J. theor. Biol. (2001) **208,** 361–382
doi:10.1006/jtbi.2000.2225, available online at http://www.idealibrary.com on IDEAL®

A New Method for Assembling Metabolic Networks, with Application to the Krebs Citric Acid Cycle

JAY E. MITTENTHAL*†, BERTRAND CLARKE‡, THOMAS G. WADDELL§ AND GLENN FAWCETT‖

*Department of Cell and Structural Biology, University of Illinois, 601 S. Goodwin St., Urbana, IL 61801, U.S.A., †Department of Statistics, University of British Columbia, Leonard S. Klink Building, Vancouver, BC, Canada V6T 1Z2, §Department of Chemistry, University of Tennessee at Chattanooga, Chattanooga, TN 37403, U.S.A. and ‖Caracal Consulting Inc., #106-1670 West 8th Ave, Vancouver, BC, Canada V6J 1V4

(*Received on* 19 *April* 2000, *Accepted in revised form on* 28 *October* 2000)

To understand why a molecular network has a particular connectivity one can generate an ensemble of alternative networks, all of which meet the same performance criteria as the real network. We have generated alternatives to the Krebs cycle, allowing group transfers and B_{12}-mediated shifts that were excluded in previous work. Our algorithm does not use a reaction list, but determines the reactants and products in generic reactions. It generates networks in order of increasing number of reaction steps. We find that alternatives to the Krebs cycle are very likely to be cycles. Many of the alternatives produce toxic or unstable compounds and use group transfer reactions, which have unfavorable consequences. Although alternatives are better than the Krebs cycle in some respects, the Krebs cycle has the most favorable combination of traits.

1. Introduction

To characterize a molecular network one specifies its topology—the connectivity among reactions—and its kinetics—the values of rate constants that govern the time-dependence of concentrations. To understand why a molecular network has a particular topology one can generate an ensemble of alternative networks that meet the same performance criteria as the real network but that have diverse topologies. The alternatives may be useful for exploring the evolution or the optimality of real networks, or for designing networks to new specifications. The alternatives may also be useful to experimenters for generating hypotheses about unknown aspects of a network under investigation. Here we present a new method for generating an ensemble of metabolic networks and use it to study the optimality of the Krebs citric acid cycle.

1.1. ALGORITHMS FOR ASSEMBLING NETWORKS

Algorithms to assemble networks from reactions have been proposed and used for metabolism and signal transduction (Mavrovouniotis *et al.*, 1990; Happel *et al.*, 1990; Mittenthal, 1996; Nuño *et al.*, 1997). These algorithms require explicit lists of allowed compounds and reactions. The algorithms can accomodate local constraints–restrictions on individual compounds or reactions. For example, a compound may be

† Author to whom correspondence should be addressed.

0022–5193/01/030361 + 22 $35.00/0

required, allowed, or excluded as an input, output, or intermediate of networks that the algorithm generates.

In addition to local constraints there may also be global constraints, such as minimizing the number of reaction steps in the network or the number of kinds of enzymes used. The preceding algorithms have to generate all possible networks that meet the local constraints, even if most of these networks are far from meeting the global constraints. So, it is desirable to develop algorithms that are responsive to global as well as local constraints. Furthermore, the compounds and reactions in molecular networks other than metabolism are often incompletely known. It would be useful to construct networks with generic reactions, in which the operation of the algorithm places constraints on reactants, products, and topology. With this motivation we have implemented a new algorithm that uses generic reactions rather than a reaction list, and that generates networks with fewer reaction steps earlier. We evaluate the alternative networks using performance criteria.

1.2. PERFORMANCE CRITERIA FOR THE KREBS CYCLE

We have used our method to generate alternatives to the Krebs cycle. Previous research has specified some performance criteria that the Krebs cycle probably meets. The number of reaction steps is likely to be minimal, because a network with fewer steps tends to use less of the genome's limited coding capacity (Kirkwood *et al.*, 1986), to use less of the limited concentration of proteins in the cytoplasm (Brown, 1991), and to have a greater flux for given concentrations of the input and output metabolites (Meléndez-Hevia *et al.*, 1994). Given these concentrations, the rate at which the network produces ATP (adenosine 5′-triphosphate) is likely to be maximal. Heinrich *et al.* (1997) and Meléndez-Hevia *et al.* (1997) have shown that in a linear pathway the rate of producing ATP is maximal if reactions near the beginning of the pathway are exergonic and those near the end are endergonic, as in glycolysis. This expectation seems reasonable for a cyclic pathway, taking the reaction joining the substrate to a feeder from the cycle as the first step in the pathway.

The Krebs cycle produces energy for cellular metabolism, but also makes carboxylic acids that are substrates for the biosynthesis of amino acids. In such a multifunctional pathway, reuse of reaction steps in meeting multiple constraints is an important criterion of performance. Reuse tends to reduce the number of steps needed to meet all the constraints, and increases the selection pressure for retaining steps.

1.3. ASSESSING THE OPTIMALITY OF METABOLIC PATHWAYS

To find the best alternatives in a large ensemble of alternative networks, there are several approaches. One can use evolutionary computation to mutate and select alternative networks, as has been done for glycolysis (Heinrich *et al.*, 1999; Stephani *et al.*, 1999). One can classify the possible networks, list the possible mechanisms for implementing each class of network, and evaluate the chemical feasibility and biological plausibility of each mechanism. This approach has been used to demonstrate the optimality of the pentose phosphate pathway (Meléndez-Hevia, 1990; Meléndez-Hevia *et al.*, 1994; Mittenthal *et al.*, 1998), glycolysis (Heinrich *et al.*, 1997; Meléndez-Hevia *et al.*, 1997; Stephani & Heinrich, 1998). Using classification and evaluation, Meléndez-Hevia *et al.* (1996) concluded that the Krebs cycle is optimal in having the fewest steps and the greatest yield of ATP. We have generated a large ensemble of alternatives to the Krebs cycle, expanding the set of allowed reactions to include reactions not admitted by Meléndez-Hevia *et al.* (1996)—rearrangements mediated by vitamin B_{12} and group transfer reactions.

2. Overview of the Method

Here we summarize our method. The appendix provides details.

2.1. CONSTRAINTS ON STOICHIOMETRY AND COMPOUNDS

The method constructs alternatives to a metabolic network that mediates specific overall reactions, which are expressed as stoichiometric constraints. We assume that any alternative to the Krebs cycle mediates three overall reactions:

(I) Convert pyruvate to three CO_2 molecules. That is, convert a three-carbon compound to three one-carbon compounds: $3 \rightarrow 1 + 1 + 1$. (II) Convert two pyruvates to 2-ketoglutarate and CO_2: $3 + 3 \rightarrow 5 + 1$. (III) Convert two pyruvates to oxalacetate and two CO_2: $3 + 3 \rightarrow 4 + 1 + 1$. Reactions I–III use pyruvate to produce energy, and to produce four- and five-carbon skeletons for the biosynthesis of amino acids.

2.2. GENERIC REACTIONS AND C-NETS

Our procedure does not use a list of reactions, but instead uses generic reactions (g-reactions) with two inputs and two outputs. The inputs and outputs are unspecified initially, but become increasingly characterized through the procedure. The rationale for using a g-reaction as the elementary process is that it represents a binary interaction. In general, any set of interactions among molecules is decomposable to a network of binary interactions.

A g-reaction may represent a cleavage; then it has one null input, so it converts one input to two outputs. Or, a g-reaction may represent an addition; then it has a null output, so it converts two inputs to one output. If both inputs and both outputs are non-null, the g-reaction is a group transfer reaction of the form

$$A + B\text{–}G \rightarrow A\text{–}G + B.$$

A network of g-reactions with specified connectivity and with the number of carbon atoms in each compound determined will be called a C-net. Figure 1 shows the C-nets for the portions of the Krebs cycle that meet constraints I–III. Note that the C-net for constraint I has a recurrent compound—a compound used early in the network that must be produced later or supplied exogenously. Each recurrent in a network forms a cycle.

2.3. CONSTRUCTING AN ENSEMBLE OF C-NETS FOR EACH STOICHIOMETRIC CONSTRAINT

The method constructs an ensemble of C-nets that are compatible with each stoichiometric constraint, as follows.

(1) Let N be the number of g-reactions. Increase N, starting with the smallest N that can meet the stoichiometric constraint. For each N, determine all possible connectivities among g-reactions.

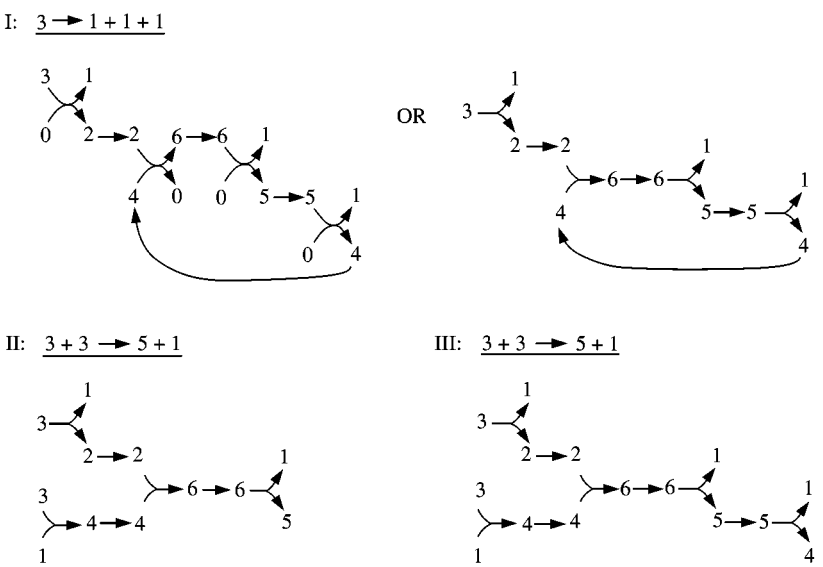


FIG. 1. C-nets from the Krebs cycle that meet stoichiometric constraints I–III. Generic reactions are shown explicitly for I (left), and in a simplified form for I (right), II, and III. The numbers represent the number of carbon atoms in compounds: **2** = acetate; **3** = pyruvate; **4** = succinate, fumarate, malate, or oxaloacetate; **5** = α-ketoglutarate; **6** = citrate or isocitrate. Recurrents for 1-carbon compounds are not shown in this and subsequent figures.

(2) For each connectivity, assign inputs and outputs from the stoichiometric constraint to particular g-reactions in all possible ways. Set all remaining inputs and outputs of g-reactions to zero.

(3) In each of the resulting networks, determine the number of carbon atoms in the inputs and outputs for each g-reaction.

(4) In the preceding steps, avoid the following conditions in any g-reaction:

(4A) The inputs must not be the same as the outputs, because such a g-reaction does not accomplish any transformation.

(4A1) One input must not be the same as one output, because the other input must then be the same as the other output.

(4A2) (0, 0) and (0, 1) are not allowed as inputs or outputs.

(4B) One input must not be greater than the sum of both outputs, or one output greater than the sum of both inputs. Such g-reactions would violate the conservation of carbon atoms.

(5) Remove all but one representative of each set of isomorphic C-nets. To define the isomorphism of two C-nets, let us regard a C-net as a set of g-reactions, each with its own label. Two C-nets are isomorphic if and only if the g-reactions in one of them can be relabeled to coincide with the g-reactions in the other.

(6) Remove futile cycles in C-nets. A futile cycle is a subset of the reactions of a connected C-net that has the null reaction, $0 \rightarrow 0$, as its net reaction. Appendix A, Section A.1.6 gives our method of finding futile cycles. Section 3.1 gives our rationale for removing futile cycles.

2.4. CONSTRUCTING AN ENSEMBLE OF PARANETS COMPATIBLE WITH ALL STOICHIOMETRIC CONSTRAINTS

The above procedure produces a sequence of C-nets for each stoichiometric constraint, ordered by increasing N. If there are several constraints, the next step is to produce a sequence of networks that meet all the constraints, again in order of increasing N. We do this by looking at the intersections of C-nets produced from the individual constraints. When we combine these C-nets the resulting networks are not C-nets, strictly speaking; we call them paranets. In a paranet, a given compound can be used in two or more different ways, depending on the constraint the paranet is meeting. Figure 2 shows the paranet for the Krebs cycle.

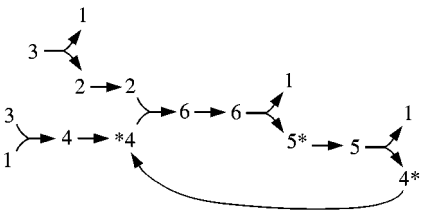


FIG. 2. Paranet for the Krebs cycle. Numbers designate compounds as in Fig. 1. An asterisk (∗) preceding a compound means that there are alternative sources for it; an asterisk following a compound signifies alternative sinks consuming it. The 4 with a ∗ preceding it may be synthesized from $3+1$, or obtained from a recurrent 4. The 4 and 5 followed by a ∗ may be used as shown, or pass into pathways that are not shown.

The paranets generated so far are C-paranets; they are assembled from C-nets, and they represent each compound only by the number of carbons in its skeleton. By hand, we elaborated the best C-paranets into realistic networks of biochemical compounds—R-paranets.

2.5. EVALUATING PARANETS

Criteria are needed for choosing which of the many C-paranets to convert to R-paranets. Previous studies of metabolic networks, cited above, suggest that better paranets will tend to have fewer g-reactions and to reuse g-reactions in meeting more than one stoichiometric constraint. Our measure of reuse in a paranet is

$$P = \frac{\text{total number of reactions in all networks of the paranet}}{(\text{number of constraints}) \times N}.$$

P ranges from 1/(number of constraints) if all networks of the paranet are disjoint, to one if all the networks are identical.

To choose C-paranets for conversion to R-paranets we used the quality $Q = N/P$, because favorable paranets have small N and/or large P. For a C-paranet the quality is $Q_C = N_C/P_C$. A better paranet has a smaller value of Q_C; it has fewer reactions and/or uses more reactions that help to meet at least two of the stoichiometric constraints.

2.6. CONVERTING A C- TO AN R-PARANET

The g-reactions of a C-paranet redistribute the carbon atoms in the reacting metabolites. Each of these metabolites is only specified by the number of carbon atoms it contains. Conversion of a C- to an R-paranet proceeds through the following stages, here as in our work on the pentose phosphate pathway (Mittenthal *et al.*, 1998).

(a) For each metabolite of every g-reaction in a C-paranet, specify the arrangement of carbon atoms as a carbon skeleton. In a paranet two metabolites with the same number of carbon atoms can be assigned different carbon skeletons.

(b) Each g-reaction then mediates a transformation of carbon skeletons. Choose an enzyme compatible with this transformation, and assign the functional groups required for operation of the enzyme to the relevant carbon atoms. If the same g-reaction occurs more than once in a paranet, different exemplars of it can use different enzymes and functional groups.

(c) Typically, additional reactions must then be added to the network. In general, a sequence of these connector steps converts the output of one g-reaction to the input of another or to a final product, without altering carbon skeletons. Thus, two or more metabolites with the same carbon skeleton can have different patterns of functional groups, in a sequence of connector steps.

(d) Some functional groups may not yet be specified at this stage. We assigned these functional groups so as to minimize the total number of functional group changes, compatible with the assignment of enzymes and with constraints on input and output metabolites of the R-paranet.

Through these stages we converted C- to R-paranets—networks of realistic reactions among metabolites with all functional groups specified. We used our knowledge of enzymes that change carbon skeletons by cleavage, addition, or group transfer reactions, and that modify functional groups without changing a carbon skeleton. In implementing this process for the Krebs cycle our aims were those of Meléndez-Hevia *et al.* (1996):

(1) Use pyruvate as the three-carbon input, CO_2 as the one-carbon output, oxalacetate as the four-carbon output, and 2-ketoglutarate as the five-carbon output.

(2) Use few steps: try to minimize the number of steps in the R-paranet, N_R, by the choice of enzymes for g-reactions and by minimizing the number of changes in functional groups between g-reactions. Note, in counting steps we regard a compound that does not leave the active site of an enzyme as occurring within a single step. For example, in the Krebs cycle, citrate → aconitate → isocitrate. Aconitate does not leave the active site; it is an intermediate in the mechanism of the aconitase reaction, which we treat as one step (Zubay, 1998, p. 257). In another example from the Krebs cycle, succinyl CoA → succinate involves succinyl acyl phosphate as an intermediate that does not leave the active site.

(3) Reactions should be plausible in terms of organic chemistry and biochemistry.

(4) Use fragments of known biochemical pathways to go between compounds, where possible.

(5) Avoid toxic, unstable, and reactive compounds. Toxic compounds included acetoin, formaldehyde, and acetaldehyde. Some compounds, such as 1,2 diketones, are unstable and may decompose to toxic compounds.

(6) Avoid compounds for which processing requires many steps. Where possible, avoid hydroxylations, which require NADH (nicotinamide adenine dinucleotide, reduced form) and so rob the cell of three ATPs, and which involve free radicals. (Compounds with methyl groups require many steps and hydroxylations, in pathways with oxaloacetate and α-ketoglutarate.)

We allowed two classes of reactions not admitted by Meléndez-Hevia *et al.* (1996):

(7) Allow use of vitamin B_{12} for 1,2-H,OH shifts and 1,2-H,COOH shifts.

(8) Allow group transfer reactions. These are organically feasible, and they occur in the metabolism of carbohydrates but not of carboxylic acids.

In cases where a compound had six or seven carbon atoms, many alternative choices for the topology of the carbon skeleton and the arrangement of functional groups were often possible. Our choices were reasonable but may not have been optimal, in terms of minimizing number of steps, maximizing stability of compounds, and so forth.

With these considerations, a C-paranet with N_C g-reactions and metric P_C produces a most favorable R-paranet with N_R g-reactions, reuse index P_R, and quality $Q_R = N_R/P_R$.

3. Results

With a computer program that implemented the algorithms in Appendix A, we initially obtained a list of the best 100 C-paranets that had five or fewer g-reactions and six or fewer carbon atoms per compound. Many of the resulting R-paranets had unfavorable properties. Consequently, we excluded some classes of C-paranets, as discussed below, and examined the best allowed 100 C-paranets.

3.1. CLASSES OF PARANETS EXCLUDED

A recurrent compound is a compound generated within a metabolic network but required in an earlier reaction to run the network. The available concentration of a recurrent limits the flux through the network. A network is less likely to evolve, the more recurrents it requires, because an adequate source for each recurrent must evolve with the network. We recognized recurrents using the algorithm described in Appendix A, and discarded C-paranets with two or more recurrents.

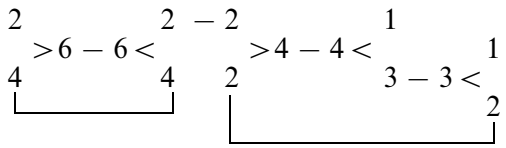
Many of the best 100 paranets had two or more group transfer reactions. Group transfer reactions within monosaccharides, such as transaldolase and transketolase mediate, are very easy and convenient. However, similar reactions on non-carbohydrates are clumsy at best, and they lead to very inconvenient structures and longer pathways. In this work, some group transfer g-reactions could not be converted to reasonable biochemical reactions, given reasonable substrates. Among these group transfer g-reactions are $3+1 \rightarrow 2+2$, $2+2 \rightarrow 3+1$, $5+3 \rightarrow 4+4$, and $5+1 \rightarrow 4+2$. Some group transfer reactions produce compounds that are toxic or unstable, or that have methyl groups. Some produce compounds with difficult-to-manipulate structures or with branching structures, elimination of which requires multistep rearrangements. Furthermore, group transfers are not oxidations, and so will not yield energy. In view of these drawbacks, we ignored all C-paranets with two or more group transfer reactions. The majority of such C-paranets in the best 100 also had two or more recurrents.

Futile cycles can be added to C-nets, and thus to paranets, in an infinite variety of ways. We eliminated futile cycles to keep the results manageable. For the constraints of six or fewer g-reactions, six or fewer carbon atoms per compound, 0 or 1 recurrents, and 0 or 1 group transfer reactions, eliminating futile cycles reduced the estimated number of C-paranets to be generated by a factor of 470, from 1.2×10^{10} to 2.5×10^{7}.

Because we excluded futile cycles, the forward and reverse reactions of a reversible reaction are not both represented explicitly within any of the C-nets that we generated. However, a reaction in an R-net made from a C-net can be reversible. The reversible reaction runs in either the forward or the reverse direction, as the concentrations of reactants determine through mass action. Operating in this way, according to thermodynamics, a reversible reaction is not a futile cycle in which one reaction undoes what another does.

Because we ignored C-nets with futile cycles, we did not generate some R-nets that lack futile cycles. For example, Meléndez-Hevia *et al.* (1996) presented an alternative to the Krebs cycle in their Fig. 6. The C-paranet for this alternative has

a futile cycle:



So, our computation did not generate this C-net.

3.2. THE BEST 100 R-PARANETS WITH ZERO OR ONE GROUP TRANSFER REACTION AND ZERO OR ONE RECURRENT

After we discarded all C-paranets with two or more group transfer reactions or with two or more recurrents, there were 4.5×10^7 allowed C-paranets. All used six or fewer g-reactions, and used compounds with seven or fewer carbon atoms. We examined the best 100 of these allowed C-paranets. Among the first 50 the poorer ones, #12–#50, had one group transfer reaction. Five of these 39 C-paranets could not be converted to R-paranets because they had an unsuitable group transfer reaction. Four of the 39 had no recurrent, and so were not cycles. Of these four, two could not be converted, one produced an unstable 2,3 diketone, and one produced acetalydehyde. For the 34 (out of 39) C-paranets that could be converted, the best R-paranet had 11 or more steps (in some cases, more than 20), and $Q_R > 14.4$. Many of these R-paranets produced toxic compounds and/or had a poor energy yield. The poor quality of #12–#50, which had a group transfer reaction, led us to ignore the remaining 49 paranets with a group transfer reaction, which had even larger Q_C. These were #51–#81 with six g-reactions, and #83–#100 with five g-reactions.

Thus, we focus on C-paranets #1–#11 and #82, which are shown in Fig. 3. Paranet #8 is the Krebs cycle. Paranets #1–#5, #7, and #9–#11 resemble the Krebs cycle in adding two compounds to give a larger one, and then cleaving 1- or 2-carbon fragments in a sequence of reactions. Each of the four C-paranets with one group transfer reaction (#1–#4) is a condensed version of a C-paranet with no group transfer reaction (#7, #8, #11, #9, respectively). #2 is the condensed form of the Krebs cycle; the pair of reactions $4 + 2 \rightarrow 6$ and $6 \rightarrow 5 + 1$ is condensed to the group transfer reaction $4 + 2 \rightarrow 5 + 1$. Such condensation reduces Q_C, and so improves the ranking of the condensed C-paranet relative to its uncondensed partner. (Appendix A Section 3 provides a proof of this statement.) However, such improvement does not occur when condensation deletes a 5-carbon compound that met constraint II in the uncondensed paranet, as is the case for #5 and #10 in Fig. 3. An additional reaction then has to be added to the condensed paranet to meet this constraint, thereby increasing its Q_C.

For paranets #1–#11 and #82, Table 1 summarizes the characteristics N_C, N_R, P_C, P_R, Q_C, and Q_R. The step ratio N_R/N_C is the average number of steps in the R-paranet needed to implement a g-reaction and to make the transition to a neighboring g-reaction. Table 1 also summarizes the energy yield for the R-paranets, measured as the number of molecules produced or consumed throughout an R-paranet, for $FADH_2$ (flavin adenine dinucleotide, reduced form), NADH, ATP, and GTP (guanosine 5′-triphosphate).

The only R-paranets with Q_R less than or equal to 14.4 are #2, #5, #7, #9, #82, and the Krebs cycle, #8. These competitors equal or exceed the Krebs cycle in some attributes, but none is as good overall. All five of these R-paranets join two smaller compounds to make a larger one that is converted to 2-ketoglutarate by reactions that include a B_{12}-mediated rearrangement.

- #2 (Fig. 4) is the condensed form of the Krebs cycle. It illustrates the reduction in performance that typically accompanies condensation. Cleavage of pyruvate produces the toxic chemical acetaldehyde, which is used in a group transfer reaction to produce a five-carbon compound. #2 has 11 steps and a step ratio of 2.75, despite the condensation from 5 to 4 g-reactions. Its $Q_R = 13.4$, and it has the same energy yield as the Krebs cycle.
- #5 (Fig. 5) resembles the Krebs cycle, but condenses acetyl CoA with pyruvate rather than oxaloacetate. Note that the aconitase-like reaction is counted as one step. #5 has 11 steps, a step ratio of 2.75 and $Q_R = 14.0$. Its energy yield (1 $FADH_2$, 4 NADH, 0 ATP, 1 GTP) is

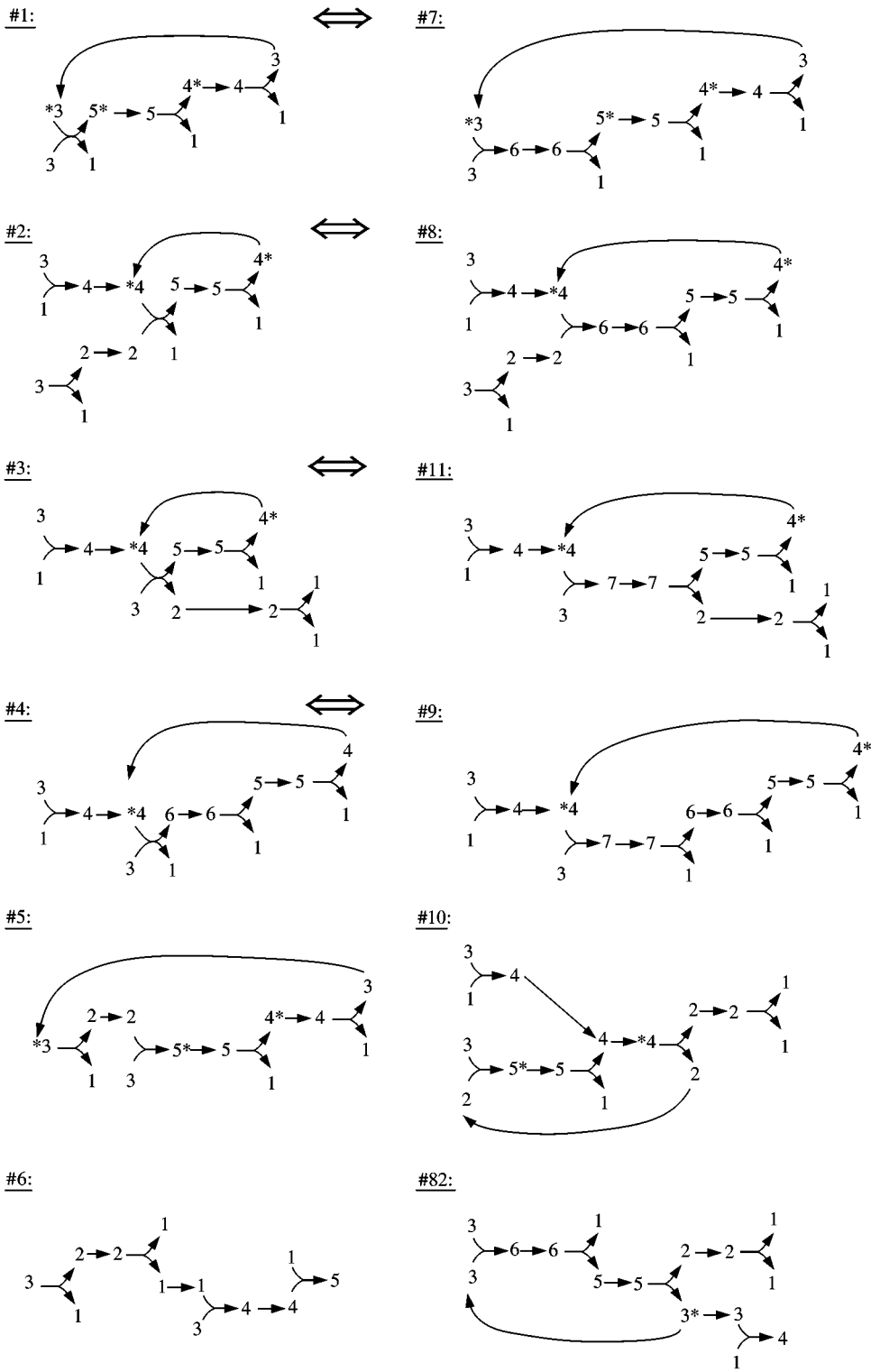


FIG. 3. Twelve C-paranets with zero or one group transfer reaction and zero or one recurrent. The numbers in the g-reactions represent the number of carbon atoms in compounds, but do not stand for particular compounds. The # labels (e.g. #5) designate the ranking of a C-paranet according to its value of N_C/P_C. The symbol ⇔ designates a pair of C-paranets in which the left-hand one is a condensed version of the right-hand one.

TABLE 1

Properties of the best 11 *paranets and* #82. *The paranets are listed in the order in which they are displayed in Fig.* 3. *For a C-paranet,* N_C *is the number of g-reactions and* P_C *is the performance metric. The most favorable R-paranet from that C-paranet has* N_R *g-reactions and metric* P_R. $Q_C = N_C/P_C$; $Q_R = N_R/P_R$.

Paranet #	N_C	N_R	N_R/N_C	P_C	P_R	Q_C	Q_R	$FADH_2$	NADH	ATP	GTP	Chemistry
Four pairs of the best C-paranets												
#1	3	14	4.7	0.67	0.57	4.5	24.5	1	4	0	1	Acetoin; B_{12}
#7	4	10	2.5	0.75	0.77	5.3	13.0	1	4	0	1	Poor K_{eq}; B_{12}
#2	4	11	2.8	0.83	0.82	4.8	13.4	1	4	−1	1	Acetaldehyde; B_{12}
#8	5	10	2.0	0.87	0.8	5.8	12.5	1	4	−1	1	Krebs cycle
#3	4	18	4.5	0.83	0.89	4.8	20.2	1	3	−2	0	Acetaldehyde
#11	5	18	3.6	0.87	0.89	5.8	20.2	1	3	−2	0	
#4	4	12	3.0	0.83	0.81	4.8	14.9	1	4	−1	2	B_{12}
#9	5	11	2.2	0.87	0.82	5.8	13.4	1	4	−1	2	Poor K_{eq}
Condensed form lacks a 5 *to satisfy constraint II*												
#5	4	11	2.8	0.75	0.79	5.3	14.0	1	4	−1	2	Krebs-like; B_{12}
#10	5	18	3.6	0.87	0.91	5.8	19.8	1	4	−1	0	Glyoxylic acid
No recurrents												
#6	4	11	2.8	0.75	0.73	5.3	15.1	0	1	−2	0	Carboxylations
The only acceptable paranet among #50–#100												
#82	5	9	1.8	0.73	0.78	6.8	11.6	0	3	−2	0	Poor K_{eq}; B_{12}

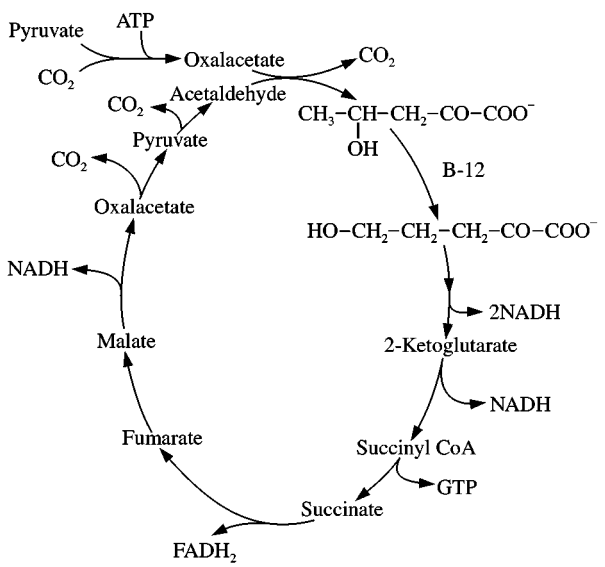


FIG. 4. R-paranet #2. In Figs 4–8, if a sequence of reactions is the same as in the Krebs cycle, an arrow designates each reaction, but only the first and last compounds in the sequence are named or displayed.

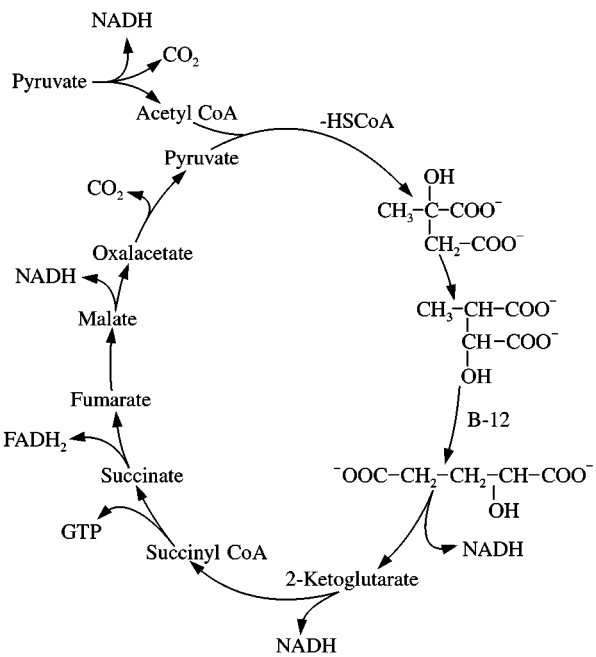


FIG. 5. R-paranet #5.

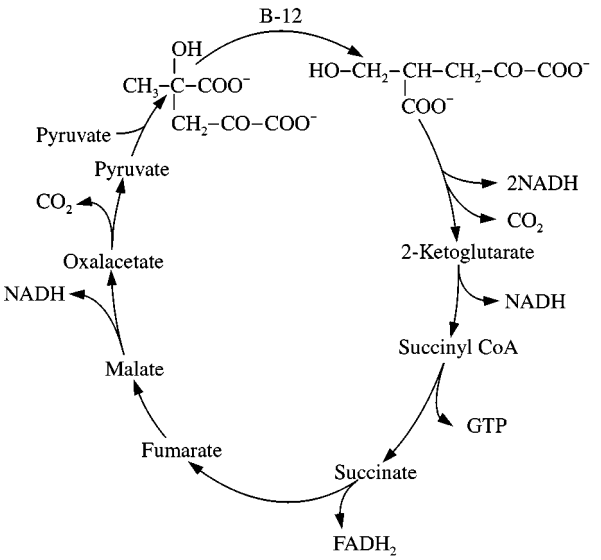


FIG. 6. R-paranet #7.

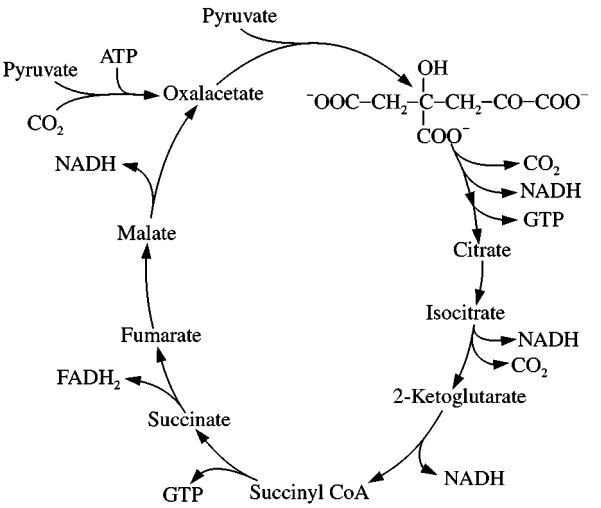


FIG. 7 R-paranet #9.

better than that of the Krebs cycle, in that #5 does not use an ATP to carboxylate pyruvate to oxaloacetate.

#7, #9, and #82 begin with an aldol addition that has an unfavorable equilibrium constant.

- #7 (Fig. 6) has an aldol addition of two pyruvates to make a 6-carbon compound, which undergoes rearrangement and decarboxylation to make 2-ketoglutarate. Its condensed form is #1. #7 has ten steps, as does the Krebs cycle, but it has four g-reactions and a step ratio of 2.5. Its $Q_R = 13.0$, and its energy yield is better than the Krebs cycle, as in #5.
- #9 (Fig. 7) has an aldol addition of pyruvate and oxaloacetate to make a 7-carbon compound, which undergoes an aconitase-like rearrangement and two decarboxylations to make 2-ketoglutarate. Its condensed form is #4. #9 has 11 steps, five g-reactions, a step ratio of 2.2, $Q_R = 13.4$, and the same energy yield as the Krebs cycle.
- #82 (Fig. 8), as #7, converts two pyruvates to 2-ketoglutarate. The five-carbon skeleton is hydroxylated, oxidized, and cleaved to pyruvate and oxalic acid—a very toxic compound. Its condensed form is #13. #82 has nine steps, five g-reactions, a step ratio of 1.8, and $Q_R = 11.6$—parameters better than the Krebs cycle. However, its energy yield (3 NADH, −2 ATP, no $FADH_2$ or GTP) is poorer than that of the Krebs cycle.

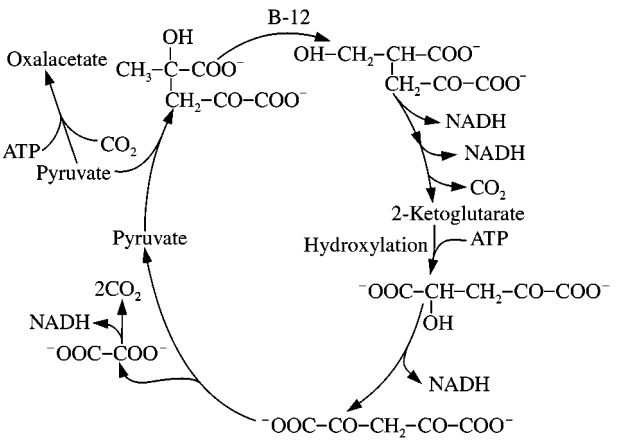


FIG. 8 R-paranet #82.

The other six of the best 11 R-paranets (#1, 3, 4, 6, 10, 11) have 12 or more steps, $Q_R > 14.4$, and a step ratio of at least 2.4. Although #1 is the smallest C-paranet, and the only one with three g-reactions, its R-paranet requires 14 steps and its $Q_R = 24.5$, reflecting a low overlap in the usage of reactions for meeting the three constraints ($P_R = 0.57$). #6 has no recurrents. It requires two carboxylations, which use energy.

3.2. ARE THERE OTHER GOOD ALTERNATIVES TO THE KREBS CYCLE?

In focusing on the best 100 of the 4.5×10^7 allowed C-paranets, we obviously selected a small sample. Might there be good alternatives to the Krebs cycle among the remaining paranets? Several arguments suggest that this possibility is very unlikely. We can estimate an upper bound on Q_C for good alternatives, using the relation

$$Q_C = Q_R(N_C/N_R)\,(P_R/P_C).$$

Potentially good alternative R-paranets have $Q_R \leqslant 14.4$. The maximum value for N_C/N_R is found from the minimum of the step ratio, N_R/N_C, which is unlikely to be less than its value in #82, 1.8. The maximum of P_R/P_C is found by taking the maximum value for P_R, 1, and the minimum value for P_C, 1/3. Thus, the largest possible value for P_R/P_C is 3. Therefore, a good alternative is unlikely to occur among the many C-paranets with Q_C greater than 24 ($14.4 \times 3/1.8$). A realistic upper bound on Q_C for good alternatives is probably much lower.

It is also unlikely that a good alternative has more than five g-reactions. A paranet with $N_C = 6$ is likely to have $N_R > 10.8$, because the minimum step ratio that we observed is 1.8. If $N_R > 10.8$, P_R would have to be >0.75 to get $Q_R < 14.4$. A step ratio as small as 1.8 is extremely rare, and a P_R as large as 0.75 occurs infrequently except among the paranets in Table 1. By the same argument, a paranet with $N_C \geqslant 7$ is likely to have $N_R > 12.6$, and so is unlikely to be a good alternative.

Among the paranets with five or fewer g-reactions, the paranets with one group transfer reaction are less likely to be good alternatives than those with no group transfer reactions. Recall the adverse effect of a group transfer reaction in the best 50 allowed paranets: #12–#50 all had a group transfer reaction, and none was a good alternative. #1–#4 had group transfer reactions, and were condensed versions of paranets without group transfer reactions. In these four pairs of reactions, the condensed paranet has a $Q_R \geqslant$ the Q_R for the uncondensed paranet.

These considerations led us to examine the best 50 paranets of the 6×10^4 with no group transfer reactions, six or fewer g-reactions, and seven or fewer carbon atoms per compound. Among these, the best eight are #5–#11 and #82 discussed above. The next five paranets had four or five g-reactions; with Q_R's ranging from 19.5 to 33.4 they are not good alternatives. The next 34 paranets had six g-reactions and so were not considered. The remaining three paranets had five g-reactions; with Q_R's ranging from 22.0 to 37.4 they are not good alternatives. These arguments and results strengthen our conclusions that the only alternatives to the Krebs cycle worth considering are #2, 5, 7, 9 and 82. The Krebs cycle is better overall than any of these paranets.

4. Discussion

4.1. IS THE KREBS CYCLE OPTIMAL?

We have examined favorable alternatives to the Krebs cycle and found none that surpasses it in overall performance, though some are better in certain respects. The Krebs cycle uses no group transfer reactions, and so avoids their unfavorable associations. It does not use vitamin B_{12}, and it produces no toxic compounds, unlike some favorable alternatives. With the exception of #82, the Krebs cycle has the minimum number of steps (ten), the best value for the quality index N_R/P_R (12.5), and the lowest value for the step ratio N_R/N_C (2.0). With respect to these parameters the Krebs cycle is better than its closest competitor, #5 (Fig. 5). #5 resembles the Krebs cycle, but adds acetyl CoA to pyruvate rather than oxalacetate. #5 has a slightly better stoichiometric energy yield than the Krebs cycle, but most alternatives have a poorer yield. The profile of free energy changes along the pathway will favor a greater rate of ATP production in the Krebs cycle than in many alternatives. In the following sections, we consider some of these issues in more detail.

4.1.1. *Economy*

The Krebs cycle performs well in reusing steps to meet different constraints; its reuse index, P_R, is relatively high. This reuse is a form of economy.

Economy is also evident in the step ratio, which measures the number of steps that change functional groups without changing carbon skeletons, per step that changes carbon skeletons. For the Krebs cycle the step ratio is 2.0, nearly the minimum among the alternatives we examined. This minimality also obtains for the non-oxidative pentose phosphate pathway, for which Mittenthal *et al.* (1998) generated alternatives. Values of N_R/N_C for the most favorable alternatives to that pathway can be calculated from their Table 4, where the first column gives N_C and the last gives N_R. Among those alternatives the real pentose phosphate pathway has the minimum step ratio, $13/7 = 1.86$. Thus, the step ratio is very low in the Krebs cycle and the pentose phosphate pathway, relative to alternatives that meet the same constraints. This may be the case in other metabolic pathways.

Note that the ratio N_R/N_C overestimates the step ratio. We have regarded a change in the number of carbon atoms in a compound as a change in its carbon skeleton, but we have neglected the rearrangements of carbon atoms that occur in the Krebs cycle and in many alternatives to it. Directly counting, in an R-paranet, the number of steps that change functional groups without changing carbon skeletons, per step that changes carbon skeletons, will give a lower number than N_R/N_C if carbon rearrangements are included among the skeleton-changing steps.

4.1.2. *Constraints and Cycles*

We imposed three stoichiometric constraints on alternatives to the Krebs cycle, requiring that they produce energy and carboxylic acids that can be converted to amino acids. However, the Krebs cycle is embedded in a larger network of reactions. Glycolysis, anaplerosis, and catabolism of fatty acids and amino acids provide fuel for the Krebs cycle. The Krebs cycle provides precursors for the synthesis of carbohydrates, amino acids, purine and pyrimidine nucleotides, and porphyrins. We limited our investigation of alternatives to the Krebs cycle in order to address a manageable problem. By requiring pyruvate, oxalacetate, and 2-ketoglutarate we have taken many of the fueling reactions into account, since these α keto-acids link directly to the corresponding amino acids and to glycolysis. It was unclear how to formulate analogous constraints for some of the larger networks that include the Krebs cycle.

Evidently, the Krebs cycle is the hub for a hub-and-spoke network that can interconvert several key metabolites. Pathways between the hub and the key metabolites are the spokes; some of these are partially joined to form branching pathways. We expect hub-and-spoke organization in a network that interconverts several key metabolites because this organization is optimal in a simplified model of metabolism (Mittenthal *et al.*, 1993). It seems plausible that a hub for the biosynthesis and degradation of several key metabolites might contain a cycle of reactions. However, for the three constraints we used, we found non-cyclic alternatives. Baldwin & Krebs (1981) suggested that a cycle is a more efficient way to oxidize acetate while extracting energy than non-cyclic alternatives. They considered only alternative networks involving glycollate. In agreement with their proposal, the non-cyclic alternatives that we found perform less well than the Krebs cycle.

4.1.3. *Aldol Addition*

In a linear pathway the rate of ATP production is maximal if reactions near the beginning of the pathway are exergonic, promoting flux in the forward direction, and reactions near the end are endergonic (Heinrich *et al.*, 1997; Meléndez-Hevia *et al.*, 1997). The Krebs cycle and alternative #5 have this pattern, but other alternatives that we examined do not. As Heinrich *et al.* (1997) noted, two exergonic steps occur early in the Krebs cycle—an aldol addition coupled to thioester hydrolysis in the citrate–synthase reaction, and the isocitrate–dehydrogenase reaction. R-paranet #5 also uses an aldol addition coupled to irreversible thioester hydrolysis. However, #7, #9, and #82 do not have early exergonic steps; these three alternatives all begin with an aldol addition that has an unfavorable equilibrium. The latter steps in the Krebs cycle and in #5, #7, and #9—succinyl CoA synthetase, succinic dehydrogenase, furnarase, and malate dehydrogenase—each have ΔG in the cell near zero (Garrett & Grisham, 1995).

4.1.4. *Vitamin B_{12}*

In the Krebs cycle, as in some of the best alternatives, an H,OH shift follows an aldol addition. Aconitase catalyses an H,OH shift in the Krebs cycle and in #9, whereas B_{12} mediates the shift in other good alternatives. In alternative #5 shifts occur in both ways: an aconitase-like reaction makes an H,OH shift, and B_{12} then mediates an H,COO^- shift. The differences between these processes of rearrangement are striking. An aconitase-like reaction does an easy dehydration, involving an acidic alpha proton, to give a stable, conjugated ene dione. Then, a favorable Michael addition of water to the C=C gives a resonance stabilized anion, all at the active site. (We treat this process as one step.) No coenzyme is required. By contrast, in a B_{12} shift the complicated B_{12} coenzyme generates free radical intermediates and cobalt complexes of different oxidation levels.

Free radical intermediates are acceptable for anaerobes with short life cycles, but are less tolerable for aerobes with longer life cycles. Most of the roughly 15 reactions known to require B_{12} occur in a few bacterial species and perform specialized fermentations. B_{12} shifts do not occur in primary energy-producing metabolism. In mammalian metabolism, B_{12} is used to a significant extent only in the oxidation of odd-C fatty acids, which are rare, and for synthesizing methionine from homocysteine (Mathews *et al*., 2000).

4.1.5. *Evolutionary Considerations*

The Krebs cycle may have evolved in anaerobic prokaryotes as two branches. The oxidative branch from pyruvate to 2-ketoglutarate could be used for biosynthesis of glutamate. The reductive branch from oxalacetate to succinate would have oxidized NADH produced in glycolysis, regenerating NAD^+ without trapping pyruvate as lactate, and would have provided succinate for biosyntheses. In this scenario, 2-ketoglutarate dehydrogenase evolved later, allowing transformation of 2-ketoglutarate to succinyl-CoA and producing an oxidative cycle. With the evolution of photosynthesis, photoreduced compounds were available to drive a reductive citric acid cycle that fixes CO_2 (Weitzman, 1985; Gest, 1987; Buchanan & Arnon, 1990). An alternative scenario posits the early evolution of a reductive citric acid cycle and its later cooption for oxidation (Wächtershäuser, 1990; Morowitz, 1999).

The best alternative paranets that we found—#2, #5, #7, and #9—could have evolved as two branches because they have the same kind of design as the Krebs cycle. In these paranets the reactions in the latter half of the cycle could have evolved from reduction of oxalacetate to succinate, and the reactions in the earlier half oxidize pyruvate to 2-ketoglutarate. However, #82 could not have evolved in an analogous way, because its latter half could not have reduced oxaloacetate to succinate. Moreover, it produces oxalic acid, which is very toxic. #82 is a poor competitor; it has an initial aldol addition with an unfavorable equilibrium, a B_{12} shift, an energy-robbing hydroxylation, and a poor energy yield.

Historically, might networks that used B_{12} have been competitors for the Krebs cycle? Morowitz (1999) proposed that metabolism evolved through the sequential addition of shells to a core (shell A) which consisted of the Krebs cycle, glycolysis, and fatty acid synthesis. The amination of 2-ketoglutarate was the gateway to shell B, the synthesis of most amino acids. In shell C, sulfur was incorporated into cysteine and methionine. The gateways to shell D, ring closure and synthesis of nitrogen and dinitrogen heterocycles, gave access to purines, pyrimidines, and many cofactors, including B_{12}. This scenario suggests that compounds in shell D evolved after enzymes (derived from shell B) and were not a part of pre-biotic chemistry. However, Ksander *et al.* (1987) showed that a pre-biotic synthesis of a corrin template for B_{12} synthesis is feasible, by synthesizing uroporphyrinogen III from glutamine nitrile under anaerobic conditions. Methanogenic archaebacteria that make B_{12} date to about 3.8 billion years ago (see Scott, 1993), so there has been ample time for competition between the Krebs cycle and alternatives that use B_{12}.

4.2. OUR METHOD

Our method in this work arose from the idea that a list of allowed reactions is not essential for constructing an ensemble of alternative metabolic networks. One can use stoichiometric

constraints on the networks' performance, and generic reactions without compounds specified, to find possible network topologies and stoichiometries of reactions, and so to generate an ensemble of networks that transform carbon skeletons—C-nets. We ranked C-nets according to their reuse of reactions to meet more than one stoichiometric constraint. We then converted the best C-nets to realistic networks involving fully specified compounds—R-nets—by hand, choosing carbon skeletons and adding functional groups. As Mittenthal *et al.* (1998) showed for the pentose phosphate pathway, functional groups can be added automatically to C-nets, using the principle that the functional groups change minimally as the network transforms starting reactants to final products. This principle is a special case of the principle of minimum structure change in chemical reaction networks, known from the 19th century (see Temkin *et al.*, 1996, pp. 25–27). It seems likely that the latter principle could be used to automate the transformation and rearrangement of carbon skeletons in the conversion of C-nets to R-nets.

Our method may be more widely useful. We believe that knowledge of initial and final compounds can be used with generic reactions to design molecular networks other than those in metabolism. In metabolic networks enzymes catalyse transformations of metabolites, but the enzymes are not transformed. However, in other networks macromolecules catalyse the modification of other macromolecules. Macromolecules may also associate with, and dissociate from each other without covalent modification. Such macromolecular networks mediate most of the processes in a cell, including signal transduction, gene regulation, the movement of processive enzymes along filamentous macromolecules, and the flow of molecules through channels in membranes. Because macromolecular networks are currently under intensive investigation, it would be useful to have a method for assembling them from reactions that were characterized to various extents, from generic to fully specified. A suitable method might be an extension of our approach.

B. Clarke gratefully acknowledges funding from NSERC Operating Grant OGP-0138122. The authors would like to express their gratitude to Gary Catt of Caracle Consulting Inc. for his generosity while guiding us through the programming labyrinth.

REFERENCES

BALDWIN, J. E. & KREBS, H. (1981). The evolution of metabolic cycles. *Nature* **291,** 381–382.

BROWN, G. C. (1991). Total cell protein concentration as an evolutionary constraint on the metabolic control distribution in cells. *J. theor. Biol.* **153,** 195–203.

BUCHANAN, B. B. & ARNON, D. I. (1990). A reverse Krebs cycle in photosynthesis: consensus at last. *Photosynth. Res.* **24,** 47–53.

GARRETT, R. H. & GRISHAM, C. M. (1995). *Biochemistry*, pp. 604–605. New York: Saunders.

GEST, H. (1987). Evolutionary roots of the citric acid cycle in prokaryotes. *Biochem. Soc. Symp.* **54,** 3–16.

HAPPEL, J., SELLERS, P. H. & OTAROD, M. (1990). Mechanistic study of chemical reaction systems. *Ind. Eng. Chem. Res.* **29,** 1057–1064.

HEINRICH, R., MONTERO, F., KLIPP, E., WADDELL, T. G. & MELÉNDEZ-HEVIA, E. (1997). Theoretical approaches to the evolutionary optimization of glycolysis. Thermodynamic and kinetic constraints. *Eur. J. Biochem.* **243,** 191–201.

HEINRICH, R., MELÉNDEZ-HEVIA, E., MONTERO, F., NUÑO, J. C., STEPHANI, A. & WADDELL, T. G. (1999). The structural design of glycolysis: an evolutionary approach. *Biochem. Soc. Trans.* **27,** 294–298.

KIRKWOOD, T. B. L., ROSENBERGER, R. F. & GALAS, D. J. (1986). *Accuracy in Molecular Processes.* New York: Chapman & Hall.

KSANDER, G., BOLD, G., LATTMAN, R., LEHMANN, C., FRÜH, T., YIANG, Y.-B., INOMATA, K., BUSER, H.-P., SCHREIBER, J., ZASS, E. & ESCHENMOSER, A. (1987). Chemie der α-Aminonitrile. 1. Mitteilung. Einleitung und Wege zu Uroporphyrinogen-octanitrilen. *Helv. Chim. Acta* **70,** 1115–1172.

MATHEWS, C. K., VAN HOLDE, K. E. & AHERN, K. G. (2000). *Biochemistry*, 3rd Edn, New York: Addison-Wesley.

MAVROVOUNIOTIS, M. L., STEPHANOPOULOS, G. & STEPHANOPOULOS, G. (1990). Computer-aided synthesis of biochemical pathways. *Biotechnol. Bioeng.* **36,** 1119–1132.

MELÉNDEZ-HEVIA, E. (1990). The game of the pentose phosphate cycle: a mathematical approach to study the optimization in design of metabolic pathways during evolution. *Biomed. Biochim. Acta* **49,** 903–916.

MELÉNDEZ-HEVIA, E., WADDELL, T. G. & MONTERO, F. (1994). Optimization of metabolism: the evolution of metabolic pathways toward simplicity through the game of the pentose phosphate cycle. *J. theor. Biol.* **166,** 201–220.

MELÉNDEZ-HEVIA, E., WADDELL, T. G. & CASCANTE, M. (1996). The puzzle of the Krebs citric acid cycle: assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways during evolution. *J. Mol. Evol.* **43,** 293–303.

MELÉNDEZ-HEVIA, E., WADDELL, T. G., HEINRICH, R. & MONTERO, F. (1997). Theoretical approaches to the evolutionary optimization of glycolysis. Chemical analysis. *Eur. J. Biochem.* **244,** 527–543.

MITTENTHAL, J. E. (1996). An algorithm to assemble pathways from processes. In: *Pacific Symposium on Biocomputing '97* (Altman, R. B., Dunker, A. K., Hunter, L. & Klein, T. E., eds), pp. 292–303. Singapore: World Scientific.

MITTENTHAL, J. E., CLARKE, B. & LEVINTHAL, M. (1993). Designing bacteria. In: *Thinking About Biology* (Stein, W. D. & Varela, F. J., eds), Lecture Notes, Vol. III, Santa Fe Institute Studies in the Sciences of Complexity, pp. 65–103. Reading, MA: Addison-Wesley.

MITTENTHAL, J. E., YUAN, A., CLARKE, B. & SCHEELINE, A. (1998). Designing metabolism: alternative connectivities for the pentose phosphate pathway. *Bull. Math. Biol.* **60,** 815–856.

MOROWITZ, H. (1999). A theory of biochemical organization, metabolic pathways, and evolution. *Complexity* **4,** 39–53.

NUÑO, J. C., SÁNCHEZ-VALDENEBRO, I., PÉREZ-IRATXETA, C. & MELÉNDEZ-HEVIA, E. (1997). Network organization of cell metabolism: monosaccharide interconversion. *Biochem. J.* **324,** 103–111.

SCOTT, A. I. (1993). How nature synthesizes vitamin B_{12}—a survey of the last four billion years. *Angew. Chem. Int. Ed. Engl.* **32,** 1223–1243.

STEPHANI, A. & HEINRICH, R. (1998). Kinetic and thermodynamic principles determining the structural design of ATP-producing systems. *Bull. Math. Biol.* **60,** 505–543.

STEPHANI, A., NUÑO, J. C. & HEINRICH, R. (1999). Optimal stoichiometric designs of ATP-producing systems as determined by an evolutionary algorithm. *J. theor. Biol.* **199,** 45–61.

TEMKIN, O. N., ZEIGARNIK, A. V. & BONCHEV, D. (1996). *Chemical Reaction Networks.* Boca Raton, FL: CRC Press.

WÄCHTERSHÄUSER, G. (1990). Evolution of the first metabolic cycles. *Proc. Natl Acad. Sci. U.S.A.* **87,** 200–204.

WEITZMAN, P. D. J. (1985). Evolution in the citric acid cycle. *FEMS Symp.* **29,** 253–275.

ZUBAY, G. (1998). *Biochemistry*, 4th Edn. Boston: WC Brown.

APPENDIX A

This appendix provides additional information about the algorithms we used. Our program to implement these algorithms is available on request from Jay Mittenthal (*mitten@life.uiuc.edu*) or Bertrand Clarke (*bertrandclarke@hotmail.com*).

Recall that a generic reaction (g-reaction) has as many as two non-zero inputs and two non-zero outputs. These inputs and outputs are unspecified initially but become increasingly characterized through the procedure. A g-reaction may have one null input, so it converts one input to two outputs, or a null output, so it converts two inputs to one output. We want to link g-reactions to form networks that represent a sequence of carbon skeleton-changing reactions, so the inputs and outputs of each g-reaction will be positive integers representing the number of carbons in a compound. The resulting network must be designed so that its operation will satisfy a pre-specified stoichiometric constraint.

The overall objective of the following algorithm is to produce a sequence of paranets, in order of increasing size. Each of these paranets satisfies a finite collection of stoichiometric constraints. The algorithm has two phases: in phase 1, we construct for each constraint an ensemble of C-nets that meet it. In phase 2, we construct paranets by combining C-nets, one from each ensemble.

A.1. Phase 1: Construction of an Ensemble for a Constraint

A.1.1. DEFINITIONS

The single constraint of interest can be written as

$$\sum_{i=1}^{k} n_i C(i) \to \sum_{i=1}^{k'} n_i' C(i).$$

The notation $C(i)$ represents any input or output compound containing i carbons. The number of inputs with i carbons is n_i and the number of outputs with i carbons is n_i'. The maximum number of carbons is k in an input compound and is k' in an output compound. Thus, conservation of carbons implies $\sum n_i C(i) = \sum n_i' C(i)$.

We distinguish between specified and unspecified inputs and outputs. We say an input (output) of a g-reaction is specified if it is one of the k inputs (k' outputs) of the constraint. The other inputs and outputs of the g-reactions are unspecified initially. They may be null (zero carbons) inputs or outputs. A non-null unspecified input or output is called internal. Our procedure will connect each internal output to an internal input. This connection is called a link. So, the number of links equals the number of internal inputs or equivalently the number of internal outputs.

Let N be the number of g-reactions in a C-net. Each of these g-reactions has two input tips and two output tips. For each g-reaction there is a conservation of carbon (cc) equation, because the total number of carbon atoms is the same for the inputs and outputs. For

example, the cc equation is $a + b = c + d$ for the g-reaction

$$\begin{matrix} a & c \\ & \mathrm{X} \\ b & d \end{matrix}$$

This g-reaction can also be represented by a C-net table:

	in		out	
X:	a	b	c	d

Now, for N g-reactions there are $2N$ input tips and $2N$ output tips. Both of these numbers of tips are sums of three terms. For input tips $2N$ is the sum of the specified, internal, and null inputs. Correspondingly, for output tips, $2N$ is the sum of specified, internal and null outputs.

A.1.2. BOUNDS ON THE NUMBER OF G-REACTIONS AND NUMBER OF LINKS

For the given constraint, we first identify a minimal value of N, N_{min}, and then give the minimal and maximal number of links for each $N \geqslant N_{min}$. We denote the number of links in a C-net by L.

To find N_{min} note that the g-reactions must provide enough input tips for the number of specified inputs, and enough output tips for the number of specified outputs. That is, the smallest possible value of N is $N_{min} = 1/2 \max(\lceil \sum n_i \rceil, \lceil \sum n_i' \rceil)$. For example, in the constraint $3 \to 1 + 1 + 1$, $\sum n_i = n_3 = 1$ and $\sum n_i' = n_i' = 3$, so $N_{min} = 2$.

We require that each C-net be connected. A connected C-net with N g-reactions must have at least $N - 1$ links between the g-reactions. That is, $L_{min} = N - 1$. Note that there will not, in general, be a connected C-net with the minimal N. (Consider the constraint $4 \to 1 + 1 + 1 + 1$, which has $N_{min} = 2$. All four output tips are used for specified outputs, so none remain to form links.)

For a given N there is also an upper bound on L: $L_{max} = 2N - \max(\sum n_i, \sum n_i')$. This is so because the number of links that can form is the lesser of the number of unspecified input tips and the number of unspecified output tips, which is

$$\min\left(2N - \sum n_i, 2N - \sum n_i'\right)$$

$$= 2N - \max\left(\sum n_i, \sum n_i'\right).$$

Thus, we have bounds for L in terms of N:

$$N - 1 = L_{min} \leqslant L \leqslant L_{max}$$

$$= 2N - \max\left(\sum n_i, \sum n_i'\right).$$

Restricting the maximum number of group transfer reactions, $N_{gtr\,max}$, decreases the maximum number of links. This is so because each group transfer reaction has two inputs and two outputs, so it can form more links than other g-reactions, which have two inputs and one output or two outputs and one input. Using a lower L_{max} appropriate to the limit on $N_{gtr\,max}$ avoids much needless computation.

Here we describe the upper bound on the number of links. First, in the absence of extra information the upper bound on L is $L_{absmax} = 2N - \max(\sum n_i, \sum n_i')$. However, we can improve this bound by using an upper bound on the number of group transfer reactions, the number of input constraints and the number of output constraints. This continues to be a worst-case upper bound, equal to or larger than the actual number of links in all cases.

To find an upper bound on the number of links, it is enough to show how to distribute the minimum number of nulls optimally over the available input and output locations.

First, since a group transfer reaction has two inputs and two outputs it has no null inputs or outputs. Thus, if $N_{gtr\,max}$ is less than N, there must be some null inputs or outputs, one null for each of the $N - N_{gtr\,max}$ non-group transfer reactions.

Second, we can reduce this number of nulls by the absolute magnitude of the difference between the number of input molecules and the number of output molecules because this is the number of

nulls that can be placed without affecting the number of links.

Third, we subtract half the number of nulls because there are half as many links as there are nulls.

Finally, we get an upper bound on the number of links of the form

$$\text{maxlinks} = (2*[N - N_{gtr\,max}])$$

$$-|\text{input\_molecules} - \text{output\_molecules}|$$

$$-\text{ceil}(\text{nulls}/2).$$

Here the factor of 2 in the first term on the right acts as if each of the three-tipped g-reactions actually has four tips, as in the formula for $L_{abs\,max}$ above. The number of nulls in the last term is the number of blanks left after assigning the nulls that are required to have the same number of inputs as outputs (middle terms).

To clarify the reasoning behind this result consider a simplified example. Suppose we have five reactions (with ten input tips and ten output tips), three inputs and one output. We can draw this as

Inputs: K K K i i i i i i i (max links = 7),

Outputs: K o o o o o o o o o,

where K indicates a molecule. We can add two nulls (labeled E, empty) without reducing the maximal number of links (connections from i's to o's) because the number of nulls that can be added without effect is the difference between the number of input molecules and the number of output molecules. This gives

Inputs: K K K i i i i i i i (max links = 7),

Outputs: K E E o o o o o o o.

Adding any one null reduces the number of links by one, but having added one null adding a second has no effect because we have already removed one end of the link. Thus, we get:

Inputs: K K K E i i i i i i (max links = 6),

Outputs: K E E E o o o o o o.

It is seen that because links join inputs and outputs, the optimal distribution of nulls assigns the same number to inputs as outputs, leading to half as many links as there are places for nulls.

A.1.3. PROCEDURE FOR SOLUTION IN PHASE 1

Loop over N, starting with N_{min}. For each N, loop over L, proceeding from L_{min} to L_{max}. For each N and L construct an ensemble of C-nets through the following four construction steps.

1. Put in links in all possible ways. Remove all isomorphic duplicates. Label both ends of a link with the same letter. The letter represents the number of carbon atoms in the compounds at the link. Use a different letter for each link.
2. Put in specified inputs and outputs in all possible ways. Remove all isomorphic duplicates.
3. Set all remaining input and output tips to zero.
4. This algorithm avoids most, but not all, isomorphism. The remaining isomorphism is removed by a duplicate-removal process which considers two C-nets to be identical if they have the same number of reactions, and if the reactions of one can be permuted in such a way that all the carbon values match those of the second. The actual connections of links are not compared.

At any of these stages some of the C-nets may be unacceptable for the reasons indicated in the disqualifier rules (4) of Section 2.3.

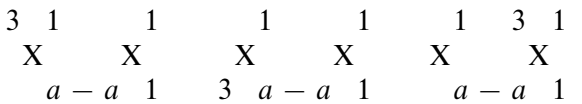
The preceding steps of construction label every tip of every g-reaction with the number of carbons in the compound at the tip. The number of compounds at some input and output tips are specified from the constraint, and other tips are null. The remaining tips are internal; they participate in links. The number of carbons in the compounds at these tips can be determined by solving the set of N cc equations.

5. Write the N cc equations and simplify them to find a unique solution or a class of solutions (if they exist). If there is a class of solutions, it can be expressed in terms of parameters equal in number to the number

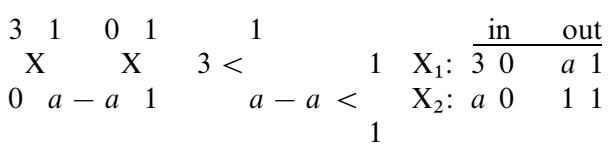
of links minus the number of linearly independent cc equations.

A.1.4. EXAMPLE OF PROCEDURE FOR SOLUTION IN PHASE 1: GLOBAL REACTION I, 3 + 1 + 1 + 1

Here, $N_{min} = 2$ as noted above. For $N = N_{min}$, $L_{min} = L_{max} = 1$ so there is only one case—that of one link—to be considered. We proceed through the construction steps. The first step is to put in the link between the two g-reactions. Then, there is only one way to put in the outputs and there are three ways to put in the inputs:



The first two of these C-nets are isomorphic, and the third one is ruled out by D2. Setting all other tips to zero, the resulting C-net can be displayed in three forms. In the third form, X_1 and X_2 are g-reactions.



	in	out
X_1:	3 0	a 1
X_2:	a 0	1 1

From any of these forms we get two cc equations:

$$3 = a + 1,$$

$$a = 1 + 1.$$

So, there is a unique solution with $a = 2$. Here we introduced one parameter and got one independent solution as expected.

For the case $N = 3$, the class of solutions is much larger. We have $L_{min} = N - 1 = 2$ and $L_{max} = 2N - \max(1,3) = 3$, and for each of the two values of L we get many candidate C-nets to examine. It can be verified that for $N = 3$ and $L = 2$ all of the candidates violate at least one of the disqualifier rules, so that none of the candidates is viable.

For $N = 3$ and $L = 3$, we illustrate the application of each construction step to a subset of the C-nets generated by the previous construction step. Putting in the first two links, which are required to get a connected C-net, gives the following three cases:

Case A:

	in	out
X_1:		a
X_2:	a	b
X_3:	b	

Case B:

	in	out
X_1:		a
X_2:		b
X_3:	a b	

Case C:

	in	out
X_1:		a b
X_2:	a	
X_3:	b	

In each of these C-nets the third link cannot connect an input tip to an output tip within the same X. A free input tip on any of the 3 X's can connect to a free output tip on any other X. If a X has two free input tips, or two free output tips, one of the tips can be ignored in making the third link. (If the ignored tip were used, it would make networks isomorphous to those using the other tip.)

In Case A, each of the X's has at least one free input tip and one free output tip. Hence, the third link can be added by taking the 3 X's two at a time, in 3!/(2!1!) = 3 ways. For each of these ways one can connect the input tip of the first X to the output tip of the second X, or vice versa. Thus, Case A gives six arrangements of three links, of which two are

#1	in	out
X_1:		a c
X_2:	a	b
X_3:	b c	

#2	in	out
X_1:	c	a
X_2:	a	b
X_3:	b	c

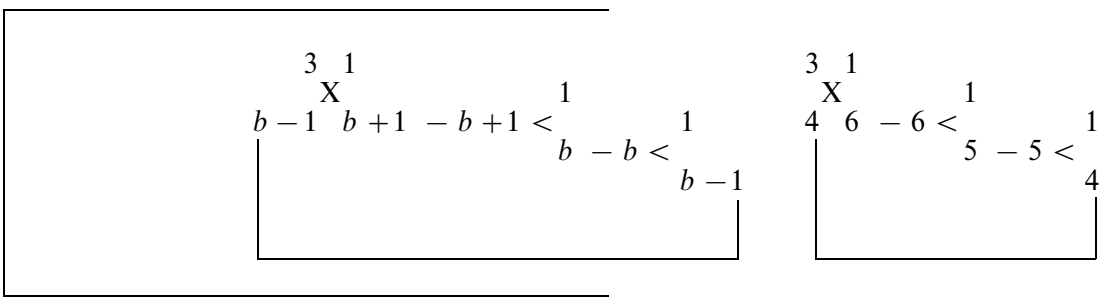
In two of the four arrangements not shown, both outputs of one g-reaction are linked to both inputs of another g-reaction. Effectively, these two g-reactions are functioning as a single g-reaction, reducing the effective N of the C-net to 2. Hence,

we do not need to consider these C-nets further, because they can be elaborated from C-nets with $N = 2$. However, it is worth noting that such C-nets can be biologically important; for example, in the pentose phosphate pathway two g-reactions are linked in this way.

Putting in inputs, outputs, and zeros for the second of these C-net tables gives the following three C-net tables:

#2.1	in		out	#2.2	in		out	#2.3	in		out
X_1:	c	3	a 1	X_1:	c	0	a 1	X_1:	c	0	a 1
X_2:	a	0	b 1	X_2:	a	3	b 1	X_2:	a	0	b 1
X_3:	b	0	c 1	X_3:	b	0	c 1	X_3:	b	3	c 1

atoms are not allowed. When $b = 1$, $c = 0$, so no transformation occurs in the third reaction. The system degenerates to two reactions, a case that would have been studied under $N = 2$. When $b = 2$, $c = 1$ and $a = 3$, so no transformation occurs in the first reaction; again the system degenerates to two reactions. As regards larger values of b, the upper limit on the size of compounds is 6 carbon atoms. So, b must be 3, 4 or 5. This is a family of C-nets which can be represented in general on the left and for $b = 5$ on the right, thus,



Note that the C-nets derived from these tables cannot be interconverted by permuting the variables a, b, and c because these variables represent links at non-isomorphic locations in the C-net.

Table A.1 implies three cc equations of which two are linearly independent:

$$c + 3 = a + 1,$$

$$a = b + 1,$$

$$b = c + 1.$$

With b as a parameter, the solutions are $a = b + 1, b = b, c = b - 1$. For various reasons, $b = 0, 1, 2$ do not yield viable C-nets: when $b = 0, c = -1$, but negative numbers of carbon

The C-net on the right is the portion of C-paranet #4 that meets stoichiometric constraint I, $3 \rightarrow 1 + 1 + 1$, as comparison with Fig. 3 shows.

A.1.5. RECURRENTS

Intuitively, a feed in a C-net is a link from an output of an earlier-occurring reaction to an input of a later-occurring one. A recurrent is a link from an output of a later-occurring reaction to an input of an earlier-occurring one. However, these definitions are not satisfactory, in that earlier, later, and the apparent numbers of feeds and recurrents depend on the order of reactions in

a diagram of the C-net, and the diagram can be drawn in various ways.

The link between recurrents and the sequence of reactions arises because each recurrent is in a cycle through the C-net. A cycle is a sequence of compounds that ends at the compound where it starts, and has no other repeated compound. The identification of a cycle considers only the path from one compound to the next, not the g-reactions in which the compounds participate. Since any compound in a cycle can be a recurrent, it is necessary to specify a way to choose recurrents; we do this next.

will be called incompatible if the precedences in the subset cannot all be realized in any sequence of its g-reactions. That is, the links in an incompatible subset cannot all be feeds, so that one of these links must be a recurrent. Two incompatible subsets are independent, i.e. disjoint, if they do not share any precedences. The number of independent incompatible subsets is the minimum number of recurrents.

As an example, consider the following C-net. Letters designate links between g-reactions. The small numbers to the right of the X's designate reactions:

$$\begin{array}{cccc} 3\ 5 & 3\ a=4 & c=3\ 1 & d=2\ b=1 \\ X_1 & X_2 & X_3 & X_4 \\ a=4\ d=2 \quad b=1 & & e=2 & e=2\ c=3 \end{array}$$

It is seen there are two recurrents because there are exactly two disjoint cycles. One cycle passes leftward (lw) from g-reaction X_4 to X_3, and then rightward (rw) from X_3 to X_4:

$$\begin{array}{cc} \text{lw from} & \text{rw from} \\ X_4 \text{ to } X_3 & X_3 \text{ to } X_4 \end{array}$$

$$\begin{array}{ccccccccc} c=3 \text{ out} \rightarrow & c=3 \text{ in} & \rightarrow & e=2 \text{ out} \rightarrow & e=2 \text{ in} \rightarrow & c=3 \text{ out} \\ \text{from } X_4 & \text{to } X_3 & & \text{from } X_3 & \text{to } X_4 & \text{from } X_4. \end{array}$$

The other cycle passes between X_4, X_2, and X_1:

$$\begin{array}{ccc} \text{lw from} & \text{lw from} & \text{rw from} \\ X_4 \text{ to } X_2 & X_2 \text{ to } X_1 & X_1 \text{ to } X_4 \end{array}$$

$$\begin{array}{cccccccccccc} b=1 \text{ out} \rightarrow & b=1 \text{ in} \rightarrow & a=4 \text{ out} \rightarrow & a=4 \text{ in} \rightarrow & d=2 \text{ out} \rightarrow & d=2 \text{ in} \rightarrow & b=1 \text{ out} \\ \text{from } X_4 & \text{to } X_2 & \text{from } X_2 & \text{to } X_1 & \text{from } X_1 & \text{to } X_4 & \text{from } X_4 \end{array}.$$

We operationalize finding the minimum number of recurrents over the connectivities of a C-net by drawing the C-net with the g-reactions in a horizontal line. Number the reactions arbitrarily from 1 to N, the number of reactions in the C-net. For each link between two g-reactions, tabulate which of the g-reactions must precede the other one in order that the link be a feed rather than a recurrent. Call this relation the precedence for the link. A subset of the precedences

Note that each cycle has at least one edge in each direction. The cycles can be represented more conveniently as a precedence table:

link	precedence
a	reaction 2 before reaction 1
b	4 before 2
c	4 before 3
d	1 before 4
e	3 before 4.

Evidently, the precedences for links *c* and *e* are an incompatible subset, as are the precedences for links *a*, *b*, and *d*. These two subsets are independent, so there are two disjoint cycles. The precedence relations reflect the ordering within a cycle as a function of the labeling of the g-reactions. The linear sequence of reactions 4, 3, 2, 1 displays the minimum number of recurrents, which is two; in this sequence the recurrent links are *d* and *e*.

A.1.6. FUTILE CYCLES

As noted in Section 2.3, a futile cycle is a subset of the reactions of a connected C-net that has the null reaction, $0 \rightarrow 0$, as its net reaction. Within a futile cycle each output of one reaction is the same as an input of another. We search for futile cycles by ignoring the connections among reactions, looking at each subset of reactions, and looking for its net reaction. A subset is futile if, for each type of molecule (number of carbons) occurring in the subset, the number of molecules produced as outputs is the same as the number of molecules consumed as inputs. A C-net is futile if any of its subsets are futile.

A futile subset of size 1 is a reaction in which the outputs are the same as the inputs. A futile subset of size 2 has a reaction and its reverse, such as

$$\begin{array}{cccc} 7 & 6 & 6 & 7 \\ & X & & X \\ & 1 & 1 & \end{array} \quad .$$

Here is a futile subset of size 3:

$$\begin{array}{cccccc} 5 & 2 & & 1 & 2 & \\ & X & & X & & X \\ 3 & 6 & 6 & 5 & 1 & 3 \end{array} \quad .$$

We ignored C-nets containing futile cycles.

A.2. Phase 2: Construction of an Ensemble of Paranets, Each of Which Meets a Collection of Constraints

A.2.1. ALGORITHM FOR CONSTRUCTING THE ENSEMBLE OF PARANETS

In phase one we get a finite collection of C-nets for each *N* and *L*. These can be ordered in a sequence. Taken together, all of these sequences comprise an ensemble of C-nets that meet a stoichiometric constraint. Consequently, we can regard the entire output of phase 1 as a sequence of C-nets.

We will construct paranets from the sequences that meet the several constraints of interest. Recall that a paranet is a kind of network produced by identifying the largest possible overlaps among finitely many networks. We consider the case of three constraints, since that is the number of constraints that we assume the Krebs cycle meets; other numbers of constraints can be treated similarly. Consider the table

constraint	C-net
1	$G_{11}, G_{12}, G_{13}, \ldots$
2	$G_{21}, G_{22}, G_{23}, \ldots$
3	$G_{31}, G_{32}, G_{33}, \ldots$

in which G_{ij} is the *j*-th C-net in the ordered sequence for constraint *i*. We proceed by using triples of the form (G_{1i}, G_{2j}, G_{3k}), ordered by the sum of the number of X's in the three C-nets. We break ties arbitrarily. (Thus, we obtain all possible sums of the numbers of X's in C-nets, one from each row of the table, and order these from smallest to largest.)

Fix a triple (G_{1i}, G_{2j}, G_{3k}). For each triple there are three pairs of C-nets. Consider one such pair—say, (G_{1i}, G_{2j}). The overlap between G_{1i} and G_{2j} is the g-reactions shared between them. It is well known that the overlap can be represented as a finite collection of connected components. If there is only one connected component, possibly the case of greatest interest, then we regard it as a module. It is a module in the sense that it is used in both C-nets, which meet two constraints. For the g-reactions within a module, we require the connectivity of their occurrence in meeting the two constraints to be isomorphic. In fact, our program only searches for overlaps, ignoring their connectivity. We have found empirically that overlaps of one g-reaction are typical in smaller C-nets. Disconnected overlaps, possibly with non-isomorphic connectivities, only appear for larger C-nets.

A.2.2. ALGORITHM FOR FINDING THE OVERLAPS

To find the overlap between two C-nets, say G_1 and G_2, represent them by their C-net tables. Convert each table to dictionary order. That is, within the inputs and within the outputs of each g-reaction put the smaller entry first. Then, within each table, put the rows in dictionary order. Now, it is enough to go through the rows of each table to find the cases where they match. Each case is a g-reaction in the overlap.

Note that the C-net tables we described in Section A.1.4 include the connectivity of the C-net, which the present algorithm ignores. If we want to characterize patterns of connectivity in the overlaps then the present algorithm must be elaborated to determine whether matching rows have isomorphic connectivities. Strict module structure would require them to be isomorphic. However, it is possible to have two C-nets that share several g-reactions with non-isomorphic connectivities. In such cases, it may or may not be possible to modify the connectivities to make them isomorphic.

A.3. Effect of Condensing two g-Reactions into one, on the Value of N_C/P_C for a Paranet

If a g-reaction that joins two compounds to form one is followed by a g-reaction that cleaves the first output into two other compounds, the two g-reactions can be condensed into one group transfer reaction. Here we show the cases in which such a condensation reduces the quality index $Q_C = N_C/P_C$ for C-paranets of interest, thereby improving the ranking of the condensed paranet relative to its uncondensed partner.

There are two cases: first, suppose the initial g-reaction of the pair to be condensed produces a 5-carbon compound that meets constraint II ($3+3 \rightarrow 5+1$), or a 4-carbon compound that meets constraint III ($3+3 \rightarrow 4+1+1$). Then condensing the two g-reactions deletes the 5- or 4-carbon compound. This condensation does not reduce Q_C because an additional reaction must be added to the condensed paranet to meet constraint II or III, thereby increasing its Q_C.

Second, suppose the initial g-reaction does not produce a compound that meets a constraint. Now, every C-net in the paranet must either include or exclude both g-reactions (in sequence), or the group transfer reaction derived from them. In this second case, let N be the number of reactions in the condensed paranet, and let S be the total number of reactions in all three C-nets derived from the condensed paranet. We see that S satisfies $N+2 \leqslant S \leqslant 3N$: the lower bound obtains because each g-reaction in the paranet must be used in at least one of the three C-nets, and connectedness of the paranet requires that at least two of the three possible pairs of C-nets must share at least one g-reaction; hence, a lower bound on S is $N+2$. The upper bound of $3N$ on S is achieved when all three C-nets are identical and use all of the g-reactions. Now, let K be the number of C-nets derived from the condensed C-paranet that include the group transfer reaction. K may be 1, 2, or 3.

Now, we show that $Q_{uncond}/Q_{cond} > 1$—that is, the condensed paranet is better—for K, S, and N of interest. Observe that

$$Q_{uncond}/Q_{cond} = (N+1/N)P_{cond}/P_{uncond}$$

$$= (N+1/N)(S/3N)(3(N+1)/S+K).$$

This follows by definition of the Q's in terms of the P's and the substitutions $P_{cond} = S/3N$, $P_{uncond} = (S+K)/[3(N+1)]$. The latter follows because the uncondensed paranet has one more g-reaction than the condensed paranet, and has $S+K$ as the total number of reactions in all three C-nets derived from the condensed paranet.

To find the values of S for which the ratio Q_{uncond}/Q_{cond} is greater than 1, we rearrange the above expression to get

$$S > N\{K/[2+(1/N)]\}.$$

We set $K=3$ as the worst-case scenario. Now, for $N=4$ we get $S > 12/2.333 = 5.14$ and the lower bound on S when $N=4$ is $N+2=6$. When $N=5$, we get $S > 15/2.2 = 6.81$ and the lower bound on S when $N=5$ is $N+2=7$. Clearly, when the lower bound on S satisfies the above inequality in the worst-case scenario for a given N, all other values of S also satisfy it. Thus, for $N=4$ and 5, Q_{uncond}/Q_{cond} is greater than 1 for all allowed values of S. However, when $N=6$, S must be greater than $18/2.166 = 8.3$ but the lower bound on S is $N+2=8$, so the bound fails. For $N \geqslant 6$ there will be, in general, values of S for which the ratio $Q_{uncond}/Q_{cond} < 1$.