

A Computational Model for the Identification of Biochemical Pathways in the Krebs Cycle

JOSEPH S. OLIVEIRA,¹ COLIN G. BAILEY,² JANET B. JONES-OLIVEIRA,³
DAVID A. DIXON,⁴ DEAN W. GULL,⁵ and MARY L. CHANDLER⁶

ABSTRACT

We have applied an algorithmic methodology which provably decomposes any complex network into a complete family of principal subcircuits to study the minimal circuits that describe the Krebs cycle. Every operational behavior that the network is capable of exhibiting can be represented by some combination of these principal subcircuits and this computational decomposition is linearly efficient. We have developed a computational model that can be applied to biochemical reaction systems which accurately renders pathways of such reactions via directed hypergraphs (Petri nets). We have applied the model to the citric acid cycle (Krebs cycle). The Krebs cycle, which oxidizes the acetyl group of acetyl CoA to CO₂ and reduces NAD and FAD to NADH and FADH₂, is a complex interacting set of nine subreaction networks. The Krebs cycle was selected because of its familiarity to the biological community and because it exhibits enough complexity to be interesting in order to introduce this novel analytic approach. This study validates the algorithmic methodology for the identification of significant biochemical signaling subcircuits, based solely upon the mathematical model and not upon prior biological knowledge. The utility of the algebraic-combinatorial model for identifying the complete set of biochemical subcircuits as a data set is demonstrated for this important metabolic process.

Key words: Krebs cycle, TCA cycle, molecular reaction, complex formation, enzyme reaction, network, Petri net, hyperdigraph, directed graph, circuit, cycle, unique minimal cycle, spanning tree, signaling subcircuit, null space, equilibrium, Gaussian, distribution.

1. INTRODUCTION

WE HAVE DEVELOPED A COMPUTATIONAL APPROACH (see Oliveira *et al.*, 2001), which depicts sequences of biochemical reactions via specialized directed graphs (cf. Berge, 1973). The creation

¹Radiological & Chemical Sciences Group, National Security and Technology Division, National Security Directorate, Pacific Northwest National Laboratory, Richland, WA.

²School of Mathematical and Computing Sciences, Victoria University of Wellington, Wellington, New Zealand.

³International Technology Assessments Group, National Security and Technology Division, National Security Directorate, Northwest National Laboratories, Richland, WA.

⁴Battelle Fellow, Pacific Northwest National Laboratory, Richland, WA.

⁵PNNL summer student, Southern Utah University, Cedar City, UT.

⁶PNNL summer graduate student, University of Southern California, Los Angeles, CA.

of such network models for biochemical systems enables the elucidation and quantification of the system response to a perturbation. Such models are the first steps toward a goal of facilitating transformation of a biochemical system by a computational model. The methodology is intended to generate a complete database or listing of all of the minimal circuits that principally decompose, or rebuild, the network. Given the complete listing of unique minimal circuits, we can formulate hypotheses regarding the characteristics of the flow of information in the network in the presence or absence of unknown and/or deleted species. We have selected the familiar Krebs cycle to demonstrate the utility, validity, and verification of our underlying approach as well as our computational algorithm.

Graphical network models provide a computational framework for identifying key circuits, and, subsequently, potential oscillatory behaviors and system responses to biochemical perturbation. The model presented here represents the set of all mass-flux balance-conserving pathways or circuits for a given biochemical reaction sequence. The size and complexity of the problem of identifying all such allowable paths and combinations of paths has significant computational complexity, which for large systems will require large computational resources. We have extended previous approaches to this problem (cf. Alberty, 1996, 1994, 1992, 1991, 1991; Clark, 1988; Kauffman, 1971a, 1993; Kohn and Letzkus, 1983; Kohn and Lemieux, 1991; Karp *et al.*, 1998; Seressiotis and Bailey, 1988; Mavrovouniotis and Stephanopoulos, 1990; Reddy *et al.*, 1996; and Schilling *et al.*, 1998, 1999, 1999) by the formulation of a combinatorial geometric model known as an oriented matroid. A thorough explanation and development of the fundamental algebraic-combinatorial mathematics has recently been published (Oliveira *et al.*, 2001).

A directed graph is a data structure that conveys connectivity and direction. It is composed of two sets: a vertex set and an edge set (cf. Oliveira *et al.*, 2001). Circles most often represent vertices; arrows represent edges. A vertex contains a data element, and an edge specifies a rule and direction for the relationship between any two vertices.

A Petri net (cf. Reisig, 1985) is an extension of this notion in which edges (like transitions) are allowed many input and output vertices (like places). In modeling a biochemical pathway with a Petri net, we represent a chemical species by a place in the net, and a chemical event by a transition. By *chemical event*, we refer to any interaction that converts one molecule into another or even processes such as a complex formation. A chemical event could be a classical chemical reaction (e.g., condensation, phosphorylation), an enzymatic process (e.g., substrate binding, release of product), and/or various reactant and product interactions where the chemical nature of the interacting partners is not changed, (e.g., complex formation).

The Petri net model introduced here provides us with an extensive set of combinatorial tools for deducing the qualitative control logic of biochemical networks. This approach defines states in the system to be marked places, which combinatorialists refer to as *boxes*; and the tokens, that are colored with markings that symbolically represent biomolecular species such as metabolites, enzymes, and cofactors, etc., are called *colored balls*. The systematic nature of this modeling approach studies the circuit arrangements or partitions of a biochemical network as functions of marked balls (biochemical species) being arranged into marked places, subject to a set of process control rules defined by the transition conditionals of the Petri net. The tokens are symbolic representations of biomolecular species.

The topology of a Petri net is completely specified by its incidence matrix, whose rows are places and columns are transitions of the Petri net. All matrix entries are either 0, 1, or -1 ; these quantities specify the absence or presence of a connecting edge between two places, as well as its direction.

In addition to its topology or connectivity, a Petri net at a given time has a state or marking that is specified by the number of tokens (here, molecules) in each place. When a reaction or series of reactions takes place, the corresponding transitions are said to have fired, and the token numbers then change commensurate with the stoichiometry of the reaction. For example, if a firing sequence of the Petri net that describes the fumarase reaction fires once, there will subsequently be one less molecule of fumarate, one more molecule of L-malate, and the same number of molecules of fumarase, as shown by the reaction scheme given in Table 1.

Should the number of fumarate molecules be zero, this particular reaction will be unable to fire until some other reaction sequence replenishes this molecule. A useful feature of the Petri net approach that we are taking is that we need not have a detailed, balanced chemical reaction. Rather, we can model the transformation of fumarate to L-malate without detailed knowledge of how the process has occurred at the molecular level. This approach clearly indicates that uncertain parameter information can be effectively included in the network model as long as we know the connectivity.

TABLE 1. FUMARATE + FUMARASE \longleftrightarrow L-MALATE + FUMARASE

fumarate + fumarase \longleftrightarrow fumarate : fumarase,	known as association
fumarate : fumarase \longleftrightarrow L-malate : fumarase,	known as transformation
L-malate : fumarase \longleftrightarrow L-malate + fumarase,	known as dissociation.

Discrete approaches to modeling kinetics, such as the one under consideration, are critical as cells deal with small numbers of molecules in terms of molar quantities. There are often only thousands to millions of a protein in a cell and this is far less than even pM quantities (10^{-12} moles = 6×10^{11} molecules). Each chemical entity reacts in a unit fashion governed by the stoichiometry of the process. Fractions of proteins do not react—whole ones do. Thus, modeling the protein reactions in a living cell requires an integer-based formulation of the reaction network, and Petri nets are totally appropriate for such modeling.

Our current model is qualitative, as it is focused on the connectivity of species through chemical processes which reflect the flow of information in the network. The model can be modified to reflect detailed stoichiometry, enzyme kinetics, enzyme inhibition or activation, and regulatory feedback mechanisms. The Petri net model of biochemical reaction networks satisfies laws of mass conservation by allowing only those firing sequences that obey the specified transition rules. Since the model can specify exact numbers of molecules, it is not subject to round-off errors that might occur with continuous solution approaches for differential rate equation models, e.g., Michaelis–Menton type kinetic enzymatically defined models (cf. Weiss, 1996). Such errors can become consequential when dealing with chemical species for which the total numbers present in a living cell are small.

In order to demonstrate the capabilities of this approach, we have chosen to model a well-understood and well-established biological network. The tricarboxylic acid (TCA) or Krebs cycle is a series of biochemical reactions central to energy production in all eukaryotic cells (cf. Elliott and Elliott, 1997). Substrates for the cycle include the products of glycolysis, protein, and lipid catabolism. Products of the TCA cycle are nucleotide reducing equivalents whose entry into the electron transport chain allows complete oxidation of molecules with concomitant production of adenosine triphosphate (ATP), the major energy molecule of the cell. We constructed our graphical model of the TCA cycle with a focus on its biological and key organic components. We do not include in the model small molecular species such as inorganic phosphate, carbon dioxide, and water, as they are not likely to have limiting concentrations in a living cell. Therefore, reactions involving these components, in some instances, will not be represented although there is nothing in the current approach that precludes us from including them in the model.

2. PETRI NETS

The phenomenological model of a kinetic reaction network can be defined in all generality by a Petri net. The mathematics underlying our approach have previously been described in detail (cf. Oliveira *et al.*, 2001). For a complex network, it is not feasible to completely search a Petri net for its principal circuits. Obtaining operational insight into the network forces us to construct a generalization of the Petri net that defines a hyperdigraph.

Petri nets are hyperdigraphs (cf. Oliveira *et al.*, 2001). The vertices represent the components of the network and are referred to as *places*. The edges represent a set of rules, or conditions, that must be true before transitioning between places; the edges are referred to as *transitions*. When a transition is executed, it has been *fired*. A set of transitions being fired is called a *firing sequence*. As with all hypergraphs, Petri net edges may connect a set of vertices with cardinality greater than two (cf. Berge, 1973).

Associated with a Petri net is a *marking space*. Markings can be thought of as a vector of token counts representing information. A place producing the information is referred to as a *source*; and a place consuming the information is a *sink*. Flux conservation is achieved when the rate at which tokens are being produced equals the rate at which tokens are being consumed. When the flux, for a given firing sequence, starts and ends at the same point, it is called a *circuit*. These circuits are of interest because they represent the paths by which the network is passing as well as conserving information (cf. Oliveira *et al.*, 2001).

Circuits are found by generating an incidence matrix. The rows are labeled by the places, and the columns are labeled by the transitions. Each entry in the incidence matrix will be a 0, 1, or -1 , as noted

above. If information is flowing from a place to a transition, that entry is -1 , representing a loss. If the flow is from a transition to a place, the entry is $+1$, representing a gain. The entry is zero otherwise. Once the incidence matrix N is generated, the left nullspace can be found by finding solutions of $N^t v = 0$, where n corresponds to the number of reacting species within the biochemical system (places), m corresponds to the number of reactions that are taking place (transitions), and v is the n -dimensional solution vector.

Stencils used in biochemistry

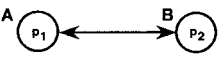
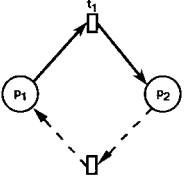
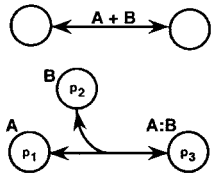
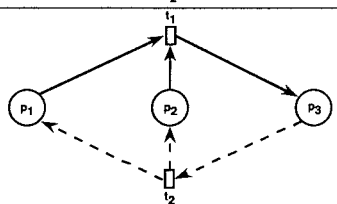
Complex biochemical processes can be considered to be composed of two fundamental types of biochemical building blocks: molecular reaction and complex formation. These fundamental building blocks, summarized in Table 2, can be considered as Petri net stencils. Each of these building blocks has associated with it a Petri net representation, which pictorially depicts the communication pathways, and its incidence matrix, which mathematically specifies which places are receiving and/or transmitting information to each other subject to the transition rules. Forward paths are denoted by solid lines with directional arrows, and backward paths are denoted by dashed lines with directional arrows. In the incidence matrix, a place $[p]$ would have a $+1$ entered for transition $[t]$ if information is propagating subject to the transition node into the place node; similarly, a place would have an entry -1 for the transition if the information is propagating from the place node subject to the transition node.

Description of an enzymatic reaction

In our approach, an enzyme-catalyzed reaction, as described in Table 3, is modeled as a series of separate steps (not necessarily the actual physical process): (1) interaction of substrates, (2) association of the interacting substrates with the enzyme, (3) transformation of substrates to products by the enzyme, (4) dissociation of products from the enzyme, and (5) separation of the products from one another. Our model's representation of these processes as separate, sequential steps is a mathematical construct, which does not compromise the actual physical results. Our focus is on accounting for each molecular species and its corresponding set of possible reactions. The biochemical reaction is viewed as an edge or subset of edges in a graph, depicting flow of information rather than detailed reaction mechanisms.

In order to identify the circuit decomposition of a general network, it is necessary that the network be represented as a graph. Accordingly, the hyperdigraph representation of the Petri net must be transformed in order to satisfy the more general condition of being a graph. There are several ways to obtain a faithful graph representation of the hyperdigraph. One can either transform the hyperdigraph into a directed bipartite graph (cf. Narayanan, 1997) or, as we have chosen, transform the hyperdigraph into an undirected graph

TABLE 2. FUNDAMENTAL PETRI NET STENCILS^a

Reaction Type	Chemistry Representation	Reaction Representation	Petri Net Representation	Incidence Matrix Representation												
Molecular Reaction	$A \leftrightarrow B$			<table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <td></td> <td>t_1</td> <td>t_2</td> </tr> <tr> <td>p_1</td> <td>-1</td> <td>1</td> </tr> <tr> <td>p_2</td> <td>1</td> <td>-1</td> </tr> </table>		t_1	t_2	p_1	-1	1	p_2	1	-1			
	t_1	t_2														
p_1	-1	1														
p_2	1	-1														
Complex Formation	$A + B \leftrightarrow A : B$			<table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <td></td> <td>t_1</td> <td>t_2</td> </tr> <tr> <td>p_1</td> <td>-1</td> <td>1</td> </tr> <tr> <td>p_2</td> <td>-1</td> <td>1</td> </tr> <tr> <td>p_3</td> <td>1</td> <td>-1</td> </tr> </table>		t_1	t_2	p_1	-1	1	p_2	-1	1	p_3	1	-1
	t_1	t_2														
p_1	-1	1														
p_2	-1	1														
p_3	1	-1														

^aThe transition nodes are physically identified for reversible reactions. The presence of the t_1 and t_2 transitions guarantees microscopic reversibility.

TABLE 3. ENZYMATIC REACTION PETRI NET STENCIL^{a,b,c}

Reaction Type	Chemistry Representation	Reaction Representation	Petri Net Representation	Incidence Matrix Representation																																										
Enzyme Reaction	$S + E \leftrightarrow S : E$ $S : E \leftrightarrow P : E$ $P : E \leftrightarrow P + E$			<table border="1"> <thead> <tr> <th></th> <th>t_1</th> <th>t_2</th> <th>t_3</th> <th>t_4</th> <th>t_5</th> <th>t_6</th> </tr> </thead> <tbody> <tr> <th>p_1</th> <td>1</td> <td>-1</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>p_2</th> <td></td> <td>1</td> <td>-1</td> <td></td> <td>-1</td> <td>1</td> </tr> <tr> <th>p_3</th> <td></td> <td></td> <td></td> <td>1</td> <td>-1</td> <td>1</td> </tr> <tr> <th>p_4</th> <td></td> <td></td> <td></td> <td>-1</td> <td>1</td> <td>-1</td> </tr> <tr> <th>p_5</th> <td></td> <td></td> <td></td> <td>1</td> <td>-1</td> <td>1</td> </tr> </tbody> </table>		t_1	t_2	t_3	t_4	t_5	t_6	p_1	1	-1					p_2		1	-1		-1	1	p_3				1	-1	1	p_4				-1	1	-1	p_5				1	-1	1
	t_1	t_2	t_3	t_4	t_5	t_6																																								
p_1	1	-1																																												
p_2		1	-1		-1	1																																								
p_3				1	-1	1																																								
p_4				-1	1	-1																																								
p_5				1	-1	1																																								

(a) An enzyme reaction is composed of one molecular reaction and two complex formations.

edge index	1	2	3	4	5	6	7	8	9	10
connectivity	$e_1[4, 1]$	$e_2[1, 4]$	$e_3[4, 5]$	$e_4[5, 4]$	$e_5[2, 5]$	$e_6[5, 2]$	$e_7[4, 3]$	$e_8[3, 4]$	$e_9[3, 5]$	$e_{10}[5, 3]$
transition	t_1	t_2	t_3	t_4	t_5	t_6	t_1	t_2	t_5	t_6
place										
vertex										
p_1	1	-1	0	0	0	0	0	0	0	0
p_2	0	0	0	0	-1	1	0	0	0	0
p_3	0	0	0	0	0	0	1	-1	-1	1
p_4	-1	1	-1	1	0	0	-1	1	0	0
p_5	0	0	1	-1	1	-1	0	0	1	-1

(b) Incidence matrix for the digraph representation of the enzyme reaction.

^aThe Petri net representation and incidence matrix are presented for a standard enzymatic reaction. Also, the transformed enzymatic reaction incidence matrix is presented, which corresponds to Fig. 1.

^bTop: An enzyme reaction is composed of one molecular reaction and two complex formations.

^cBottom: Incidence matrix for the digraph representation of the enzyme reaction.

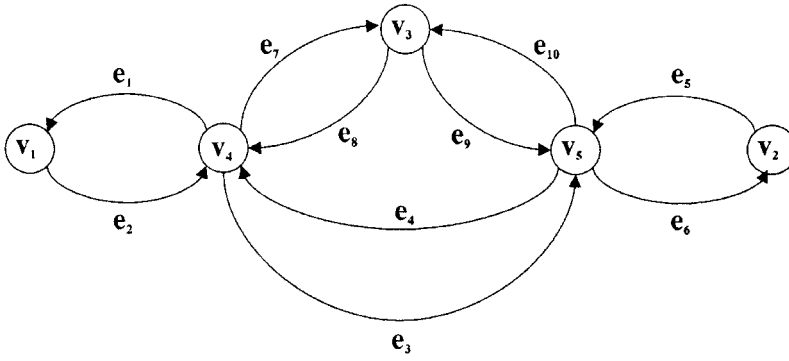


FIG. 1. The enzyme reaction digraph.

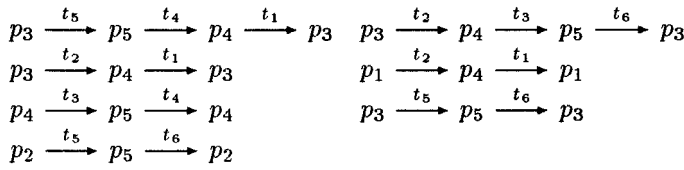


FIG. 2. The complete labeled pathway listing of the transformed enzyme reaction. Note that there are 5 two-cycles and 2 three-cycles.

while maintaining a record of the added edges and compressed multi-edges and all of the directional information (cf. Berge, 1973).

We observe that the reason for this need to transform the hyperdigraph is because there may be multiple inputs and/or outputs associated with the transitions and our current mathematical machinery requires a true graph. The molecular reaction presented in Table 2, which is a simple reversible reaction path, is referred to as a simple circuit in the network model and a two-cycle in the graph representation. The complex formation is a nonsimple example. As can be seen in the Petri net representation of complex formation, transition $[t_1]$ contains the rule combining the flow of information from places $[p_1]$ and $[p_2]$ to $[p_3]$; similarly, transition $[t_2]$ contains the rule splitting the flow of information from place $[p_3]$ into places $[p_1]$ and $[p_2]$. Our representation assumes that all transitions are simply unary, where the regulation cannot be specified. We remedy this problem by associating a graph with a Petri net that has exactly the same cycles as the net has circuits.

Consider one such example involving nonsimple reactions, the enzyme reaction presented in Table 3(a). The incidence matrix of the Petri net representation yields an associated family of edge-sets with only six minimal elements corresponding to the three two-cycles: $\{t_1, t_2\}$, $\{t_3, t_4\}$, and $\{t_5, t_6\}$. Actually, the first and last of these each give rise to two distinct cycles. We are also missing two three-cycles. It is clear that this approach does not capture all of the information.

The readily constructed graph for the enzyme reaction is presented in Fig. 1, with its incidence matrix as presented in Table 3(b).¹ We have introduced a slight change in notation from that of the places $[p]$ to the vertices $[v]$ and from transitions $[t]$ to edges e such that $[p_4] \rightarrow [t_3] \rightarrow [p_5]$ is denoted by e_{10} . In this way, we may distinguish between multiple paths passing through a single transition. For example, the communication of information through transition $[t_1]$ is now split such that $p_4 \xrightarrow{t_1} p_3$ becomes e_2 and $p_4 \xrightarrow{t_1} p_1$ becomes e_7 .

The complete list of correspondences between transitions and digraph edges is provided in Table 3(b). The basis size was determined to be six. The seven unique minimal cycles or paths we found are presented in Fig. 2.

¹This previously published example is repeated here in order to correct a minor error.

Computational identification of the minimal cycles

To reduce the size of the nullspace to a finite number, we have devised a search algorithm, which we call “successive simplification.” The algorithms are based on the fundamental mathematics previously described by us. We first transform the hyperdigraph to obtain a directed graph, which has cycles in one-to-one correspondence with the circuits of the directed bipartite graph representation of the Petri net. Then, we use the fact that the cycles of an undirected graph form a \mathbb{Z}_2 -vector space, which is the nullspace of the incidence matrix considered as a \mathbb{Z}_2 -matrix. Hence, we remove the directions on the digraph, condense repeated edges into a single edge, and find all cycles of the undirected graph. The vector space of cycles has dimension $m - n + 1$ where m is the number of edges of the transformed matrix, and n the number of vertices. Hence the space has size 2^{m-n+1} . A basis for this vector space may be found from a spanning tree for the graph. A list of all cycles in the digraph may then be generated with a simple backtracking search algorithm. A simple bookkeeping procedure is used to keep track of the labels, i.e., the assignments of specific numerical entries to the appropriate place and transition. In all generality, this can be a tedious process, but, a linear computational complexity order is still provably guaranteed. Our reversibility assumption simplifies this process considerably—every undirected cycle gives rise to exactly two positive cycles—one in each direction. Of course, our list will contain nonminimal cycles—because, for example, figure eights may arise—and so we avoid listing nonminimal cycles, not by checking for minimality, but by directly listing a minimal subcycle of every cycle we find. This generates repetitions, which are easily removed at a later time, retaining a linear complexity order in both time and space. The final list of minimal unique cycles is the smallest set of cycles required to compose any physically meaningful cycle that the network system is capable of performing. This listing is guaranteed to be complete due to the exhaustive nature of the search.

3. THE KREBS CYCLE

The Krebs cycle is the second stage in glucose oxidation. The Krebs cycle takes the products of glycolysis, which are pyruvates, and converts two of the carbons in pyruvate to CO_2 and also transfers electrons to electron carriers. As part of this process, three molecules of nicotinamide adenine dinucleotide (oxidized form) [NAD] (where NAD is NAD^+) are reduced to nicotinamide adenine dinucleotide H (reduced form) [NADH] and one molecule of flavin adenine dinucleotide (oxidized form) [FAD] is converted to flavin adenine dinucleotide H_2 (reduced form) [FADH₂].

Fifty-nine places and 86 transitions are required to specify the Krebs cycle, in a simplified form where we have not tracked the number of water molecules nor the CO_2 released. From an operational control systems theory point of view, the transitions $[T]$ act as the control laws for the system whereas the places $[P]$ act as the place variables in the Petri net representations of the biochemical reactions. The Petri net representation is a combinatorial abstraction of the molecular interactions defined over a chemical reaction space where the transitions define the operational conditions or rules that must be satisfied for a reaction to occur. In this graphical representation, a chemical species moves from one place to another subject to a transition rule based on chemical equilibria (thermodynamics) or kinetics. The chemical species do not pass formally through the transition but rather are subject to the rules described by the transition.

Our primary concern is with the connectivity of the network made up from these biochemical stencils. Two place nodes are connected subject to a transition node simply if it is possible for the second place to be reached from the first place through some physical/chemical mechanism which is reversible. Of course, the actual amount of chemical system reversibility may be very small and is dependent on the equilibrium constant and/or kinetic rate constants. Because all of the reactions under consideration are potentially reversible, these paired sets of transitions are identified explicitly. The implications of the complexity that can be incorporated in the transition rules is deferred to a subsequent paper, and, in this discussion, only the existence of the species moving through the places of the pathways is indicated. In other words, we do not assign weights or probabilities to the paths of the reactions; we only indicate the existence of the reaction paths and their directions, i.e., how information flows in the network. We are not considering the quantities of molecular species present, the equilibrium constants, their reaction rates, or the timed sequence of events.

Forward paths are designated by directed solid lines, and backward paths are designated by directed dashed lines. Forward and backward paths are given equal weight in terms of describing the possible paths

along which information can flow. We consider that two places, or even subreactions, are connected regardless of whether that communication of information is via a forward or a backward path or combinations thereof, indicating only whether a physical connection is possible. Our assumption of reversibility of the system requires adding extra edges to ensure that every connection between two places is both forward and backward; i.e., a pair of edges always exist.

Specification of the places

The Krebs cycle is described by 59 places, one for each species in the cycle. The list of places is presented in Table 4. The places have been color coded according to the subreactions comprising the graph cycle. The places printed in bold are those later identified by the mathematical model to be critical. The places represent the various key chemical species, other than H^+ , H_2O , and CO_2 , produced or consumed in the Krebs cycle.

The Krebs cycle and its subreactions

The Krebs cycle was modeled after the information given by Darnell *et al.* (1990) and Stryer (1988). Our Petri net representation of the Krebs cycle is presented in Fig. 3. (See www.biomolecular.org/publications/Krebs/index.html). The nucleotide cofactor places are indicated in red, the enzymes in green, and all other species are indicated in blue. The details of the figure are obscured by the requisite scaling down of the figure to accommodate the published page, so we have included larger renditions of the subreactions as presented in Figs. 3 through 10. However, despite the reduced size of Fig. 3, the cross-couplings—involving places P57 or NADH, P58 or NAD, and P59 or coenzyme A [CoA]—are particularly noteworthy. These nucleotide cofactors are involved in interactions between the subreactions which provide alternative pathways through the Krebs cycle. The physical limitations associated with including this figure in a readable format actually emphasizes the issue of complexity in biological networks. The Krebs cycle is a relatively simple set of biochemical reactions; however, even its complexity is such that a complete description becomes quite large. Our approach is that once a level of confidence in this procedure is established, the mathematics will be relegated to the computational domain, which will produce data sets which can then be analyzed with the confidence that the true complexity has been properly modeled.

Having established the Petri net, the incidence matrix is then constructed. In fact, the incidence matrix may be constructed directly from the biochemistry without the need for physically drawing the network. The incidence matrix is presented in Table 5. (See www.biomolecular.org/publications/Krebs/index.html.) Blocks of color, which match the color coding of the Krebs cycle Petri net subreactions, are visible. Note the tremendous amount of natural structure in this nearly block-banded matrix. The cross-couplings of places P57 [NADH], P58 [NAD], and P59 [CoA] present the exceptions to the otherwise block-banded structure. However, it is just these cross-couplings, together with the coupling overlaps, that provide the network's richness, adaptability, and combinatorial flexibility. Again, to fit the entire table onto a single page in order to stress the inherent structure of the matrix makes it difficult to read. Magnified submatrices associated with some of the subreactions will follow the magnified subreactions.

Before we begin a discussion of the individual subreactions of the Krebs cycle, there are interesting features to be noted regarding the nucleotide cofactors: P57 [NADH], P58 [NAD], and P59 [CoA]. All of the forward paths go into P57 (from Subreactions 4, 5, and 9) and all of its backward paths leave from P57 (to other species within Subreactions 4, 5, and 9). Conversely, all of the forward paths lead from P58 (from Subreactions 4, 5, and 9) and all of its backward paths lead into P58 (again, interaction with other species within Subreactions 4, 5, and 9). P59 is different. CoA has forward paths leading into it from Subreactions 1 and 6, and its third forward path leads from CoA into Subreaction 5. CoA has backward paths to Subreactions 1 and 6 and from Subreaction 5. The direction of cross-coupling flow has significant ramifications to the importance of specific species in the analysis of the Krebs cycle system.

As discussed earlier, our interest is in the connectivity of the species, without prejudice to forward path, backward path, timing, etc. If a physical mechanism exists that allows information from place 1 to flow to place 2, it is included and given equal weight. The set of minimal cycles is determined and then examined to investigate which places and which transitions appear most and least often in the minimal cycle set. These noteworthy places and transitions may then be examined in terms of their biochemical importance. We emphasize that our listing of cycles is guaranteed to be the complete set of minimal cycles.

TABLE 4. KREBS CYCLE PETRI NET PLACE DEFINITIONS

Place	Species	Type
P1	Acetyl CoA	Nucleotide cofactor
P2	Oxaloacetate : acetyl CoA	
P3	Oxaloacetate : acetyl CoA : citrate synthase	
P4	Citrate : CoA : citrate synthase	
P5	Citrate synthase	Enzyme
P6	Citryl CoA : citrate synthase	
P7	Citrate : CoA	
P8	Citrate	
P9	Citrate : aconitase	
P10	Isocitrate : aconitase	
P11	Aconitase	Enzyme
P12	cis-Aconitate : aconitase	
P13	Isocitrate	
P14	Isocitrate : isocitrate dehydrogenase	
P15	α -Ketoglutarate : isocitrate dehydrogenase	
P16	Isocitrate dehydrogenase	Enzyme
P17	Isocitrate : NAD : isocitrate dehydrogenase	
P18	Oxalosuccinate : NADH : isocitrate dehydrogenase	
P19	Oxalosuccinate : isocitrate dehydrogenase	
P20	α -Ketoglutarate	
P21	α -Ketoglutarate : α -ketoglutarate dehydrogenase complex	
P22	Succinyl CoA : α -ketoglutarate dehydrogenase complex	
P23	α -Ketoglutarate dehydrogenase complex	Enzyme
P24	α -Ketoglutarate : NAD : α -ketoglutarate dehydrogenase complex	
P25	α -Ketoglutarate : NAD : CoA : α -ketoglutarate dehydrogenase	
P26	Succinyl CoA : NADH : α -ketoglutarate dehydrogenase complex	
P27	Succinyl CoA	
P28	Succinyl CoA : succinyl CoA synthetase	
P29	Succinate : succinyl CoA synthetase	
P30	Succinyl CoA synthetase	Enzyme
P31	Succinyl phosphate : succinyl CoA synthetase : CoA	
P32	Succinyl phosphate : succinyl CoA synthetase	
P33	Succinate : phosphate : succinyl CoA synthetase	
P34	Guanosine diphosphate [GDP]	Nucleotide cofactor
P35	Succinate : phosphate : GDP : succinyl CoA synthetase	
P36	Succinate : GTP : succinyl CoA synthetase	
P37	Guanosine triphosphate [GTP]	Nucleotide cofactor
P38	Succinate	
P39	Succinate : succinate dehydrogenase	
P40	Fumarate : succinate dehydrogenase	
P41	Succinate dehydrogenase	Enzyme
P42	Flavin adenine dinucleotide (oxidized form) [FAD]	Nucleotide cofactor
P43	Succinate : FAD : succinate dehydrogenase	
P44	Fumarate : FADH ₂ : succinate dehydrogenase	
P45	Flavin adenine dinucleotide H ₂ (reduced form) [FADH ₂]	Nucleotide cofactor
P46	Fumarate	
P47	Fumarate : fumarase	
P48	L-malate : fumarase	
P49	Fumarase	Enzyme
P50	L-malate	
P51	L-malate : malate dehydrogenase	
P52	Oxaloacetate : malate dehydrogenase	
P53	Malate dehydrogenase	Enzyme
P54	L-malate : NAD : malate dehydrogenase	
P55	Oxaloacetate : NADH : Malate dehydrogenase	
P56	Oxaloacetate	
P57	Nicotinamide adenine dinucleotide H (reduced form) [NADH]	Nucleotide cofactor
P58	Nicotinamide adenine dinucleotide (oxidized form) [NAD]	Nucleotide cofactor
P59	Coenzyme A [CoA]	Nucleotide cofactor

^aCommonly used abbreviations are indicated in brackets.

Petri nets of the Krebs cycle subreactions. Subreaction 1, depicted in Fig. 3, is connected to Subreaction 9 via a forward path and from Subreaction 9 via a backward path. The citrate in Subreaction 1 is connected to, or may even be considered a part of, the combined Subreactions 2/3 via a forward path and from Subreactions 2/3 via a backward path. Finally, citrate:CoA is connected to CoA, which is connected to Subreaction 5 in a forward path (and from a backward path) and to Subreaction 6 in a backward path (and from a forward path). Note that CoA is included as a complex formation, similar to acetyl CoA. It is not drawn the same way because CoA is shared and appears in the network in several other subreactions. Note that Subreactions 2 and 3 were combined because the aconitase enzyme catalyzes two consecutive Krebs cycle reactions. Subreactions 2/3, depicted in Fig. 5, as previously stated, are connected to the citrate of Subreaction 1 via a backward path and from the same citrate via a forward path. The isocitrate of Subreactions 2/3 is also a species of Subreaction 4. Subreaction 4, depicted in Fig. 6, shares isocitrate with Subreactions 2/3. It also shares α -Ketoglutarate with Subreaction 5. Tremendous complexity and system versatility is provided in Subreaction 4 from its communication via cofactors NAD and NADH. Subreaction 5, depicted in Fig. 10, appears later in the text due to its orientation when extracted from Fig. 3. Note that Subreaction 5 shares α -ketoglutarate with Subreaction 4 and it shares succinyl CoA with Subreaction 6. Further, it couples to both Subreaction 1 and the combined Subreactions 2/3 and to Subreaction 6 through P59 [CoA]. It also couples to Subreactions 4 and 9 through both NAD and NADH. Subreaction 6, depicted in Fig. 7, involves nucleotide cofactors guanosine diphosphate [GDP] and guanosine triphosphate [GTP]. It shares succinyl CoA with Subreaction 5 and succinate with Subreaction 7. Its communications through cofactor CoA have already been discussed. Subreaction 7, depicted in Fig. 8, involves nucleotide cofactors FAD and FADH. It shares succinate with Subreaction 6 and fumarate with Subreaction 8. Subreaction 8, depicted in Fig. 9, shares fumarate with Subreaction 7 and L-malate with Subreaction 9. Lastly, Subreaction 9, also depicted in Fig. 10, completes the cycles with its sharing of oxaloacetate with Subreaction 1. Its cross-couplings have already been discussed. Clearly, this system is able to represent an extensive collection of circuits via forward and backward paths. Some admissible paths, however physically improbable, could easily be missed using alternative methods of analysis of the Krebs cycle, and yet they may prove to be quite interesting biochemically and/or pharmaceutically.

Incidence matrices of the Krebs cycle subreactions. In Table 6, we expand the incidence matrix entries for Subreactions 1, the combined 2/3 and 4 from Table 5. Note the nearly block-banded matrix structure, the overlapping of the blocks associated with adjacent subreactions, and the cross-couplings accomplished by the P57, P58 and P59 cofactors.

Identification of relevant biochemical circuits

The directed connections between the transitions and the places, as specified by the Krebs cycle presented in Fig. 3, are the input for our computer code. The code first identifies 73 two-cycles because the algorithm suppresses multiple edges and would not identify them otherwise. These same two-cycles can be recognized as all of the molecular reactions in the Krebs cycle diagram, noting that each complex formation contains two two-cycles and each enzymatic reaction contains five two-cycles. The two-cycles may be counted by traversing the Krebs cycle backbone starting and ending at P56, which identifies 43 two-cycles, and then adding an additional two-cycle for each of the 14 complex formations, and two additional two-cycles for each of the 8 enzyme reactions, for the total of 73.

The code next determines that the basis size is 15. The implication is that there are then $(2^{15} - 1)$ possible minimal cycles through the undirected graph or $(2^{15+1} - 2)$ possible minimal cycles through the transformed directed fundamental network graph. It is noted that there is an empty cycle (i.e., with no edges) in the undirected graph and two empty cycles in the directed graph, which the algorithm will not count, and so the code counter is set accordingly. In theory, all of the closed circuits achievable in this network can be composed from the minimal set of circuits. In actuality, the size of the directed characterizing data set is considerably smaller than the 2^{16} limiting theoretical value. Subsequently, a script is run to identify the set of unique minimal cycles by removing the duplicates.

There are 73 two-cycles and 15,820 other cycles and the 1 zero-cycle, for a total of 15,894 unique minimal cycles. This complete listing of cycles is then placed into a database for further analysis. The total number of unique minimal cycles is considerably less than the limiting value of 2^{16} and is actually a factor of four less than the theoretical number due to the removal of nonminimal cycles. All 15,893 unique

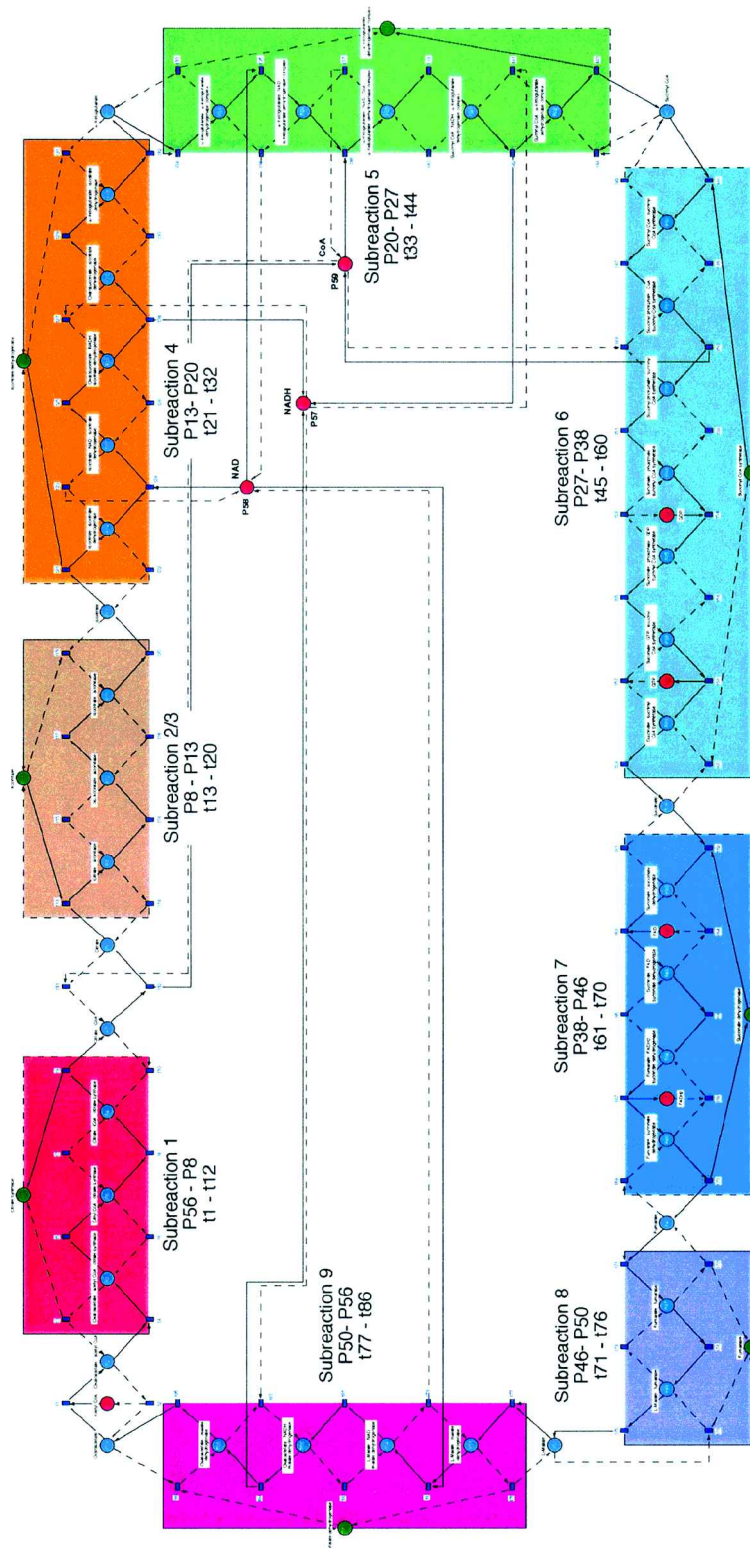


FIG. 3. Petri net of the nine-reaction Krebs cycle. Each of the subreactions are repeated in Figs. 3 through 10 for readability; however, the entire composite hyperdigraph is included in order to show how the cofactor places P57 (NADH), P58 (NAD), P59 (CoA) provide a transnetwork coupling offering significant network redundancy and complexity.

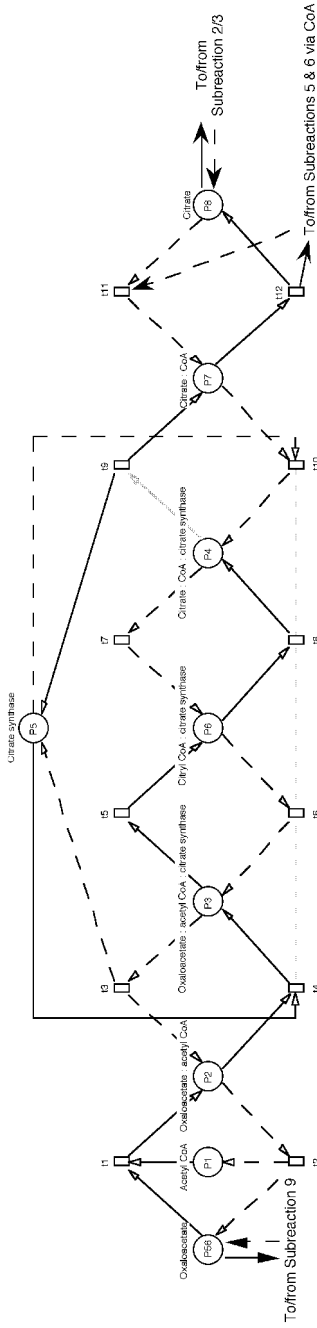


FIG. 4. Subreaction 1: An acetyl residue from acetyl CoA condenses with oxalacetate to form citrate with a catalyzed reaction by the citrate synthase enzyme.

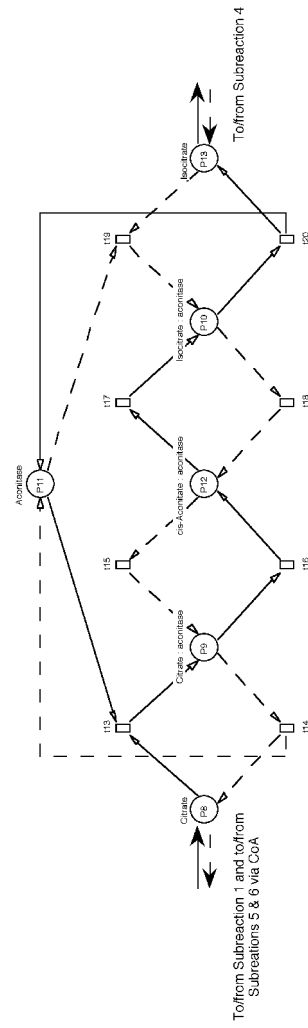


FIG. 5. Subreactions 2/3: Catalyzed reaction by the aconitase enzyme.

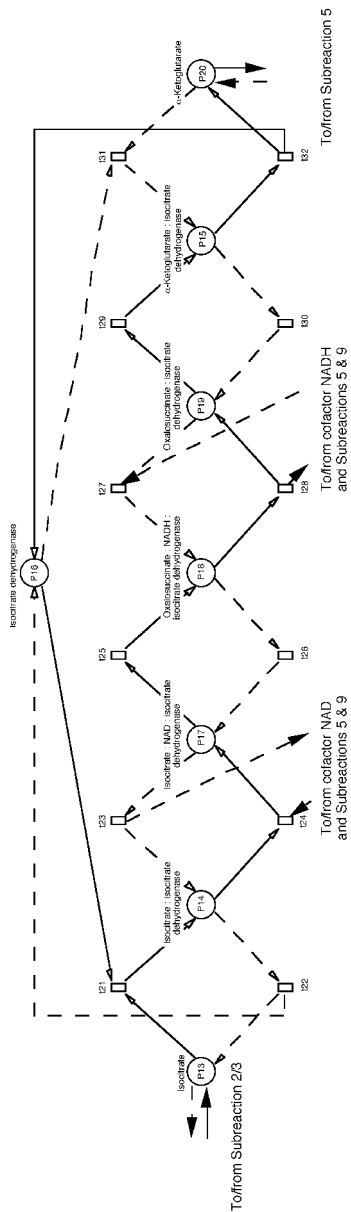


FIG. 6. Subreaction 4: Catalyzed reaction by the isocitrate dehydrogenase enzyme. Note that Subreaction 4 couples to Subreactions 5 and 9 through both P57 [NADH] and P58 [NAD].

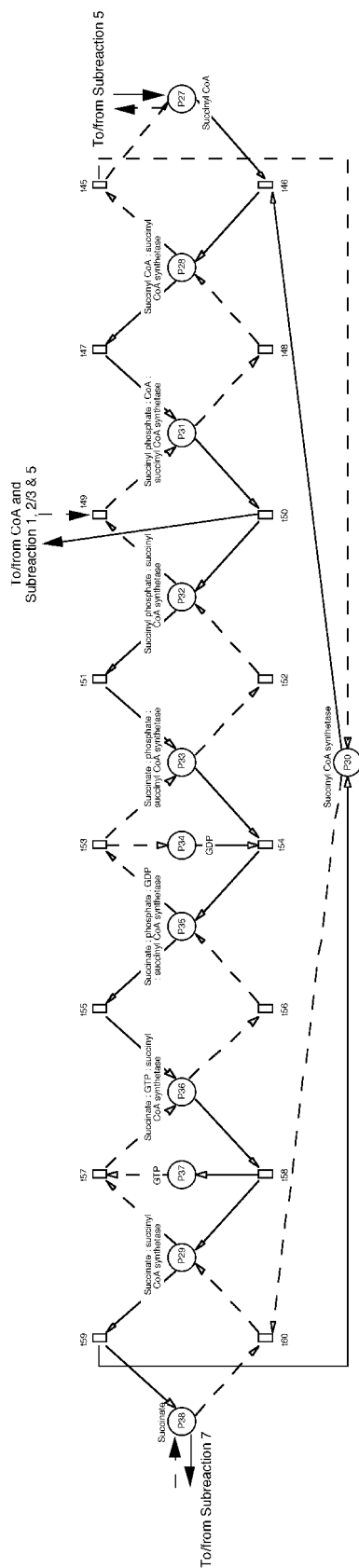


FIG. 7. Subreaction 6: Catalyzed reaction by the succinyl CoA synthetase enzyme.

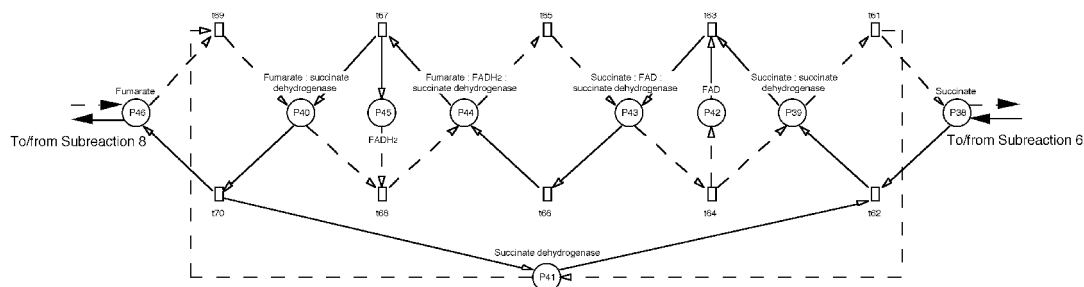


FIG. 8. Subreaction 7: Catalyzed reaction by the succinate dehydrogenase enzyme.

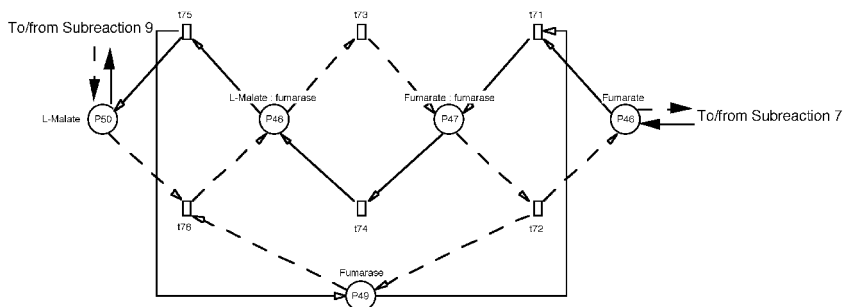


FIG. 9. Subreaction 8: Catalyzed reaction by the fumarase enzyme.

minimal circuits, corresponding to the 15,893 nonzero cycles, with their label information are required in combination to completely characterize the Krebs cycle network. It is a tremendously rich network, which has evolved with multitudes of alternative routes. There is a profound diversity provided by taking combinations of 15,893 circuits.

This collection of unique minimal pathways is the complete listing of closed walks that traverse the network from beginning to end. The unique minimal pathways are the shortest, nonrepeating, nonlooping, nonrevisited paths. Once this data set of circuits that decompose this network as a set of minimal cycles of the graph has been identified, any number of search strategies may be specified as linear objective functions over this database of minimal cycles. Once edge weights/markings are added to the hyperdigraph representation, the network can be analyzed to identify the set of paths which are of particular biochemical interest via optimization methods using path algebras or linear programming/oriented matroid programming.

As a first approach to analyzing this data set, we look for biochemically meaningful information, such as identifying the transitions or places appearing most often or least often in the set of unique minimal cycles. Other questions that can be answered by postprocessing of this data set involve the identification of alternate pathways through the Krebs cycle. For example, if a place or transition in the Krebs cycle is disabled, what alternate pathways remain from place 1 to place 2? The alternative pathways can be identified by inspecting the complete listing of circuits, eliminating those circuits containing the disabled place or transition. Frequency distributions are presented indicating the number of occurrences of each transition or place in the list of unique minimal cycles. Extreme values in the network analysis are of interest, be they valleys (i.e., infrequent occurrences) or peaks (i.e., frequent occurrence).

Information on transitions. There are 86 transitions required to specify the Krebs cycle. The transitions, the count of their occurrences in the 15,893 unique minimal cycles, and the percentage of their occurrences in the 15,893 unique minimal cycles are presented in Table 7. They are listed in order of decreasing occurrence. Bar plots for transitions 1 through 86 versus percentage of occurrence in the 15,893 unique minimal cycles are presented in Fig. 11. The graph is arranged in order of transition number. Because we are assuming that all reactions are reversible with $k = 1$, each pair of transitions occurs with the same frequency, as shown in Table 7 and Fig. 11. We have listed the transition occurrences in pairs; however,

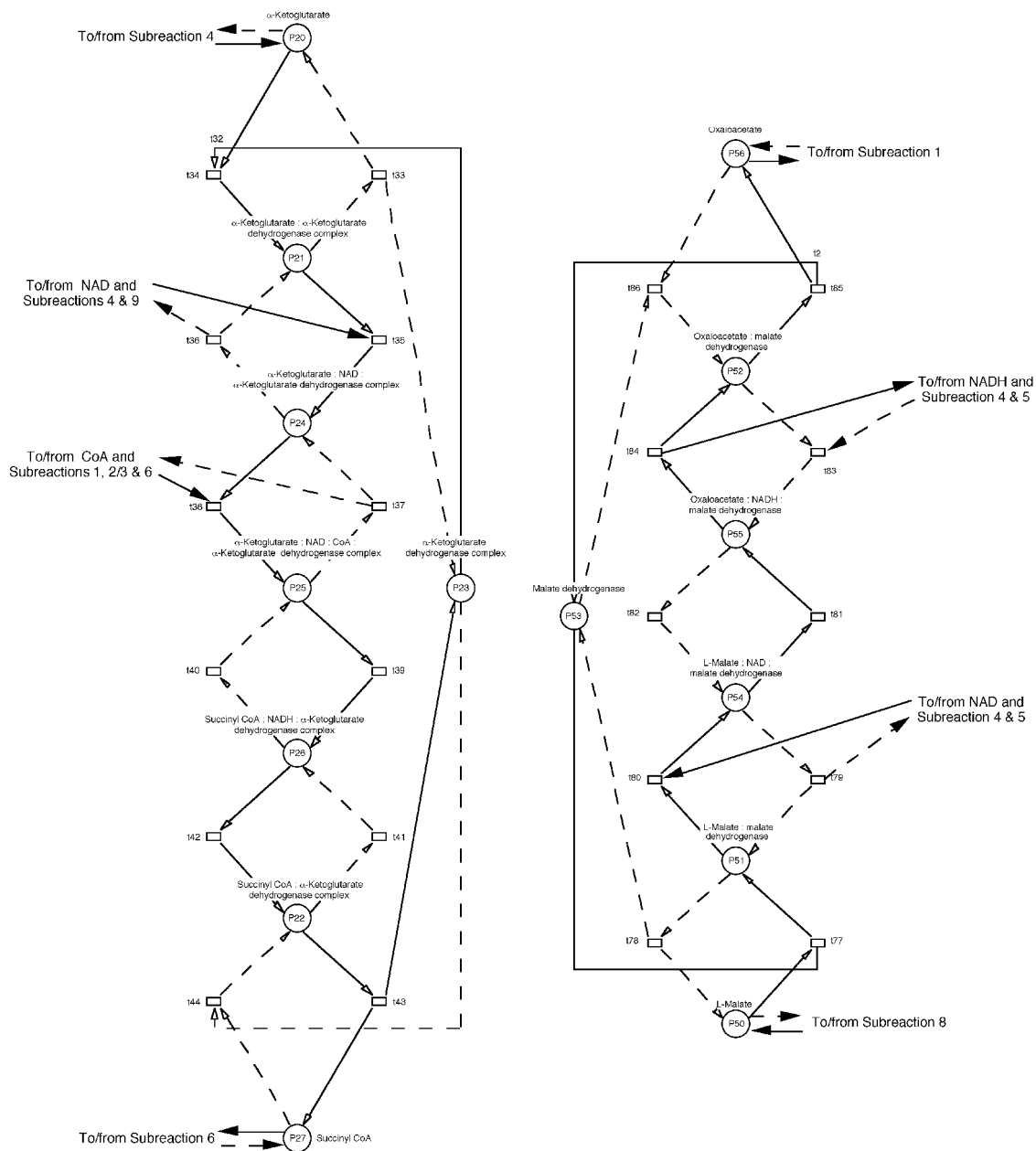


FIG. 10. Subreaction 5 (on the left) and Subreaction 9 (on the right).

in subsequent analyses for which the near equilibrium assumption will not be made, the values can and will be different.

The transition bar plot enables us to compare the low percentage of occurrence transitions with the high percentage of occurrence transitions. A low percentage of transitions means that information does not pass as readily in these parts of the cycle and a high percentage means that these transitions are readily passing information. Recall that here we assume that the equilibrium constants and the rate constants are unity.

The first set of low transition values (t_5 – t_8) corresponds to passing through the citrate synthetase, P5. The next set (t_{15} – t_{18}) corresponds to aconitase (P11) reactions. The next set of low values occur in Subreaction 4 (t_{25} , t_{26} , t_{29} , and t_{30}), corresponding to isocitrate dehydrogenase (P16) complex formation. These low values are split by large transitions coupling to NAD/NADH₂. Low values are also found at t_{39} and t_{40} , involving α -ketoglutarate dehydrogenase (P23) as well as with succinyl CoA (P27) synthetase. There are

a large number of low percentage transitions (t_{47} , t_{48} , t_{51} – t_{58}) in the CoA synthetase pathway, including the GDP \rightarrow GTP reaction. Another group of low percentage transitions (t_{63} – t_{68}) is found for succinate dehydrogenase (P41) and the formation of FADH₂ (P45) from FAD (P43). The final low percentage transitions, (t_{73} and t_{74}) and (t_{81} and t_{82}), are found for fumarase (P49) and for malate dehydrogenase (P53), respectively. The low percentage transitions (< 30%) are, in general, found in the centers of the various subreactions and involve enzyme-substrate complex formation.

There are many fewer high percentage transitions (> 60%) than low percentage transitions. The high percentage transitions involve the transitions in and out (t_{59} – t_{62}) from succinate (P38), in and out (t_{69} – t_{72}) from fumarate (P46), and in and out (t_{75} – t_{78}) of L-malate (P50). Note in Tables 5 and 6 that these are where the overlaps occur in the block-banded incidence matrix.

Whereas the lowest numbers of transitions correspond to enzyme complex formation, the highest numbers correspond to the release and use of the 4 carbon systems—succinate, fumarate, and malate. These involve substitutions of OH for H on succinate, as shown in Fig. 12. The reaction involving the next step, formation of a carbonyl from the alcohol to form oxaloacetate, is shown in Fig. 13.

Information on places. There are 59 places required to specify the Krebs cycle. The places, the count of their occurrences in the 15,893 minimal cycles, and the percentage of their occurrences in the 15,893 unique minimal cycles are presented in Table 8. The table is presented in descending order of occurrence. A bar plot for places 1 through 59 versus percentage of occurrence in the 15,893 minimal cycles is presented in Fig. 14. The graphs are presented in order of place number. Note that both the most frequently appearing places are of interest, as well as those that appear least frequently.

There are a number of places that have a high percentage of occurrences. The highest percentage is that associated with P51, which is the L-malate:malate dehydrogenase complex, followed by P29 and P7, which are the succinate:succinyl CoA synthetase complex and the citrate:CoA complex, respectively. The next highest percentages involve succinate and more of its complexes and fumarate and its complexes. These are consistent with the fact that the most transitions involve these species. An interesting place which shows up at a high percentage is the isocitrate:isocitrate dehydrogenase, P14.

An interesting species is the product CoA, P59. This product species occurs in 69% of the cycles. In addition, the nucleotide cofactor pair, NAD/NADH (P58/P57), which is involved in a number of enzyme complexes, occurs in almost 75% of the cycles, in contrast to the GDP/GTP (P34/P37) and FAD/FADH₂ (P42/P45) pairs, which occur in only one. This is not surprising as NAD is reduced to NADH three separate times in the Krebs cycle. Thus, species that are produced multiple times have a much higher probability of being in a cycle. It is interesting to speculate that the cycle could originally just have been producing NADH, but over time, the cell found that it could produce additional energetic species, FADH₂ and GTP, by small modifications of the cycle. The fact that GDP/GTP and FAD/FADH₂ appear so infrequently, as compared to the appearance of NAD/NADH in most cycles, is consistent with this conjecture.

However, the cross-coupling nucleotide cofactors—places P57 or nicotinamide adenine dinucleotide H (reduced form) [NADH], P58 or nicotinamide adenine dinucleotide (oxidized form) [NAD], and P59 or coenzyme A [CoA]—occur quite frequently at between 70% and 74%.

Interesting minimal two-cycles. There is a large separation between the places that occur least frequently, in fact only once in the listing of unique minimal cycles, and the remaining places. These five species—P1 (acetyl CoA), P34 (guanosine diphosphate [GDP]), P37 guanosine triphosphate [GTP], P42 (flavin adenine dinucleotide (oxidized form) [FAD]) and P45 (flavin adenine dinucleotide H₂ (reduced form) [FADH₂])—act as controllers in the Krebs cycle. They are sources (P1, P34, and P42) and sinks (P37 and P45).

Not surprisingly, the entry point, the acetyl-cofactor, coenzyme A, (P1) occurs only once and is the carrier of a high energy form of the pyruvate from glycolysis. This compound is converted by the cycle to CoA. The other species produced that occur only once are the two pairs of molecules involved in the reduction process, the GDP/GTP (P34/P37) couple and the FAD/FADH₂ (P42/P45) couple. This is not surprising, as the role of the cycle is to provide the energy to reduce FAD to FADH₂ only once and GTP is produced from GDP only once. Recall that complex formation is composed of 2 two-cycles involving both forward and backward pathways. Since the choice of starting-point and end-point is arbitrary, only 5 two-cycles are unique. The five two-cycles involving these five species are presented in Fig. 15. Note that solid arrows indicate forward paths and dashed arrows indicate backward paths.

TABLE 5. TRANSPOSE OF THE INCIDENCE MATRIX N^t OF THE KREBS CYCLE^a

^aThe $m = 86$ transitions, T1 through T86, label the rows and the $n = 59$ places, P1 through P59, label the columns. The incidence matrix has been color-coded to match the subreactions identified in Fig. 3.

TABLE 6. UPPER LEFT QUADRANT OF THE INCIDENCE MATRIX GIVEN IN TABLE 5^a

	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉	P ₁₀	P ₁₁	P ₁₂	P ₁₃	P ₁₄	P ₁₅	P ₁₆	P ₁₇	P ₁₈	P ₁₉	P ₂₀	...	P ₅₆	P ₅₇	P ₅₈	P ₅₉
t ₁	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₂	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₃	0	1	-1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₄	0	-1	1	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₅	0	0	-1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₆	0	0	1	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₇	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₈	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₉	0	0	0	-1	-1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₁₀	0	0	0	1	1	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₁₁	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
t ₁₂	0	0	0	0	0	0	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1
t ₁₃	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
t ₁₄	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
t ₁₅	0	0	0	0	0	0	0	0	1	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₁₆	0	0	0	0	0	0	0	0	-1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₁₇	0	0	0	0	0	0	0	0	0	1	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₁₈	0	0	0	0	0	0	0	0	0	-1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₁₉	0	0	0	0	0	0	0	0	1	-1	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0
t ₂₀	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
t ₂₁	0	0	0	0	0	0	0	0	0	0	0	0	-1	1	0	-1	0	0	0	0	0	0	0	0	0
t ₂₂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₂₃	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₂₄	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₂₅	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₂₆	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₂₇	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₂₈	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₂₉	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₃₀	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₃₁	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t ₃₂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

^aFrom top-left to bottom-right, presented are the incidence submatrix entries for Subreaction 1, the combined Subreactions 2/3, and Subreaction 4. Note the outlying cross-couplings with cofactor places P₅₇, P₅₈, and P₅₉. Place P₅₆ is shared with Subreaction 9 to complete the cycle.

TABLE 7. KREBS CYCLE TRANSITION COUNTS^a

Transition	No. Occur.	Percent Occur.	Transition	No. Occur.	Percent Occur.
t77, t78	10603	66.71	t35, t36	7655	48.16
t59, t60	10509	66.12	t27, t28	7653	48.15
t61, t62	10155	63.89	t49, t50	7617	47.92
t69, t70	10155	63.89	t1, t2	5170	32.53
t71, t72	10155	63.89	t81, t82	3834	24.12
t75, t76	10155	63.89	t39, t40	3808	23.96
t21, t22	9543	60.04	t29, t30	3790	23.84
t85, t86	9003	56.64	t25, t26	3748	23.58
t11, t12	8854	55.71	t53, t54	3741	23.53
t13, t14	8694	54.70	t57, t58	3741	23.53
t19, t20	8694	54.70	t47, t48	3740	23.53
t43, t44	8127	51.13	t51, t52	3740	23.53
t45, t46	8049	50.64	t55, t56	3740	23.53
t37, t38	7865	49.48	t63, t64	3387	21.31
t33, t34	7833	49.28	t67, t68	3387	21.31
t23, t24	7801	49.08	t65, t66	3386	21.30
t79, t80	7787	48.99	t73, t74	3386	21.30
t83, t84	7787	48.99	t15, t16	2899	18.24
t31, t32	7763	48.84	t17, t18	2899	18.24
t41, t42	7761	48.83	t5, t6	2586	16.27
t3, t4	7755	48.79	t7, t8	2586	16.27
t9, t10	7755	48.79			

^aThe paired transitions essentially split into two distinct categories: those between 48% and 67% and those between 16% and 33%.

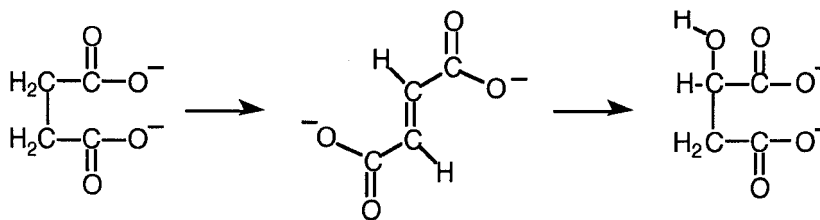


FIG. 12. The substitutions of OH for H on succinate.

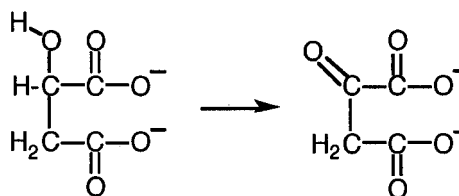


FIG. 13. The formation of a carbonyl from the alcohol to form oxaloacetate.

TABLE 8. KREBS CYCLE PLACE COUNTS^a

Place	No. Occur.	Percent Occur.	Place	No. Occur.	Percent Occur.
P51	14549	91.54	P59	11031	69.40
P29	14249	89.65	P3	10341	65.06
P7	14023	88.23	P4	10341	65.06
P39	13541	85.20	P2	10339	65.05
P40	13541	85.20	P56	10338	65.04
P47	13541	85.20	P27	8618	54.22
P48	13541	85.20	P20	8030	50.52
P38	13538	85.18	P53	7668	48.24
P46	13538	85.18	P23	7636	48.04
P50	13538	85.18	P19	7580	47.69
P14	13333	83.89	P16	7496	47.16
P52	12949	81.47	P35	7481	47.07
P22	11857	74.60	P36	7481	47.07
P28	11789	74.17	P30	7480	47.06
P57	11735	73.83	P32	7480	47.06
P21	11677	73.47	P33	7480	47.06
P25	11673	73.44	P43	6773	42.61
P58	11663	73.38	P44	6773	42.61
P54	11621	73.12	P41	6772	42.60
P55	11621	73.12	P49	6772	42.60
P9	11593	72.94	P11	5798	36.48
P10	11593	72.94	P12	5798	36.48
P8	11590	72.92	P5	5172	32.54
P13	11590	72.92	P6	5172	32.54
P26	11569	72.79	P1	1	0.006
P15	11553	72.69	P34	1	0.006
P17	11549	72.66	P37	1	0.006
P24	11425	71.88	P42	1	0.006
P18	11401	71.73	P45	1	0.006
P31	11357	71.45			

^aThe places essentially split into three categories: those between 65% and 92%, those between 33% and 54%, and the places which only appear in one cycle. The latter grouping contains the enzymes—P5, P11, P16, P23, P30, P41, P49, and P53.

The positive forward-only minimal cycles. Of particular interest are the positive (forward-paths-only) cycles, which are listed in Table 9. Of all of the 15,893 unique minimal cycles, there are but 11 forward-only positive cycles. They include the forward paths through the eight subreactions. The longest of these 11 positive cycles, with 43 places, is identified as the “backbone” of the Krebs cycle. These nine positive forward-only unique minimal cycles are somewhat obvious; however, it is important to note that the code was able to make their identification automatically. What is less obvious is that the code identified two additional forward-only cycles, which are variants of the forward paths associated with Subreaction 5 and the “backbone.” These variations are the result of the cross-coupling bypass provided by the nucleotide cofactor CoA.

Note that the five species identified in the discussion on “interesting” minimal two-cycles are not to be found in any of these 11 unique positive minimal cycles. Recall that these species—P1, P34, P37, P42, and P45—appear only once as two-cycles in the complete listing. While these species are critical to the functioning of the Krebs cycle, as sources (P1, P34, and P42, i.e., acetyl CoA, GDP and FAD) and sinks (P37 and P45, i.e., GTP and FADH₂), the Krebs cycle does not pass “through” them. They are species in complex formations. Some have forward paths into the species followed by a backward path out, or vice versa, but no forward path leads into them followed by a forward path out.

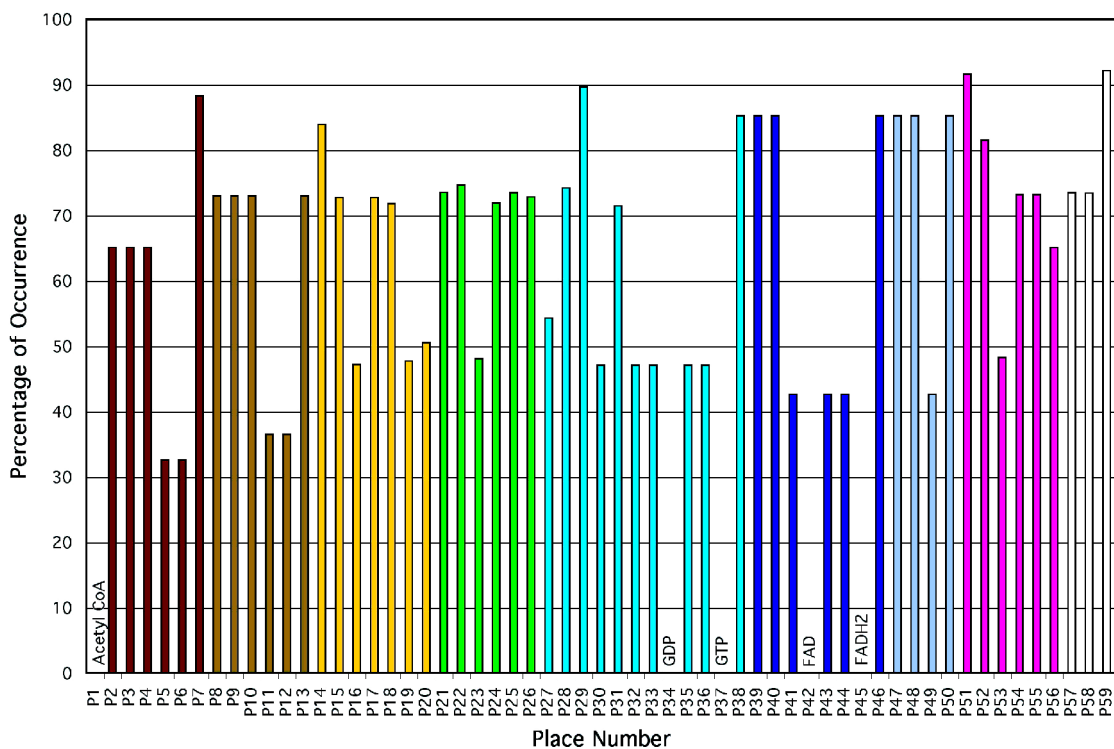


FIG. 11. Transition occurrence bar plot. Note the bipolar distribution.

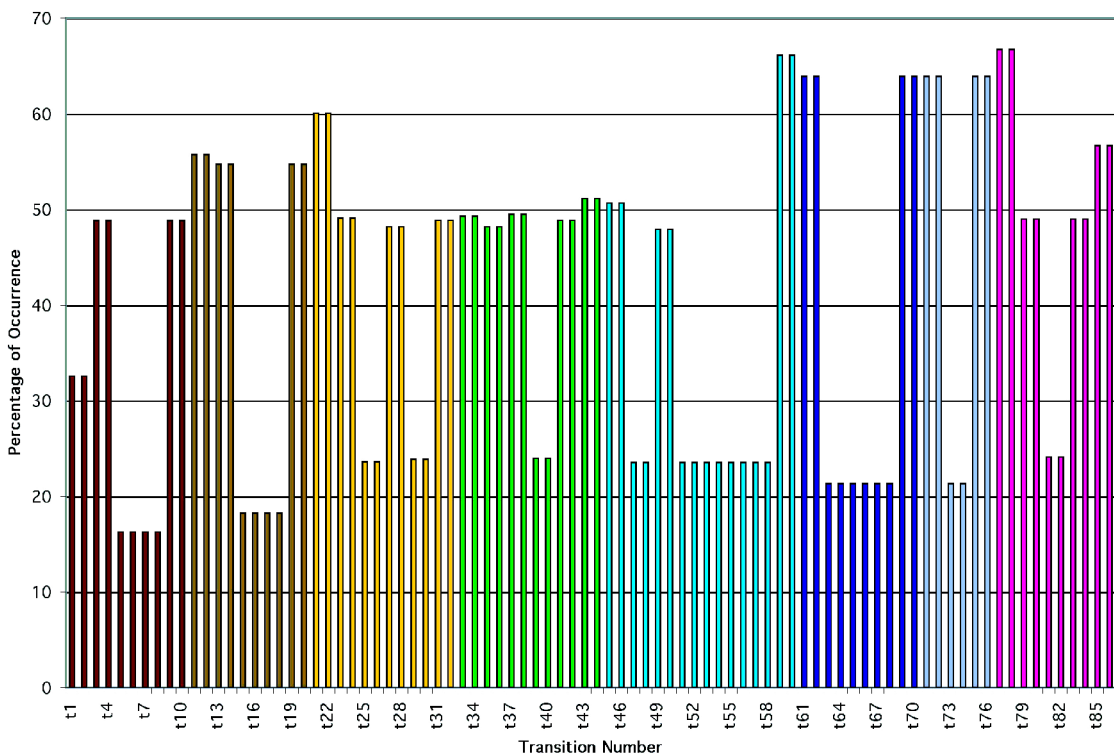


FIG. 14. Place occurrence bar plot.

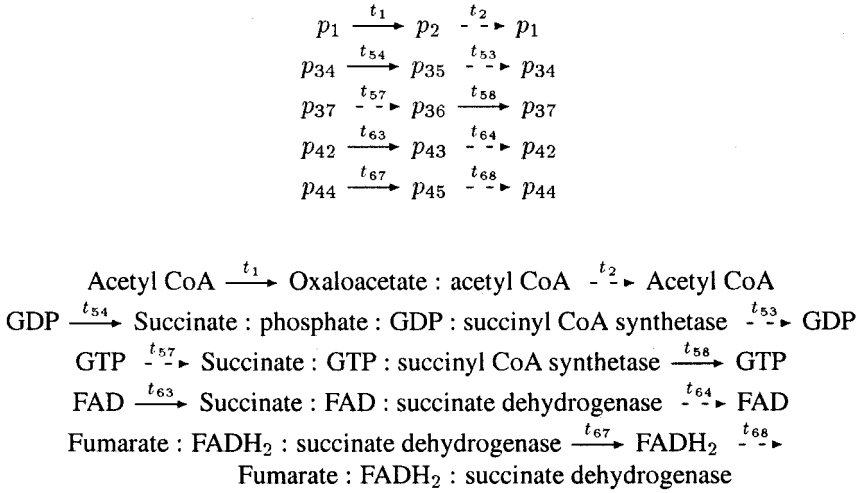


FIG. 15. The labeled pathway listing of the places which only appear in one unique minimal cycle each, specifically in two-cycles, listed first in terms of place numbers and repeated in terms of species.

Length distribution of minimal cycles. The number of cycles of a given length is represented by a Gaussian distribution with many cycles of considerable length, as shown in Fig. 16. This distribution is somewhat counterintuitive in that the distribution is skewed towards the longer length cycles rather than the shorter length cycles. The multitude of pathway options available to the functioning of the Krebs cycle is quite rich and provides ample alternatives to any pathway should such a path be inhibited.

4. CONCLUSIONS

The present study was undertaken to demonstrate that our hyperdigraph approach can be used effectively to provide insight into the behavior of a biochemical system. We demonstrate that this utility identifies biochemical pathways and circuits of principal significance. The TCA, or Krebs cycle, was selected because its reaction pathways are well known and have been extensively analyzed. The intention was to demonstrate that our mathematical procedures and software will identify that which is known as well as that which may not be known, completely and consistently. A database containing all of the 15,893 unique minimal cycles of the graph representation (which correspond to the unique minimal circuits of the network) was generated. Every pathway that passes through the Krebs cycle can be composed from circuits in the listing. This database can be used to answer hypotheses such as the implication of removing species from the network. Based on our solid mathematical foundation and with this demonstration of the validity of our approach, the algorithms can be used to analyze other important biochemical systems of current interest to the community.

We have analyzed this mathematical model to identify interesting mathematical features, which are related to interesting biochemical features. We have established a computational technique that can be used to reveal many features of biochemical networks. The Krebs cycle, because it is well known and extensively studied, demonstrates the ability of our methodology to identify patterns that reflect important biochemical events. As expected, the circuit constituting the canonical TCA cycle appears as one of the 15,893 minimal cycles/circuits identified by our technique and it is further identified as one of the 11 positive forward-only unique minimal cycles/circuits.

An unexpected result of this analysis is the Gaussian distribution of the 15,893 unique minimal cycles when plotted as a function of cycle length. One might have thought that there would be more characteristic cycles of shorter length than longer length and/or one might have expected that the number of cycles of a given length would drop off exponentially with increasing cycle length, but such was not the case. The richness of the Krebs cycle is certainly demonstrated in this observation. This critically important and robust cycle has embedded in it all of the requisite alternative pathways to ensure its proper function. It is a quite complex network, despite its simplicity.

TABLE 9. ELEVEN POSITIVE FORWARD CYCLES/CIRCUITS OF THE KREBS CYCLE^a

Cycle No.	Length	Cycle Specification	Cycle Subreaction Description
1	3	$p_{49} \xrightarrow{t_{71}} p_{47} \xrightarrow{t_{74}} p_{48} \xrightarrow{t_{75}} p_{49}$	p_{49} , Fumarase enzyme
2	4	$p_{11} \xrightarrow{t_{13}} p_9 \xrightarrow{t_{16}} p_{12} \xrightarrow{t_{17}} p_{10} \xrightarrow{t_{20}} p_{11}$	p_{11} , Aconitase enzyme
3	4	$p_5 \xrightarrow{t_4} p_3 \xrightarrow{t_5} p_6 \xrightarrow{t_8} p_4 \xrightarrow{t_9} p_5$	p_5 , Citrate synthase enzyme
4	5	$p_{39} \xrightarrow{t_{63}} p_{43} \xrightarrow{t_{66}} p_{44} \xrightarrow{t_{67}} p_{40} \xrightarrow{t_{70}} p_{41} \xrightarrow{t_{62}} p_{39}$	p_{41} , Succinate dehydrogenase enzyme
5	5	$p_{54} \xrightarrow{t_{81}} p_{55} \xrightarrow{t_{84}} p_{52} \xrightarrow{t_{80}} p_{53} \xrightarrow{t_{77}} p_{51} \xrightarrow{t_{80}} p_{54}$	p_{53} , Malate dehydrogenase enzyme
6	6	$p_{26} \xrightarrow{t_{42}} p_{23} \xrightarrow{t_{43}} p_{23} \xrightarrow{t_{34}} p_{21} \xrightarrow{t_{35}} p_{24} \xrightarrow{t_{38}} p_{25} \xrightarrow{t_{39}} p_{26}$	p_{23} , α -Ketoglutarate dehydrogenase complex enzyme
7	6	$p_{18} \xrightarrow{t_{28}} p_{19} \xrightarrow{t_{29}} p_{15} \xrightarrow{t_{32}} p_{16} \xrightarrow{t_{21}} p_{14} \xrightarrow{t_{24}} p_{17} \xrightarrow{t_{26}} p_{18}$	p_{16} , Isocitrate dehydrogenase enzyme
8	7	$p_{26} \xrightarrow{t_{42}} p_{28} \xrightarrow{t_{43}} p_{27} \xrightarrow{t_{46}} p_{28} \xrightarrow{t_{47}} p_{31} \xrightarrow{t_{50}} p_{59} \xrightarrow{t_{38}} p_{25} \xrightarrow{t_{39}} p_{26}$	Like cycle No. 6 with CoA bypass
9	8	$p_{30} \xrightarrow{t_{46}} p_{28} \xrightarrow{t_{47}} p_{31} \xrightarrow{t_{50}} p_{32} \xrightarrow{t_{51}} p_{33} \xrightarrow{t_{54}} p_{35} \xrightarrow{t_{55}} p_{36} \xrightarrow{t_{58}} p_{29} \xrightarrow{t_{59}} p_{30}$	p_{30} , Succinyl CoA synthetase enzyme
10	31	$p_{26} \xrightarrow{t_{42}} p_{22} \xrightarrow{t_{43}} p_{27} \xrightarrow{t_{46}} p_{28} \xrightarrow{t_{47}} p_{31} \xrightarrow{t_{50}} p_{32} \xrightarrow{t_{51}} p_{33} \xrightarrow{t_{54}} p_{35} \xrightarrow{t_{55}} p_{36}$ $p_{29} \xrightarrow{t_{58}} p_{38} \xrightarrow{t_{62}} p_{39} \xrightarrow{t_{63}} p_{43} \xrightarrow{t_{66}} p_{44} \xrightarrow{t_{67}} p_{40} \xrightarrow{t_{70}} p_{46} \xrightarrow{t_{71}} p_{47}$ $p_{74} \xrightarrow{t_{74}} p_{48} \xrightarrow{t_{75}} p_{50} \xrightarrow{t_{77}} p_{51} \xrightarrow{t_{80}} p_{54} \xrightarrow{t_{81}} p_{55} \xrightarrow{t_{84}} p_{52} \xrightarrow{t_{85}} p_{56} \xrightarrow{t_1} p_2$ $p_3 \xrightarrow{t_4} p_3 \xrightarrow{t_5} p_6 \xrightarrow{t_8} p_4 \xrightarrow{t_9} p_7 \xrightarrow{t_{12}} p_{59} \xrightarrow{t_{38}} p_{25} \xrightarrow{t_{39}} p_{26}$	“Backbone” with CoA bypass
11	43	$p_{18} \xrightarrow{t_{28}} p_{19} \xrightarrow{t_{29}} p_{15} \xrightarrow{t_{32}} p_{20} \xrightarrow{t_{34}} p_{21} \xrightarrow{t_{38}} p_{24} \xrightarrow{t_{38}} p_{25} \xrightarrow{t_{39}} p_{26} \xrightarrow{t_{42}} p_{22}$ $p_{27} \xrightarrow{t_{43}} p_{28} \xrightarrow{t_{46}} p_{31} \xrightarrow{t_{47}} p_{32} \xrightarrow{t_{50}} p_{33} \xrightarrow{t_{54}} p_{35} \xrightarrow{t_{55}} p_{36} \xrightarrow{t_{58}} p_{29}$ $p_{38} \xrightarrow{t_{62}} p_{39} \xrightarrow{t_{63}} p_{43} \xrightarrow{t_{66}} p_{44} \xrightarrow{t_{67}} p_{40} \xrightarrow{t_{70}} p_{46} \xrightarrow{t_{71}} p_{47} \xrightarrow{t_{74}} p_{48}$ $p_{75} \xrightarrow{t_{75}} p_{50} \xrightarrow{t_{77}} p_{51} \xrightarrow{t_{80}} p_{54} \xrightarrow{t_{81}} p_{55} \xrightarrow{t_{84}} p_{52} \xrightarrow{t_{85}} p_{56} \xrightarrow{t_1} p_2$ $p_5 \xrightarrow{t_5} p_6 \xrightarrow{t_8} p_4 \xrightarrow{t_9} p_7 \xrightarrow{t_{12}} p_8 \xrightarrow{t_{13}} p_9 \xrightarrow{t_{16}} p_{12} \xrightarrow{t_{17}} p_{10} \xrightarrow{t_{20}} p_{13}$ $p_{21} \xrightarrow{t_{21}} p_{14} \xrightarrow{t_{24}} p_{17} \xrightarrow{t_{26}} p_{18}$	“Backbone”

^aEight forward circuits are associated with the eight subreactions and three are associated with the Krebs cycle “backbone” and/or the CoA bypass.

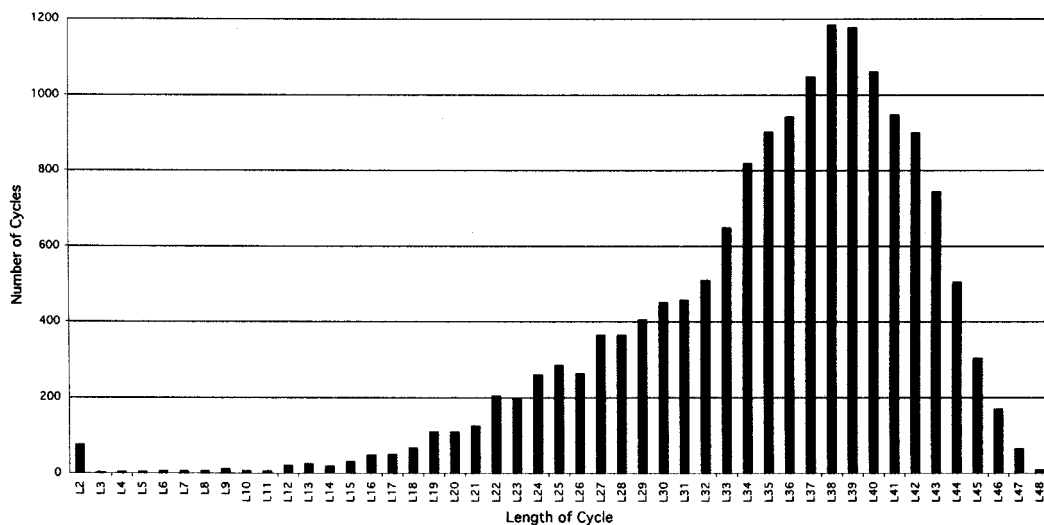


FIG. 16. Discrete distribution of the number of cycles of a given cycle length.

It is possible that the general topology of this Petri net and patterns of minimal circuits may also be found in other networks. This approach has the potential of providing us information about the evolutionary conservation of processes dealing with information flow in a cell. For example, the differences in the frequencies of appearance of GDP/GTP and FAD/FADH² as compared to NAD/NADH provide us with unique insights into how the cycle may have modified itself to optimize its output. The size of the calculation scales with the difference between the number of places and the number of transitions of the transformed graph. The number of computations that need to be performed is exponential in this difference. The number of transitions is always larger than the number of places (except in the case that the transformed graph is a tree).

This example suggests that the application of network models to less-well-characterized reaction sequences will assist in experimental efforts to elucidate their fine details and to suggest which experiments could provide the most information. The qualitative information that we obtain, concerning product species and cascade inductions, may be realized as hidden or suppressed subnetworks of biological significance. An example is the comparison of signaling mechanisms that are clearly homologous, such as those of TFG- α and epidermal growth factor. Network models can provide a framework to determine which mechanisms have a common evolutionary path in terms of how information is processed and may increase our knowledge of evolutionary divergence and speciation.

Finally, when we input equilibrium constants into the model, these will serve as weights in the Petri net and could change the significance of the key transitions. The weights might suggest which pathways amongst the many unique minimal cycles are more or less dominant and/or critical. Inclusion of rate constants will provide timing information, which could significantly change the distribution of transitions and how information flows due to timing delays. In fact, use of more detailed information about the rate constants and equilibrium constants will actually simplify the network analysis as it may cause some pathways to go to zero making the system less complex.

ACKNOWLEDGMENTS

This research was supported by the Laboratory Directed Research and Development program at the Pacific Northwest National Laboratory (PNNL) and performed in part in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) at PNNL. The EMSL is funded by the Office of Biological and Environmental Research in the U.S. Department of Energy under Contract DE-AC06-76RLO 1830. The authors would like to thank the ERULF program of the DoE for funding the undergraduate research of Dean W. Gull. The authors would like to thank Kenneth D. Jarman and Chad Scherrer for their internal inputs and the external reviewers for their invaluable review comments.

REFERENCES

- Alberty, R.A. 1991a. Chemical equations are actually matrix equations. *J. Chem. Ed.* 68, 984.
- Alberty, R.A. 1991b. Equilibrium compositions of solutions of biochemical species and heats of biochemical reactions. *Proc. Natl. Acad. Sci. USA* 88, 3268–3271.
- Alberty, R.A. 1992. Conversion of chemical equations to biochemical equations. *J. Chem. Ed.* 69, 493.
- Alberty, R.A. 1994. Review of biochemical thermodynamics. *Biochimica et Biophysica Acta (BBA)* 1207, 1–11.
- Alberty, R.A. 1996. Calculation of biochemical net reactions and pathways by using matrix operations. *Biophysical J.* 71, 507–515.
- Arkin, A., Ross, J., and McAdams, H.H. 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected *Escherichia coli* cells. *Genetics* 149, 1633–1648.
- Bailey, C.G., and Oliveira, J.S. 1998. An axiomatization for cubic algebras. In *Mathematical Essays in Honor of Gian-Carlo Rota*, Birkhäuser.
- Bailey, C.G., and Oliveira, J.S. 2001. MV-algebras and cubic algebras. Submitted to the *Annals of Combinatorics*.
- Bailey, C.G., and Oliveira, J.S. Another universal axiomatization of cubes. In preparation.
- Berge, C. 1973. *Graphs and Hypergraphs*, American Elsevier, New York.
- Björner, A., Las Vergnas, M., Sturmfels, B., White, N., and Ziegler, G. 1993. Oriented matroids. *CUP*.
- Clark, B.L. 1988. Stoichiometric network analysis. *Cell Biophys.* 12, 237–253.
- Darnell, J., Lodish, H., and Baltimore, D. 2000. *Molecular Cell Biology*, 2nd ed., Scientific American Books, W. H. Freeman, New York.
- Diestel, R. 2000. *Graph Theory*, 2nd ed., Springer, New York.
- Elliott, W.H., and Elliott, D.C. 1997. *Biochemistry and Molecular Biology*, Oxford University Press, New York.
- Garey, M.R., and Johnson, D.S. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Bell Labs., Murray Hill, NJ, W. H. Freeman, San Francisco, CA.
- Gillespie, D.T. 1979. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.* 22, 403–434.
- Henriksen, C.M., Christensen, L.H., Nielsen, J., and Villadsen, J. 1996. Growth energetics and metabolic fluxes in continuous cultures of penicillium chrysogenum. *J. Biotech.* 45, 149–164.
- Karp, P.D. 1998. EcoCyc: Encyclopedia of *Escherichia coli* genes and metabolism. *Nucl. Acids Res.* 26(1), 43–45.
- Kauffman, S.A. 1971a. Gene regulation networks: A theory for their global structure and behavior. *Current Topics in Dev. Biol.* 6, 145.
- Kauffman, S.A. 1993. *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, New York.
- Kohn, M.C., and Lemieux, D.R. 1991. Identification of regulatory properties of metabolic networks by graph theory. *J. Theor. Biol.* 150, 3–25.
- Kohn, M.C., and Letzkus, W.J. 1983. A graph-theoretical analysis of metabolic regulation. *J. Theor. Biol.* 100, 293–304.
- Mavrouniotis, M.L., and Stephanopoulos, G. 1990. Computer-aided synthesis of biochemical pathways. *Biotech. Bioeng.* 36, 1119–1132.
- McAdams, H.H., and Arkin, A. 1997. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA* 94, Biochem. 814–819.
- Metropolis, N., and Rota, G.-C. 1978. Combinatorial structure of the faces of the n-cube. *SIAM J. Appl. Math.* 35, 689–694.
- Narayanan, H. 1997. Submodular functions and electrical networks. In *Annals of Discrete Mathematics*, North-Holland, Elsevier, New York.
- Oliveira, J.S., Bailey, C.G., Jones-Oliveira, J.B., and Dixon, D.A. 2001. An algebraic-combinatorial model for the identification and mapping of biochemical pathways. *Bull. Math. Bio.* 63, 1163–1196.
- Oster, G.S., Perelson, A.S., and Katchalsky, A. 1973. Network thermodynamics: Dynamic modelling of biophysical systems. *Quarterly Rev. Biophys.* 6(1), 1–134.
- Papoutsakis, E., and Meyer, C. 1985. Equations and calculations of product yields and preferred pathways for butanediol and mixed-acid fermentations. *Biotech. Bioeng.* 27, 50–66.
- Peusner, L. 1986. Studies in network thermodynamics, vol. 5, *Studies in Modern Thermodynamics*, Elsevier.
- Pons, A., Dussap, C., Pequignot, C., and Gros, J. 1996. Metabolic flux distribution in *Corynebacterium melassecola* ATCC 17965 for various carbon sources. *Biotech. Bioeng.* 51, 177–189.
- Reddy, V.N., Liebman, M.N., and Mavrouniotis, M.L. 1996. Qualitative analysis of biochemical reaction systems. *Comput. Biol. Med.* 26, 9–24.
- Reisig, W. 1985. *Petri Nets, An Introduction*, Springer-Verlag, New York.
- Samoilov, M., Arkin, A., and Ross, J. 2001. On the deduction of chemical reaction pathways from measurements of time series of concentrations. *Am. Inst. Phys.* 11(1), 108–114.
- Schilling, C.H., and Palsson, B.O. 1998. The underlying pathway structure of biochemical networks. *Proc. Nat. Acad. Sci. USA* 95, 4193–4198.

- Schilling, C.H., and Palsson, B.O. 1999. Towards metabolic phenomics: Analysis of genomic data using flux balances. *Biotech. Prog., Am. Chem. Soc. and Am. Inst. of Chem. Eng.* 15, 288–295.
- Schilling, C.H., Schuster, S., Palsson, B.O., and Heinrich, R. 1999. Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era. *Biotech. Prog., Am. Chem. Soc. and Am. Inst. Chem. Eng.* 15, 296–303.
- Schnakenberg, J. 1979. Simple chemical reaction systems with limit cycle behavior. *J. Theoret. Biol.* 81, 389–400.
- Seressiotis, A., and Bailey, J.E. 1988. MPS: An artificially intelligent software system for the analysis and synthesis of metabolic pathways. *Biotech. Bioeng.* 31, 587–602.
- Smith, W.R., and Missen, R.W. 1982. *Chemical Reaction Equilibrium Analysis: Theory and Algorithms*, Wiley, New York.
- Stryer, L. 1982. *Biochemistry*, W. H. Freeman, New York.
- Weiss, T. 1996. *Cellular Biophysics: Transport*, vol. 1, MIT Press, Cambridge, MA.

Address correspondence to:

*J. B. Jones-Oliveira
International Technology Assessments Group
National Security and Technology Division
National Security Directorate
Pacific Northwest National Laboratory
P.O. Box 999, Msin K8-41
Richland, WA 99352*

E-mail: jjo@pnl.gov