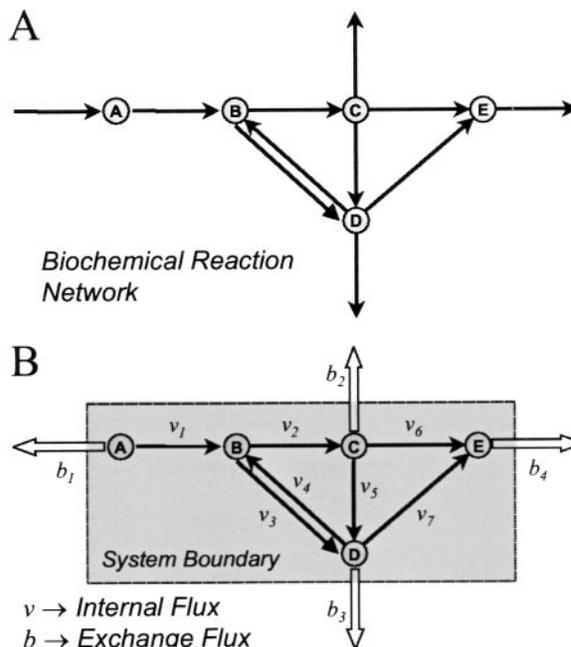


FIG. 1. (A) Chemical reaction scheme consisting of five metabolites and 11 fluxes. (B) Conversion of the reaction scheme using conventions described within the text. A system boundary is drawn with internal fluxes (v_1 – v_7) and external fluxes (b_1 – b_4) labeled.

Balance Equations:

- A: $-v_1 - b_1 = 0$
- B: $v_1 + v_4 - v_2 - v_3 = 0$
- C: $v_2 - v_5 - v_6 - b_2 = 0$
- D: $v_3 + v_5 - v_4 - v_7 - b_3 = 0$
- E: $v_6 + v_7 - b_4 = 0$



Matrix Notation
 $S \cdot v = 0$

Stoichiometric Matrix

$$S = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

$\underbrace{\hspace{150px}}_{\text{Internal Fluxes}}$
 $\underbrace{\hspace{150px}}_{\text{Exchange Fluxes}}$

FIG. 2. Representation of reaction balance equations in matrix notation. From the five balance equations, a 5×11 stoichiometric matrix is generated.

metric form, as a linear combination of the free variables in the balance equations (15). The vectors indicating the weights of the free variables then become a set of basis vectors that span the null space. Using this procedure a set of six basis vectors are constructed (Fig. 3). The basis vectors of \mathcal{B} are all linearly independent and theoretically feasible as they satisfy Eq. 1. However, what do these vectors tell us from a biochemical viewpoint?

The first seven entries of each vector correspond to the internal fluxes (v_i) whereas the last four indicate the activity of the exchange fluxes (b_j). Through inspection of each of the components of the basis vectors, we see that each vector traces out a theoretical pathway through the system in which the components of each vector describe the relative flux distributions of each of the reactions in the pathway. Notice that, in two of the vectors (\mathbf{b}_4 and \mathbf{b}_5), there are negative values representing the activity of certain internal fluxes. These vectors represent pathways that are biochemically impossible and thus are of little use in interpreting the system's biochemical structure and associated physiology. A biochemically meaningful basis is needed.

Spanning the Null Space with Biochemical Pathways

Because all of the vectors in the basis \mathcal{B} are linearly independent, these vectors can be used to create a different set of six basis vectors, which also form a spanning set of the null space. To guide the construction of these basis vectors, we impose two constraints on both the internal and exchange fluxes:

1. We constrain all internal fluxes to be positive;

$$v_i \geq 0 \text{ (all internal fluxes)} \quad [3]$$

creating a new spanning set which is both theoretically and biochemically feasible. This requirement is critical to ensure that

the basis vectors exhibit biochemical relevance to the analysis of the reaction network.

2. We selectively impose further restrictions on the values of the exchange fluxes,

$$\begin{aligned} b_j &\geq 0 \text{ (exit only)} \\ b_j &\leq 0 \text{ (enter only)}. \end{aligned} \quad [4]$$

Therefore, the value of the exchange fluxes indicates whether or not a particular metabolite is being transported into the system or being siphoned off for other purposes in each pathway. If a reactant is known to enter the system and does not exit, then we may wish to impose the constraint that the value for that particular flux must be negative.

Under these constraints, the basis \mathcal{B} is transformed into a new basis \mathcal{P} defined by the equations relating the two different bases shown in Fig. 3. The basis \mathcal{P} will be composed of six linearly independent vectors because it is a linear combination of the vectors of basis \mathcal{B} .

The transformation defined results in basis vectors that represent biochemical pathways. For the example system, the new basis is illustrated in Fig. 4. Each basis vector is depicted as a pathway, and the directions of the exchange fluxes are indicated. By virtue of the way that the system was constructed, the overall reaction balance equation for the pathways is given by the values of the exchange fluxes with negative values indicating substrates and positive values indicating products in stoichiometric ratios indicated by their respective values.

Note that by the unique representation theorem of coordinate systems, any vector (v) in NulS can be written as a unique combination of these six basis vectors, thus providing a one-to-one mapping of points in the null space. Therefore, any flux

\mathcal{B} Basis:

Theoretically Feasible

$$\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_6\} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline -1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Basis Transformation



\mathcal{P} Basis:

Theoretically & Biochemically Feasible

$$\mathcal{P} = \left. \begin{aligned} \mathbf{p}_1 &= \mathbf{b}_1 \\ \mathbf{p}_2 &= \mathbf{b}_2 \\ \mathbf{p}_3 &= \mathbf{b}_3 \\ \mathbf{p}_4 &= \mathbf{b}_4 + \mathbf{b}_1 \\ \mathbf{p}_5 &= \mathbf{b}_5 + \mathbf{b}_2 \\ \mathbf{p}_6 &= \mathbf{b}_5 + \mathbf{b}_6 \end{aligned} \right\} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline -1 & -1 & -1 & -1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

FIG. 3. Transformation of the theoretically feasible basis \mathcal{B} (spanning the null space of S) to a set of basis vectors \mathcal{P} (also spanning the null space of S), which are both theoretically and biochemically feasible. Note that all internal fluxes (above the dashed line) are positive in the basis \mathcal{P} .

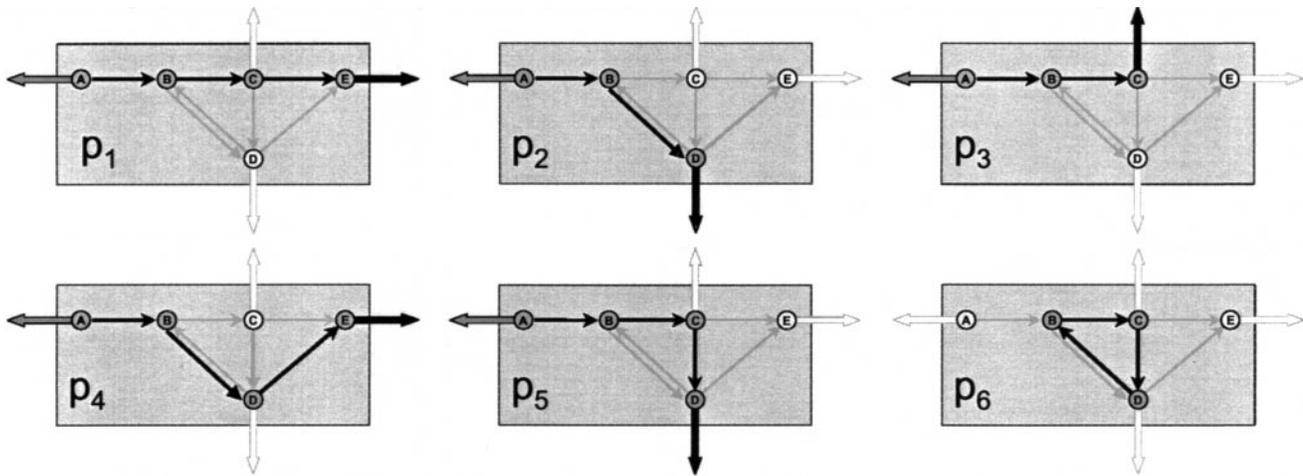


FIG. 4. Fundamental pathways traced out by each of the six linearly independent basis vectors in \mathcal{P} . Exchange flux activity that enters the system (i.e., negative) is indicated by the gray filled arrows, whereas activity that exits the system (i.e., positive) is indicated by the black filled arrows. Active internal fluxes are indicated by the black arrows.

distribution can be viewed as a linear combination of the pathways traced by the basis vectors in \mathcal{P} . If the basis vectors represent particular metabolic functions, any experimentally determined or computed solution can be represented as a linear combination of such metabolic functions. Any changes from this solution can then directly be related to the underlying systemic metabolic functions. These basis vectors thus represent a metabolically meaningful pathway structure of the biochemical reaction network under consideration. In the following application, we span the null space for a network of reactions involved in human erythrocyte metabolism demonstrating this approach for the analysis of biochemical networks.

Application to Integrated Metabolic Functions: The Human Red Blood Cell

The mature red blood cell emerges from the bone marrow enucleated and stripped of a large portion of the metabolic machinery available to most cells. These cells are incapable of protein or lipid synthesis and oxidative phosphorylation, as well as tricarboxylic acid cycling (20). The focus of the red blood cell is on gas transport and exchange. Beyond this function, the cell requires energy to maintain the plasticity of its membrane, prevent the accumulation of methemoglobin, protect its hemoglobin from oxidative denaturation, synthesize glutathione, acylate CoA, and replenish its adenine nucleotide pool using salvage pathways. The metabolic resources used in these functions are found in the high energy phosphate of ATP, or as a reducing compound in the form of reduced glutathione and the pyridine cofactors, NADH, and NADPH. Glucose is the primary carbon source used to derive the energy necessary to synthesize the energy carriers mentioned above. Through anaerobic glycolysis via the Embden–Meyerhof pathway, the red cell generates ATP. In addition, this common pathway also can produce NADH, used in methemoglobin reduction, and 2,3-bisphosphoglycerate (2,3-DPG) used for oxyhemoglobin modulation. Additionally, the red cell contains the necessary pathways for oxidative glycolysis through the pentose phosphate pathway. This pathway mainly is used to produce NADPH used in the reduction of glutathione to ultimately provide protection against oxidative damage to the cell.

Pathway Analysis of the Null Space

We will now analyze this reaction network following the approach outlined above to define systemically the metabolic pathways of the red cell and to provide an interpretation of the metabolic machinery of the human erythrocyte. The reaction network analyzed is shown in Fig. 5. These are the Embden–Meyerhof

pathway of anaerobic glycolysis, which includes the Rapoport–Luebering shunt (production of 2,3-DPG), and the pentose phosphate shunt. Following the definitions given above, this system of three interacting pathways consists of 37 internal fluxes and 14 exchange fluxes. Of the 37 internal fluxes, 32 can be paired together as 16 reversible reactions. All of the enzymes represented in the system and the fluxes for which they are responsible are listed in Table 1. The exchange fluxes consist of those for glucose, pyruvate, and lactate, which are all transported across the cell membrane, 2,3-DPG and ribose 5-phosphate, which are

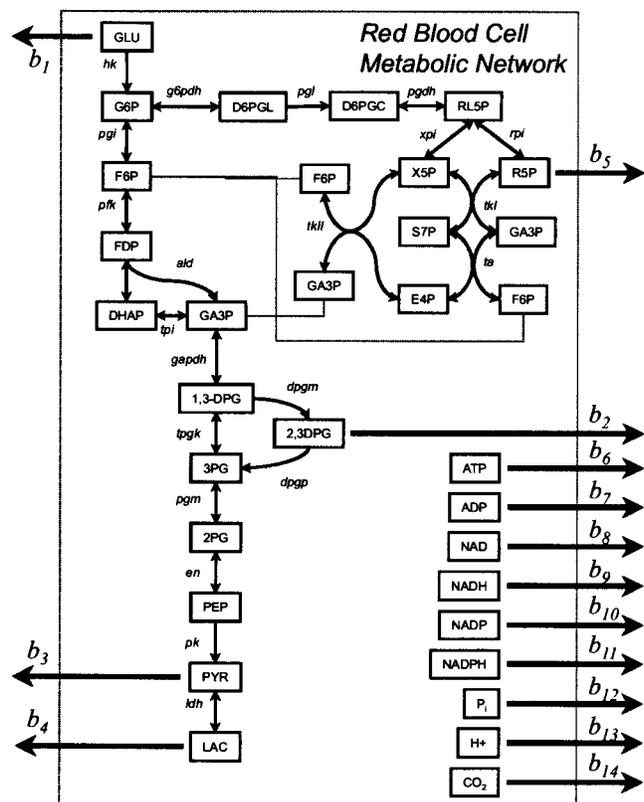


FIG. 5. Metabolic reaction network analyzed from the human erythrocyte containing 29 metabolites and the related 51 fluxes (37 internal and 14 exchange fluxes). Not shown in the figure is the involvement of ATP, ADP, NAD, NADH, NADP, NADPH, CO₂, H⁺, and P_i in associated reactions.

then used for other metabolic activities within the cell such as oxyhemoglobin modulation and adenosine metabolism, respectively. The remaining nine exchange fluxes correspond to those for ADP, ATP, NAD⁺, NADH, NADP⁺, NADPH, CO₂, H⁺, and P_i. There are 29 metabolites involved in the reaction network listed in Table 2. They are classified as internal or exchange metabolite corresponding to the presence of an exchange flux for a particular metabolite. Note that ADP, NAD⁺, NADP⁺, H⁺, and P_i could be removed from the system with no effect on the dimensions of the null space, but for completeness, we have left them in the system in this example.

Material balances on all 29 metabolites were performed to create a stoichiometric matrix (29 × 51) representing the connectivity of the network. The dimension of the null space of the matrix was then determined to be 22. Using MATHEMATICA 3.0 (Wolfram Research Inc., Champaign, IL), a set of basis vectors that spanned the null space was generated. These vectors constituted a theoretically feasible basis; however, there were numerous vectors that contained negative values for the internal fluxes. Therefore, a new set of basis vectors was constructed from the original set by using strict linear combinations of the previous set in accordance with Eqs. 3 and 4. Glucose only was allowed to enter the system whereas the exchange fluxes for pyruvate, lactate, 2,3-DPG, and ribose 5-phosphate were constrained to be positive because they could only leave the system. All of the other exchange fluxes remained unconstrained. Then, 22 linearly independent basis vectors that spanned the null space were constructed that indicated the relative activity of each flux in the pathways.

Sixteen of the 22 basis vectors were constructed to represent the cycling of reversible reactions that have no net effect on the input/output of the system as indicated by the absence in activity of any exchange fluxes. The remaining six vectors constitute true biochemical pathways through the system. The balance equations for these fundamental pathways traced by the basis vectors are shown in Table 3. The first pathway **p**₁ refers to the catabolism of glucose through the Embden–Meyerhof pathway continuing to the formation and removal of lactate. This pathway is the primary route for the formation of ATP. The second pathway **p**₂ also follows the Embden–Meyerhof pathway but uses the Rapoport–Luebering shunt in the ultimate formation and removal of pyruvate. This route can be used for the sole production of the

Table 1. Names and abbreviations of the 21 enzymes included in the red cell reaction network

Flux	Abbreviation	Enzyme
V ₁	hk	Hexokinase
V ₂ , V ₃	pgi	Phosphoglucosomerase
V ₄ , V ₅	pfk	Phosphofruktokinase
V ₆ , V ₇	ald	Aldolase
V ₈ , V ₉	tpi	Triose phosphate isomerase
V ₁₀ , V ₁₁	gapdh	Glyceraldehyde phosphate dehydrogenase
V ₁₂ , V ₁₃	pgk	Phosphoglycerate kinase
V ₁₄	dpgm	Diphosphoglyceromutase
V ₁₅	dpgp	Diphosphoglycerate phosphatase
V ₁₆ , V ₁₇	pgm	Phosphoglyceromutase
V ₁₈ , V ₁₉	en	Enolase
V ₂₀	pk	Pyruvate kinase
V ₂₁ , V ₂₂	ldh	Lactate dehydrogenase
V ₂₃	g6pdh	Glucose-6-phosphate dehydrogenase
V ₂₄ , V ₂₅	pgl	6-Phosphoglyconolactonase
V ₂₆ , V ₂₇	pgdh	6-Phosphoglyconate dehydrogenase
V ₂₈ , V ₂₉	rpi	Ribose-5-phosphate isomerase
V ₃₀ , V ₃₁	xpi	Xylulose-5-phosphate epimerase
V ₃₂ , V ₃₃	tkI	Transketolase
V ₃₄ , V ₃₅	ta	Transaldolase
V ₃₆ , V ₃₇	tkII	Transketolase

The presence of two fluxes related to an enzyme indicates the potential for reversible reactions.

Table 2. Names and abbreviations of the 29 metabolites included in the reaction network

Internal	Exchange	Abbreviation	Metabolite
	x	GLU	Glucose
x		G6P	Glucose 6-phosphate
x		F6P	Fructose 6-phosphate
x		FDP	Fructose 1,6-bisphosphate
x		GA3P	Glyceraldehyde 3-phosphate
x		DHAP	Dihydroxyacetone phosphate
x		13DPG	1,3-Diphosphoglycerate
	x	23DPG	2,3-Diphosphoglycerate
x		3PG	3-Phosphoglycerate
x		2PG	2-Phosphoglycerate
x		PEP	Phosphoenolpyruvate
	x	PYR	Pyruvate
	x	LAC	Lactate
x		D6PGL	6-Phosphogluconolactone
x		D6PGC	6-Phosphogluconate
x		RL5P	Ribulose 5-phosphate
x		X5P	Xylulose 5-phosphate
	x	R5P	Ribose 5-Phosphate
x		S7P	Sedoheptulose 7-phosphate
x		E4P	Erythrose 4-phosphate
	x	CO2	Carbon dioxide
	x	Pi	Inorganic phosphate
	x	ADP	Adenosine 5'-diphosphate
	x	ATP	Adenosine 5'-triphosphate
			Nicotinamide adenine dinucleotide
	x	NAD	Nicotinamide adenine dinucleotide
	x	NADH	Nicotinamide adenine dinucleotide (R)
	x	NADP	Nicotinamide adenine dinucleotide phosphate
			Nicotinamide adenine dinucleotide phosphate (R)
	x	NADPH	Nicotinamide adenine dinucleotide phosphate (R)
	x	H+	Hydrogen ion

Metabolites are classified as internal or exchange compounds based on the presence/absence of an exchange flux for the given metabolite.

reducing power of NADH. The third independent pathway **p**₃ is similar to **p**₁ except that the route ends in the removal of pyruvate as opposed to lactate. This route then results in the net production of ATP and NADH in equimolar amounts. Note that if lactate were dumped by the cell in **p**₂, there would have been no net gain of energy in any form. In addition, this pathway can be constructed from a linear combination of **p**₁, **p**₂, and **p**₃, whereby making it dependent on those vectors and subsequently excluding it from the spanning set of basis vectors. The fourth pathway **p**₄ is the route used for the production of 2,3-DPG, which follows straight down the Embden–Meyerhof pathway taking the first arm of the Rapoport–Luebering shunt. Notice that when 2 moles of 2,3-DPG and NADH are formed for every mole of glucose, there is an expense of 2 moles of ATP. The fifth pathway **p**₅ results in the first use of the pentose phosphate shunt for the production of 1 mole of ribose 5-phosphate and 2 moles of NADPH per mole of glucose. The final pathway **p**₆ represents the classic use of the pentose phosphate shunt for the oxidation of glucose ultimately to generate 12 moles of NADPH for every mole of glucose by cycling carbon compounds through the shunt.

In viewing the metabolic machinery as a factory for energy, redox, and precursor production, we can use these basis vectors to guide our interpretation of this metabolic system. The demands clearly are to produce various metabolic resources in different ratios based on dynamic needs of the cell, while the cell is continuously supplied with glucose from the plasma to be used as its “raw material.” Focusing attention on the activity of the exchange fluxes, we see that the cell essentially has six independent routes through which it can channel glucose with each one addressing specific needs of the cell. The remaining pathways,

Table 3. Net reaction balance equations and functional attributes of basis pathways used to span the null space of the red blood cell stoichiometric matrix

Basis pathway	Net reaction equation	Primary functional attribute
p1	glucose + 2 Pi + 2 ADP → 2 lactate + 2 ATP	ATP energy production for metabolic energy
p2	glucose + 2 NAD+ → 2 pyruvate + 2 NADH + 2 H+	NADH energy production for methemoglobin reduction
p3	glucose + 2 Pi + 2 ADP + 2 NAD+ → 2 pyruvate + 2 ATP + 2 NADH + 2 H+	ATP production for metabolic energy and NADH production for methemoglobin reduction
p4	glucose + 2 ATP + 2 NAD+ → 2 23DPG + 2 Pi + 2 ADP + 2 NADH + 2 H+	2,3-DPG production for oxyhemoglobin modulation
p5	glucose + ATP + 2 NADP+ → R5P + CO ₂ + ADP + 2 NADPH + 2 H+	Ribose 5-phosphate production for adenosine salvaging
p6	glucose + 12 NADP+ → 6 CO ₂ + 12 NADPH + 12 H+	NADPH energy production for glutathione reduction and subsequent antioxidant activity
p7–p22	(no net reaction)	(none)

representing reversible reactions have no net effect on the performance of the system, which for our purposes serves to narrow down our consideration of the null space to six dimensions. There are pathways designed strictly for energy production with the release of certain by-products (i.e., pyruvate and lactate). There are then pathways that generate metabolites (i.e., 2, 3-DPG and ribose 5-phosphate) to be fed into other reaction schemes in the cell. These pathways may cost energy to operate and thus can be balanced by other pathways. As an example, the operation of p₄ producing 2,3-DPG at the expense of ATP can be compensated by the operation of p₁ to solely produce ATP.

Discussion

Because genomics is leading to the complete definition of genotypes of an increasing number of organisms, the definition and conceptualization of biochemical pathways in the context of a whole cell has emerged as an important issue (21). An approach to the study of biochemical reaction systems and their capabilities and performance has been introduced, combining the laws of material conservation in biochemical reactions with the methods of linear algebra. A biochemical system is first balanced and then translated into mathematical representation as a stoichiometric matrix. The null space of the matrix is then spanned in a manner that is theoretically and biochemically feasible providing an explicit description of the null space. By incorporating biochemical knowledge to guide the construction of these basis vectors, we arrive at independent biochemical pathways, which operate within the defined metabolic system. The structure of the null space, the space that contains all possible metabolic phenotypes, is now given in terms of systemically defined pathways.

Treating the structure of metabolic reaction networks in the framework of pathways operating through the system affords many advantages and potential implications for additional conceptual developments. With the ability to decompose a metabolic reaction network into independently operating pathways or modalities, it is possible to gain insight into the regulatory logic implemented by the cell using a new pathway-oriented perspective (22). In addition, the use of such a pathway-oriented approach will undoubtedly assist in attempts to make sense of all of the metabolic components revealed through genomics. By using this approach to pathway definition, organisms can be studied based on the existence of fundamental metabolic pathways, which trace their ability to synthesize required precursors and cofactors.

As is the case in any coordinate translation, there are many sets of bases that can be used to span the null space of a stoichiometric matrix; thus, the selection of bases is nonunique. The development of a useful basis depends on the removal and addition of biochemical and regulatory constraints, as well as the application of interest and how one may wish to interpret the pathways and fluxes operating within a network. In developing an algorithm to perform this critical basis transformation, the challenge is to incorporate biochemical intuition as the guiding force in the

selection of a basis with biochemical importance as opposed to relying purely on the machinery of linear algebra. Algorithm construction is currently underway for the automation of this procedure for the general case.

Traditionally biochemical pathways have been defined in the context of their historical discovery, such as glycolysis, the pentose phosphate pathway, and the citric acid cycle. Through the approach developed here, we move away from the traditional definitions of biochemical pathways to a new classification of pathways—classification based on systemic function as opposed to historical discovery. Defining the functional metabolic pathways in an organism will undoubtedly impact our views of metabolism, from its capabilities to its regulation and even perhaps its evolution. Taken together, the conceptual development and analysis presented represents a step toward the establishment of fundamental concepts on which we can build as we move beyond bioinformatics. Functional analysis and representation of genomes and their characteristics represent one of the most interesting challenges that now faces bioengineering and other quantitative disciplines focused on analysis of fundamental biological phenomena.

We thank the Whitaker Foundation and their support through graduate fellowships in biomedical engineering.

- Koonin, E. V. (1997) *Genome Res.* **7**, 418–421.
- Pennisi, E. (1997) *Science* **277**, 1432–1434.
- Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., et al. (1997) *Science* **277**, 1453–1474.
- Ouzounis, C. & Casari, G. (1996) *Trends Biotechnol.* **14**, 280–285.
- Tatusov, R. L., Koonin, E. V. & Lipman, D. J. (1997) *Science* **278**, 631–637.
- Fields, S. (1997) *Nat. Genet.* **15**, 325–327.
- Metting, F. B. & Romine, M. F. (1997) *Trends Microbiol.* **5**, 91–92.
- Palsson, B. O. (1998) in *Integrating Engineering and Biology: A Revolution in the Making* (Natl. Acad. Press, Washington, DC), in press.
- Edwards, J. S., Schilling, C. H., Ramakrishna, R. & Palsson, B. O. (1998) in *Metabolic Engineering*, eds. Lee, S. Y. & Papoukakis, E. T. (Springer, New York), in press.
- Varma, A. & Palsson, B. O. (1994) *Bio/Technology* **12**, 994–998.
- Bonarius, H. P. J., Schmid, G. & Tramper, J. (1997) *Trends Biotechnol.* **15**, 308–314.
- Varma, A. & Palsson, B. O. (1994) *Appl. Environ. Microbiol.* **60**, 3724–3731.
- Varma, A. & Palsson, B. O. (1994) *Biotechnol. Bioeng.* **43**, 275–285.
- Strang, G. (1988) *Linear Algebra and Its Applications* (Harcourt Brace, Orlando, FL).
- Lay, D. C. (1997) *Linear Algebra and Its Applications* (Addison-Wesley, Reading, MA).
- Papoutsakis, E. & Meyer, C. (1985) *Biotechnol. Bioeng.* **27**, 50–66.
- Pons, A., Dussap, C., Pequignot, C. & Gros, J. (1996) *Biotechnol. Bioeng.* **51**, 177–189.
- Henriksen, C. M., Christensen, L. H., Nielsen, J. & Villadsen, J. (1996) *J. Biotechnol.* **45**, 149–164.
- Bird, R. B., Stewart, W. E. & Lightfoot, E. N. (1960) *Transport Phenomena* (Wiley, New York).
- Paglia, D. E. (1991) in *Hematology: Basic Principles and Practice*, eds. Hoffman, R., Benz, E. J., Shattil, S. J., Furie, B. & Cohen, H. J. (Churchill Livingstone, New York), pp. 269–273.
- Ash, C. (1997) *Trends Microbiol.* **5**, 135–139.
- Stephanopoulos, G. & Simpson, T. W. (1997) *Chem. Eng. Sci.* **52**, 2607–2627.