Molecular Cell
Article

Cell
PRESS

Central Carbon Metabolism as a Minimal Biochemical Walk between Precursors for Biomass and Energy

Elad Noor,[1,2] Eran Eden,[2] Ron Milo,[1] and Uri Alon[2,*]
[1]Department of Plant Sciences
[2]Department of Molecular Cell Biology
Weizmann Institute of Science, Rehovot 76100, Israel
*Correspondence: uri.alon@weizmann.ac.il
DOI 10.1016/j.molcel.2010.08.031

SUMMARY

Central carbon metabolism uses a complex series of enzymatic steps to convert sugars into metabolic precursors. These precursors are then used to generate the entire biomass of the cell. Are there simplifying principles that can explain the structure of such metabolic networks? Here we address this question by studying central carbon metabolism in *E. coli*. We use all known classes of enzymes that work on carbohydrates to generate rules for converting compounds and for generating possible paths between compounds. We find that central carbon metabolism is built as a minimal walk between the 12 precursor metabolites that form the basis for biomass and one precursor essential for the positive net ATP balance in glycolysis: every pair of consecutive precursors in the network is connected by the minimal number of enzymatic steps. Similarly, input sugars are converted into precursors by the shortest possible enzymatic paths. This suggests an optimality principle for the structure of central carbon metabolism. The present approach may be used to study other metabolic networks and to design new minimal pathways.

INTRODUCTION

The metabolic network of *E. coli* is one of the best-characterized reaction networks in biology (Neidhardt et al., 1987; Heinrich and Schuster, 1996; Price et al., 2003; Karp et al., 2007; Stryer, 1995). It is thus an excellent model system to ask whether there exist simplifying principles that can explain aspects of biological structure (Savageau, 1976; Fell and Wagner, 2000; Alon, 2007; Janga and Babu, 2008; Fell, 1996).

An example of such a study was presented by Meléndez-Hevia and Isodoro in a paper entitled "The game of the pentose phosphate cycle" (Meléndez-Hevia and Isodoro, 1985). The pentose-phosphate pathway was abstracted as a problem of converting six molecules made of five carbons (pentoses) into five molecules made of six carbons (hexoses). The game had two rules for which steps are allowed during the transformations. These rules mimic the action of the relevant enzymes: (1) transfer either two or three carbons from one molecule to another and (2) no molecule can have less than three carbons. One can now find the solution to this game with the minimal number of steps (Figure 1). Remarkably, the minimal solution is equivalent to the naturally occurring pathway in *E. coli*, where transketolase (TK), transaldolase (TA), and aldolase (AL) catalyze the reactions. The pentose phosphate game illustrates the basic properties of a theory of optimal structure. The first feature is to define rules for generating possible networks for comparison—such as the two rules mentioned above. The second feature is boundary conditions—the input and output molecules. In the case of the pentose-phosphate game, the boundary conditions were six pentoses as input and five hexoses as output. Finally, one must stipulate the function to be optimized, in this case minimizing the number of steps to bridge between the boundary conditions.

The pentose phosphate game was followed up by Meléndez-Hevia and coworkers (Meléndez-Hevia et al., 1994). Additional studies analyzed glycolysis and the TCA cycle (Meléndez-Hevia et al., 1996), the major avenues for carbon utilization in most organisms. Numerous possible designs for the pentose phosphate pathway (Mittenthal et al., 1998) and the TCA cycle (Mittenthal et al., 2001) were generated by using a larger set of rules than in the pentose-phosphate game, but still focusing on the changes to the carbon skeleton of each compound (Ebenhöh and Heinrich, 2003). Work by Heinrich and others compared the glycolysis pathway to other alternatives and studied optimality principles regarding the ordering of ATP-consuming and ATP-producing reactions in that pathway (Stephani and Heinrich, 1998; Heinrich and Schuster, 1998; Stephani et al., 1999; Heinrich et al., 1997; Ebenhöh and Heinrich, 2001).

As more systems were studied, the simplicity of the original pentose phosphate game seemed to be a special case; researchers expanded the set of rules and considerations to include energy efficiency and production of NAD(P)H and other metabolites. Different studies employed different rules, boundary conditions (choice of which piece of metabolism to study), and optimization functions. Considerable progress has been made, increasing our understanding of the complexity of the problem. In particular, these studies suggest that the chemical nature of the metabolites must be taken into account in the search for a unifying principle for metabolic structure.

A distinct approach uses optimality to understand the distribution of fluxes in metabolism, using elementary modes or extreme

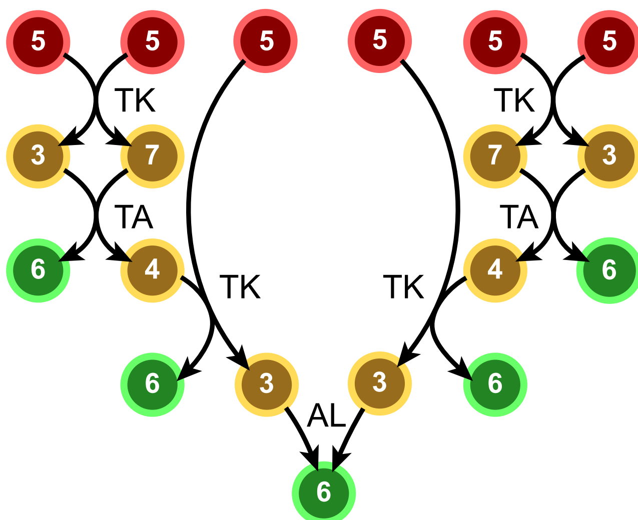

Figure 1. The Solution for the Game of the Pentose Phosphate Pathway
The pathway with the minimal number of steps as presented by Meléndez-Hevia and Isodoro (1985). This solution converts six pentoses into five hexoses using only seven steps, which is the minimal number of steps. This strategy is the one used in the actual pentose phosphate pathway of *E. coli*. Enzymes are TK (transketolase) that transfers two carbon atoms, TA (transaldolase) that transfers three carbon atoms, and AL (aldolase) that links two three-carbon molecules.

pathway analysis. In these approaches, one takes as known the stoichiometric structure of metabolism and computes fluxes that maximize objectives such as biomass production (Schuster et al., 1999; Schilling et al., 2000; Papin et al., 2004). Thus, rather than asking ''Why is metabolism built the way it is?'' elementary mode analysis seeks the fluxes corresponding to a given structure and input conditions.

Here we ask whether there exist principles that explain why metabolism has the reactions that it has, rather than other possible reactions that can bridge between metabolic inputs and outputs. We use as a model system the central carbon metabolic network of *E. coli* that catabolizes carbohydrates for growth. We define the rules using all known classes of enzymes that act on carbohydrates (Hatzimanikatis et al., 2005). Note that we use the set of all known enzyme *classes*—each representing a *type* of reaction, rather than the set of all known enzymes—each representing a specific reaction with specific substrates as in previous studies (Handorf et al., 2005; Raymond and Segre, 2006; Handorf and Ebenhöh, 2007).

The rules of the game include many more putative reactions than are biochemically feasible—for example, we do not take into account thermodynamic (Alberty, 2005) or steric constraints that can rule out certain reactions. Since we use all known enzyme classes, our set of reactions contains all feasible reactions that can appear in organisms, and additional reactions that are unfeasible. Thus, if we find that a specific pathway is minimal in the present large set of reactions, it is also minimal for the subset of biochemically feasible reactions. Additional constraints can be added to improve the present game by using, for example, recent work on estimating ΔG of reactions (Jankowski et al., 2008; Finley et al., 2009) and absolute concentrations of central metabolites (Bennett et al., 2009).

We find that the central metabolic network could have been much shorter—we find many examples of putative paths between parts of metabolites that are shorter (have fewer enzymes) than the paths that actually exist. The reactions in the real network used by the organisms share the following principle: they form a minimal walk between the well-characterized set of substances that are the origin of the cell's biomass (Neidhardt et al., 1990; Palsson, 2006). These *precursors* are the 12 compounds located at branchpoint nodes from which flux moves out of central carbon metabolism to make amino acids, nucleotides, and other biomass components, and one compound that is essential for a net gain of ATP in glycolysis. Every pair of consecutive precursors along the network is connected by a minimal biochemical path. Furthermore, we find that input sugars are converted into precursors by the shortest possible enzymatic paths.

RESULTS

Rules for Generating Pathways Based on All Relevant EC Enzyme Classes

In this study, we choose a large part of central carbohydrate metabolism of *E. coli* (Figure 2), which we denote for the present purposes as central carbon metabolism. This well-studied set of metabolic pathways, which are almost perfectly conserved across organisms, interconvert sugars and biomass. Biomass is generated by fluxes branching out from central carbon metabolism, originating at 12 well-known precursor substances (Neidhardt et al., 1990). At each such branchpoint node, flux branches out from central carbon metabolism to build the components of the cell's biomass (the precursors are listed in Table 1). In addition to providing biomass precursors, central carbon metabolism generates ATP and reducing agents. The present network includes most of the previously studied systems mentioned in the introduction: glycolysis, the TCA cycle, and the pentose phosphate pathway. We also include the best-studied systems in which environmental sugars are degraded to join the central pathway.

The central carbon metabolism network we study includes only carbohydrates and phosphate groups (and CoA) and thus is a simpler test case than full metabolism that involves nitrogen, sulfur, and other elements. We did not include anabolic pathways such as the amino acid biosynthesis pathways that branch out from the precursor nodes, because this would entail reactions that include nitrogen and sulfur, increasing the complexity of the computations beyond the present scope. The present results apply also when adding additional carbohydrate pathways such as gluconeogenesis, the glyoxylate shunt, and the Entner-Doudoroff bypass.

The metabolic network is represented in this study in the standard way, with compounds as nodes and enzymes as edges. Ubiquitous metabolites such as protons, water, ATP/ADP, and NAD(P)H/NAD(P)+ are not considered as nodes in the network, although they are taken into account in the biochemical transformation rules described next.

Figure 2. The Central Carbohydrate Metabolic Network

Solid squares show 12 precursor metabolites (as in Table 1), D-glucose-6-phosphate (G6P), D-fructose-6-phosphate (F6P), D-ribose-5-phosphate (R5P), D-erythrose-4-phosphate (E4P), D-glyceraldehyde-3-phosphate (GAP), glycerate-3P (3PG), phosphoenolpyruvate (PEP), pyruvate (PYR), acetyl-CoA (ACA), 2-ketoglutarate (2KG), succinyl-CoA (SCA) and oxaloacetate (OXA), and glycerate-1,3P (BPG), essential for a net gain of ATP in glycolysis. The purple squares contain the downstream end products of the biosynthetic pathways that branch out from central carbon metabolism.

We sought a way of generating biochemically plausible pathways between any two given compounds. For this purpose, we used all known classes of enzymes that operate on carbohydrates. To define the enzyme classes, we use the Enzyme Commission (EC) numbering system (Webb, 1992). The EC classes are a comprehensive set of biochemically possible reactions gathering together enzymes found in diverse organisms. Since we are dealing with central carbon metabolism, we included only enzyme classes whose product and substrate contain only C, O, H, P, and CoA (CoA was effectively treated as a single atom). This leads to a list of 30 enzyme classes (Table 2). Other EC classes have substrates or products with nitrogen, sulfur, and other elements not found in the network under consideration and thus do not contribute to the present study.

The EC classification has four numbers. The first three numbers define the class of reactions, and the last number defines the specific instance of the enzyme, with its specific substrate and product. Thus, EC 2.7.1 denotes hydroxyl kinases in general, and EC 2.7.1.11 is a phospho-fructo-kinase that acts on F6P (D-fructose-6P) to generate FBP (D-fructose-1,6P). The enzyme classes used for the present rules are thus defined by the first three EC numbers only.

Table 1. The 12 Precursor Metabolites for Biomass in *E. coli*

Number	Metabolite	Abbreviation	Building Blocks Produced
1	D-glucose-6-phosphate	G6P	glycogen, LPS
2	D-fructose-6-phosphate	F6P	cell wall
3	D-ribose-5-phosphate	R5P	His, Phe, Trp, nucleotides
4	D-erythrose-4-phosphate	E4P	Phe, Trp, Tyr
5	D-glyceraldehyde-3-phosphate	GAP	lipids
6	glycerate-3-phosphate	3PG	Cys, Gly, Ser
7	phosphoenolpyruvate	PEP	Tyr, Trp
8	pyruvate	PYR	Ala, Ile, Lys, Leu, Val
9	acetyl-CoA	ACA	Leu, lipids
10	2-ketoglutarate	2KG	Glu, Gln, Arg, Pro
11	succinyl-CoA	SCA	Met, Lys, tetrapyrroles (e.g., heme)
12	oxaloacetate	OXA	Asn, Asp, Ile, Lys, Met, Thr, nucleotides

In the present study, we described the action of each enzyme class as an operation on an input molecule (Hatzimanikatis et al., 2005). For example, a hydroxyl kinase (EC 2.7.1) takes as input a C-OH within a molecule and changes it to a C-O-PO3. An aldehyde dehydrogenase (EC 1.2) takes as an input O=C-OH in a molecule and changes it to O=CH, releasing a water molecule, and so on. Such syntactic rules were defined for each enzyme class (Table 2 and Hatzimanikatis et al., 2005).

Note that these rules consider the reacting groups within the molecule as independent and do not take into account the molecular context. For example, alcohol dehydrogenase (EC 1.1) changes C-OH to C=O or vice versa, without regard to the position of the carbon along the molecule and the structural context in which it occurs. These positional effects matter in real enzymes, and thus the present rules contain more reactions than are biochemically feasible.

We used these EC classification-based rules to explore all possible product molecules from a given molecule *X*. To apply the rules, one searches for a subpart in *X* that is an input substrate to an enzyme class, and accordingly modifies *X*. In this way, a large number of new product molecules can be generated from each *X* (Figure 3A).

Algorithm for Finding the Minimal Number of Enzyme Steps between Two Compounds

We next used these rules to generate paths between compounds. For any two compounds *X* and *Y*, we evaluated the minimal number of steps required to reach from one compound to the other. We begin with one compound, *X*, and generate all possible products using the EC classes. Then, in the following step, we generate in the same way all possible products for each of the molecules produced in the first step. The search is over when the required final compound *Y* was reached.

Every step multiplies the number of possible products by a factor of ~20 (depending on the chosen original molecule, see Table 3). This means that the complexity of the search grows exponentially with the distance from *X* to *Y*. In some cases the number of possibilities can reach several million. Therefore, the algorithm was made more efficient by simultaneously starting to generate products from both *X* and *Y*, and finishing the search when the ends meet—a bidirectional breadth-first search. The minimal path (or set of minimal paths) between *X* and *Y* is thus found (Figure 3B). Table 3 demonstrates the exponential growth of the number of paths between two compounds as a function of the length of the paths. A detailed description of the open-source software tool that performs these calculations and provides minimal pathways is available in the Supplemental Information and can also be accessed at http://www.weizmann.ac.il/mcb/UriAlon/Papers/pathfinder/.

Large Parts of the Central Carbon Metabolic Network Could Potentially Have a Shortcut

We applied the algorithm to find minimal paths between the compounds in central carbon metabolism. We find that the central carbon metabolism network does not always contain the shortest path between each pair of compounds. In Figure 4, we plot several of the shortcuts found by the present algorithm.

For example, glycolysis, the nine-step path from G6P to PYR, is not the shortest possible pathway. G6P can be converted to two molecules of GAP in less than four steps, e.g., by using EC classes 2.7.1 and 4.1.2 (shortcut 1 in Figure 4A). Similarly, GAP can be directly converted to 3PG using EC class 1.2 (aldehyde dehydrogenase) which means that the two-step pathway that exists in *E. coli* could in principle be made shorter (shortcut 2 in Figure 4A). The rest of glycolysis is minimal, i.e., the five steps that connect BPG to ACA are the minimum needed to do so. There are also numerous putative shortcuts that bypass many parts of the TCA cycle and the pentose phosphate pathway.

A further graphical description can be seen when plotting the pieces of metabolism within which all paths are minimal. We defined a minimality module in a given network as a subnetwork in which the path between any pair of nodes (compounds) is minimal, and where the subnetwork cannot be made larger by adding neighboring nodes. A generic example for determining minimality modules is given in Figures 4B–4D. First the existing reactions are compared to the possible reactions according to the EC classes from Table 2. Contrasting these two networks, the "real" and the "possible," defines the optimality modules—those sets of consecutive compounds that are connected by the shortest path in the "real" network. The minimality modules (Figure 4D) are the largest sets of compounds among which all paths are minimal. Thus a and b are in the same module (yellow) but c is not part of that module because a and c can be connected by a one-step pathway rather than the longer two-step pathway seen in the "real" network. The middle module (red) contains three compounds: b, c, and d. The pathway from b to d is minimal, since at least two steps

Table 2. Enzyme Classes that Operate on Carbohydrates

Number	EC Class	General Description	Schematic Reaction
1	1.1	alcohol dehydrogenase (ALC-DH)	CH–OH ↔ C=O
2	1.1.1	ALC-DH w/ carboxyl lyase	HO–CH–C–COOH ↔ O=C–CH + (CO_2)
3	1.2	aldehyde dehydrogenase (ALD-DH)	C=O + (HO) ↔ COOH
4	1.2.1	ALD-DH acetylating	O=C + (CoA) ↔ O=C–S–CoA
5	1.2.1	ALD-DH acetylating decarboxylating	O=C–COOH + (CoA) ↔ O=C–S–CoA + (CO_2)
6	1.2.1	ALD-DH phosphorylating	O=C + (O–PO_3) ↔ O=C–O–PO_3
7	1.3	C–C reductase	C–C ↔ C=C
8	2.1.1	methyl-transferase	R + (CH) ↔ R–CH
9	2.1.2	hydroxymethyl-transferase	R + (C–OH) ↔ R–C–OH
10	2.1.2	formyl-transferase	R + (C=O) ↔ R–C=O
11	2.1.3	carboxyl-transferase	R + R–COOH ↔ R–COOH + R
12	2.2.1	transketolase	O=R + OH–R–CO–COH ↔ OH–R–CO–COH + O=R
13	2.2.1	transaldolase	O=R + OH–R–COH–CO–COH ↔ OH–R–COH–CO–COH + O=R
14	2.3.1	acyl-transferase	R–CO–CH + R ↔ R + R–CO–CH
15	2.3.3	acyl-transferase with CoA	O=R + C–CO–S–CoA + (HO) ↔ OH–R–C–CO–OH + (CoA)
16	2.7.1	kinase (hydroxyl)	C–OH + (PO_3) ↔ C–O–PO_3
17	2.7.2	kinase (carboxyl)	COOH + (PO_3) ↔ COO–PO_3
18	2.7.1	kinase (phosphoenol)	C–C=O + (PO_3) ↔ C=C–O–PO_3
19	3.1.1	carboxyl-ester hydrolase	R–CO–O–R + (HO) ↔ R–COOH + R–OH
20	3.3.2	ether hydrolase	R–O–R + (HO) ↔ HO–R + R–OH
21	3.7.1	ketone hydrolase	CH–C=O + (HO) ↔ CH + COOH
22	4.1.1/6.4.1	carboxyl lyase and ligase	R–COOH ↔ RH + CO_2
23	4.1.2/4.1.3	aldehyde lyase and oxo-acid lyase	R–C–OH ↔ RH + C=O
24	4.2.1	hydro lyase	C–C–OH ↔ C=C + (HO)
25	5.1	epimerase	
26	5.3.1	isomerase (aldose to ketose)	HO–C–C=O ↔ O=C–C–OH
27	5.3.2	isomerase (keto to enol)	O=C–C ↔ HO–C=C
28	5.3.3	isomerase (enol to enol)	C–C=C ↔ C=C–C
29	5.4.2	phospho-transferase	R–O–PO_3 + R–OH ↔ R–OH + R–O–PO_3
30	6.2.1	acid-thiol ligases with CoA	O=C–S–CoA + (HO) ↔ O=C–OH + (CoA)

are required to convert one to the other. The neighboring compounds a and e are not included in the module, since each one has a possible shortcut from c, meaning that the naturally occurring pathways (a ↔ b ↔ c and c ↔ d ↔ e) are not minimal. In defining minimality modules, the direction of the reactions is not important.

All of the minimality modules in the central carbohydrate metabolism network are depicted in Figure 5A. These minimality modules are rather small. Wherever a module begins and ends, a shortcut was found by the algorithm (an explanation using two modules from glycolysis is given in Figures 5B and 5C).

Consecutive Precursor Metabolites Are Connected by Minimal Paths

We next asked, why is it that central carbon metabolism seems to employ more enzyme steps than the minimal number predicted by the shortcuts? We find that all of the putative shortcuts found by the present algorithm skip branchpoints, metabolites where carbon flows out of the network to synthesize biomass. The branchpoint metabolites are indicated by bold squares in Figure 2, along with the names of the anabolic pathways supplied by each branchpoint metabolite (Table 1).

There are 12 such branchpoint metabolites that serve as exit points for carbon from the central carbon metabolism network we consider. They were previously named precursor metabolites (Neidhardt et al., 1990) (Table 1), because they are the source material for all biomass components made by *E. coli*: amino acids, RNA and DNA nucleotides, lipids, LPS, and peptidoglycan monomers. The nonprecursor metabolites in the present network (denoted by thin rectangles in Figure 2) do not participate in forming biomass directly, but rather act as metabolic steps that connect precursors. A shortcut that skips a branchpoint metabolite loses the ability to synthesize the biomass components whose anabolic pathways originate from that metabolite. In contrast, a hypothetical shortcut that only skips nonbranchpoint metabolites would still allow the cell to make all biomass components.

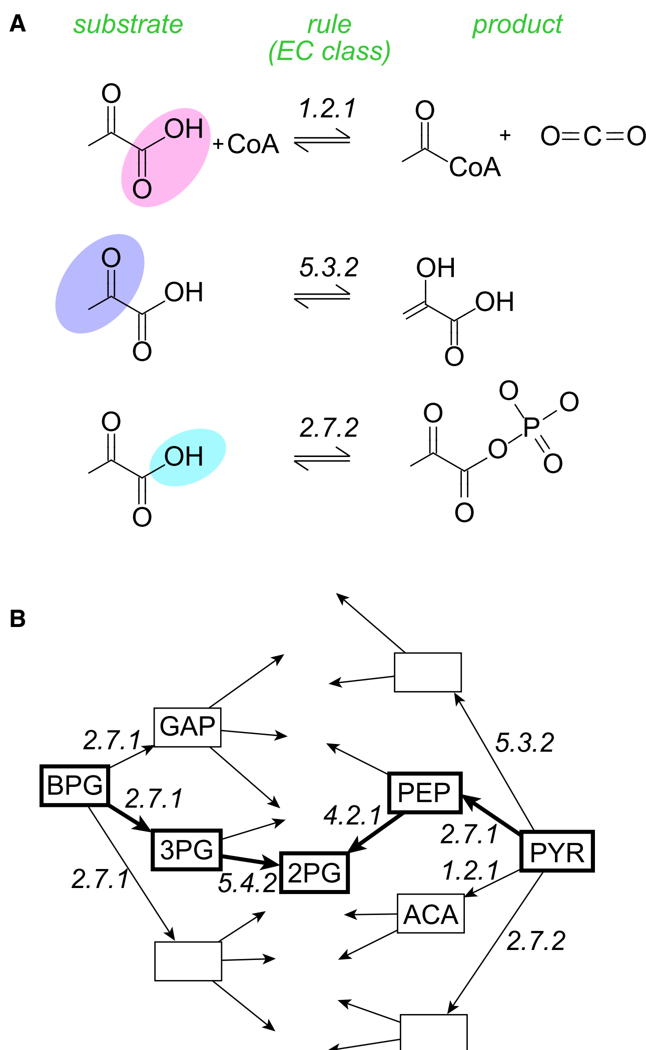

Figure 3. Rules for Finding the Shortest Possible Path that Connects between Two Compounds, According to Known Enzyme Classes

Given a compound *X*, one searches for a subpart in *X* that is an input to an enzyme class, and accordingly modifies *X*. In this way, a large number of new molecules can be generated from each *X*.

(A) For example, to generate a new molecule from PYR, enzyme class 1.2.1 can act on its COOH moiety (magenta) and add CoA to generate ACA and CO. Enzyme class 5.3.2 can act on C-C=O (violet) to generate enolpyruvate. Enzyme class 2.7.2 can act on C-OH (cyan) and add phosphate to generate PEP.

(B) The minimal path between two compounds is found by a bidirectional search on the tree of possible products generated by the enzyme classes. Each node describes a compound. Lines between nodes are reactions generated by EC enzyme classes. All possible reactions are found according to the syntactic rules of Table 2. A minimal path (in bold) is found when the trees emanating from *X* and *Y* join. The example shown is for *X* = BPG (glycerate-1,3P) and *Y* = PYR (pyruvate). Only intermediates with known names are shown: GAP (D-glyceraldehyde-3-phosphate), 3PG (glycerate-3P), 2PG (glycerate-2P), PEP (phosphoenolpyruvate), and ACA (acetyl-CoA).

Another way to state the present finding is that all of the branchpoint metabolites (precursor metabolites) are connected to each other along central carbon metabolism by minimal enzymatic paths (Table 3). In other words, for every pair of consecutive precursors, there is a minimality module that contains both. Paths between consecutive precursors cannot be made shorter with the present enzyme classes. Thus, central carbon metabolism is similar to a minimal walk between precursors.

This feature is perhaps most evident in the TCA cycle and pentose phosphate pathways in which most of the metabolites are nonprecursors. Most of the structure of these parts of metabolism is made of the minimal paths between precursors. In contrast, glycolysis is made of mostly precursors (8 out of 11 compounds are precursors), and minimal paths are easier to make.

There is only one exception in which two consecutive biomass precursors are connected by a path longer than the minimal path, but this exception illuminates the rule. The exception is the two-enzyme path in glycolysis from the precursor GAP through the non-precursor molecule BPG (glycerate-1,3P) to the precursor 3PG. A shorter possible path between the precursors GAP and 3PG exists, as mentioned above, (using an aldehyde dehydrogenase—EC class 1.2, see Figure 4). This putative path does not pass through BPG. One might thus ask why BPG exists in the network, as it is not a precursor for biomass. Although it is not a biomass precursor, BPG serves a unique and crucial role in pathway energetics: it is the only glycolytic step where inorganic phosphate is used to phosphorylate a compound and this phosphate is immediately used to produce ATP from ADP. Without this step, glycolysis would not have a net ATP gain. Thus, one may expand the rule that consecutive precursors are connected by minimal paths, if we consider BPG to have a special "energy precursor" status similar to the biomass precursors.

The principle that some shortcuts can be ruled out when we add a requirement for specific byproducts applies when we consider shunts in central carbon metabolism, such as the glyoxylate bypass. We briefly consider this to highlight the intricacies of the present computations. The glyoxylate shunt connects D-isocitrate to succinate, bypassing the two CO_2 evolving steps (D-isocitrate to 2KG and 2KG to SCA), and another ATP producing step (SCA to succinate). In this transformation the byproduct of the reaction is glyoxylate, not two molecules of CO_2. When we ask whether the path in the TCA cycle (not the shunt) from D-isocitrate to succinate is the shortest possible, it is always in the context of the byproducts (two CO_2 molecules). Therefore, the three-step path through the TCA cycle is still the shortest possible (with CO_2 byproducts), despite the fact that the first and last compounds can be connected by a shorter path with different byproducts.

Two of the predicted shortcuts correspond to actual pathways expressed in *E. coli* under special conditions. When expressed, these pathways work in parallel to the central carbon metabolic reactions. These are the Entner-Doudoroff and the methylglyoxal pathways (Figure 4, shortcuts 1 and 3). These shortcuts thus do not abolish the cell's ability to produce biomass, because central metabolism is still at work in parallel, generating the needed precursors. Both the Entner-Doudoroff and the methylglyoxal pathways are minimal solutions, under the present EC classes, to the problem of connecting their source and end products.

Cell PRESS

Table 3. The Exponentially Increasing Number of Possible Pathways and Products

The Number of Possible Products from a Substrate Compound

Substrate	Number of Steps		
	One	Two	Three
ACA	14	199	2227
Fumarate	23	460	6418
PEP	25	411	5139
PYR	25	423	5252
3PG	42	956	13084
GAP	43	987	14367
OXA	44	902	12685

The Number of Distinct Paths between Pairs of Compounds

Substrate	Product	Number of Steps			
		One	Two	Three	Four
GAP	3PG	1	5	144	1152
3PG	PEP	0	1	49	302
fumarate	OXA	0	3	13	402
SCA	OXA	0	0	0	74

Pathway Lengths between Pairs of Consecutive Precursors

Pathway	Substrate/s	Product/s	Number of Steps in Natural Path	Number of Steps in Minimal Path
Glycolysis	G6P	F6P	1	1
Glycolysis	F6P	2 x GAP	3	3
Glycolysis	GAP	3PG	2	1
Glycolysis	3PG	PEP	2	2
Glycolysis	PEP	PYR	1	1
Glycolysis	PYR	ACA + CO_2	1	1
TCA cycle	OXA + ACA	2KG + CO_2	4	4
TCA cycle	2KG	SCA + CO_2	1	1
TCA cycle	SCA	OXA	4	4
Anaplerotic	PEP + CO_2	OXA	1	1
Pentose phosphate	G6P	R5P + CO_2	4	4
Pentose phosphate	R5P + D-xylulose-5P	F6P + E4P	2	2
Pentose phosphate	E4P + D-xylulose-5P	F6P + GAP	1	1
Entner-Doudoroff	G6P	GAP + PYR	4	4
Sugar catabolism	ribitol	D-ribulose-5P	2	2
Sugar catabolism	L-xylulose	D-xylulose-5P	3	3
Sugar catabolism	L-arabinose	D-xylulose-5P	3	3
Sugar catabolism	D-xylose	D-xylulose-5P	2	2
Sugar catabolism	D-arabitol	D-xylulose-5P	2	2
Sugar catabolism	D-ribose	R5P	1	1
Sugar catabolism	D-glucose	G6P	1	1
Sugar catabolism	D-fructose	D-fructose-1,6P	2	2
Sugar catabolism	D-mannose	F6P	2	2
Sugar catabolism	acetate	ACA	1	1
Sugar catabolism	glycerol	glycerone-P	2	2

The number of possible products from a substrate compound is multiplied roughly by a factor of 20 at every step of the path leading to the final compound. The number of distinct paths between two compounds grows exponentially as a function of their length. However, pairs of consecutive precursors are minimally connected. The only exception is the path from GAP to 3PG in glycolysis, since the intermediate BPG is sometimes needed for ATP production from inorganic phosphate and ADP.

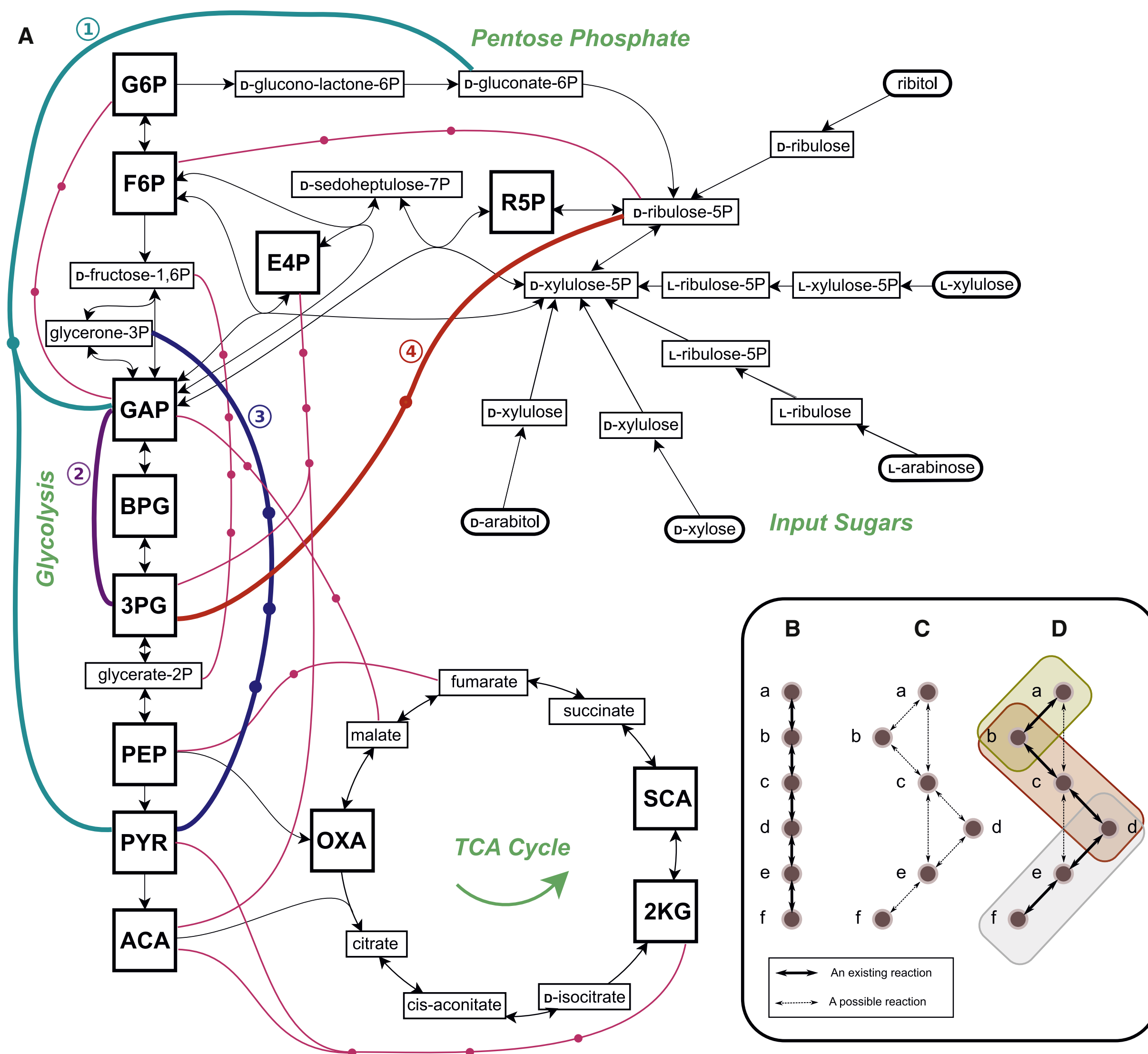

Figure 4. Some Pathways Could Potentially Have Shortcuts

(A) Putative shortcuts found by the algorithm all bypass at least one precursor. The colorful lines are shortcuts made of EC class enzymes. Precursor metabolites are in bold squares (for the full names, see Table 1). Input sugars are in bold rounded boxes. Four of the shortcuts exist in nature (thick lines and circled numbers): (1) the Entner-Doudoroff bypass, (2) nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase (*gapN*) in plants, (3) the methylglyoxal pathway, and (4) the CO_2 fixating branch in the Calvin cycle in plants. Shortcuts 1 and 3 exist in *E. coli* and act in parallel to central metabolism. They are active only under special conditions.

(B–D) Minimality modules contain groups of compounds connected by paths that cannot be made shorter. (B) Consider a hypothetical pathway between (Ba) and (Bf), with five reactions that exist in natural metabolism. (C) All possible reactions using these compounds, according to the EC classes from Table 2. Note that both the naturally occurring reactions and other possible reactions are depicted. Contrasting these two networks, the "real" and the "possible," defines the optimality modules—those sets of consecutive compounds that are connected by the shortest path in the "real" network. (D) The minimality modules (shown in shaded rectangles) are the largest sets of compounds among which all paths are minimal.

Environmentally Available Sugars Are Connected to Central Metabolism by Minimal Paths

We also studied the metabolic paths that degrade the sugars and other carbon sources on which *E. coli* can grow. Each of these carbon sources stands at the input of a pathway that feeds into central metabolism. We find that these pathways are the minimal pathways needed to connect the carbon source with central carbon metabolism. For example, L-arabinose is converted in three steps to the pentose-phosphate entry port D-xylulose-5P. No shorter path is possible within the present rules. This minimality principle applies to all environmentally available carbon sources that we tested, including glucose, fructose, mannose, acetate, and glycerol (not depicted in Figure 5).

We note that when two such pathways overlap, as when the L-xylulose and L-arabinose pathways share the same intermediate reaction (L-ribulose-5P ↔ D-xylulose-5P), the overlapping

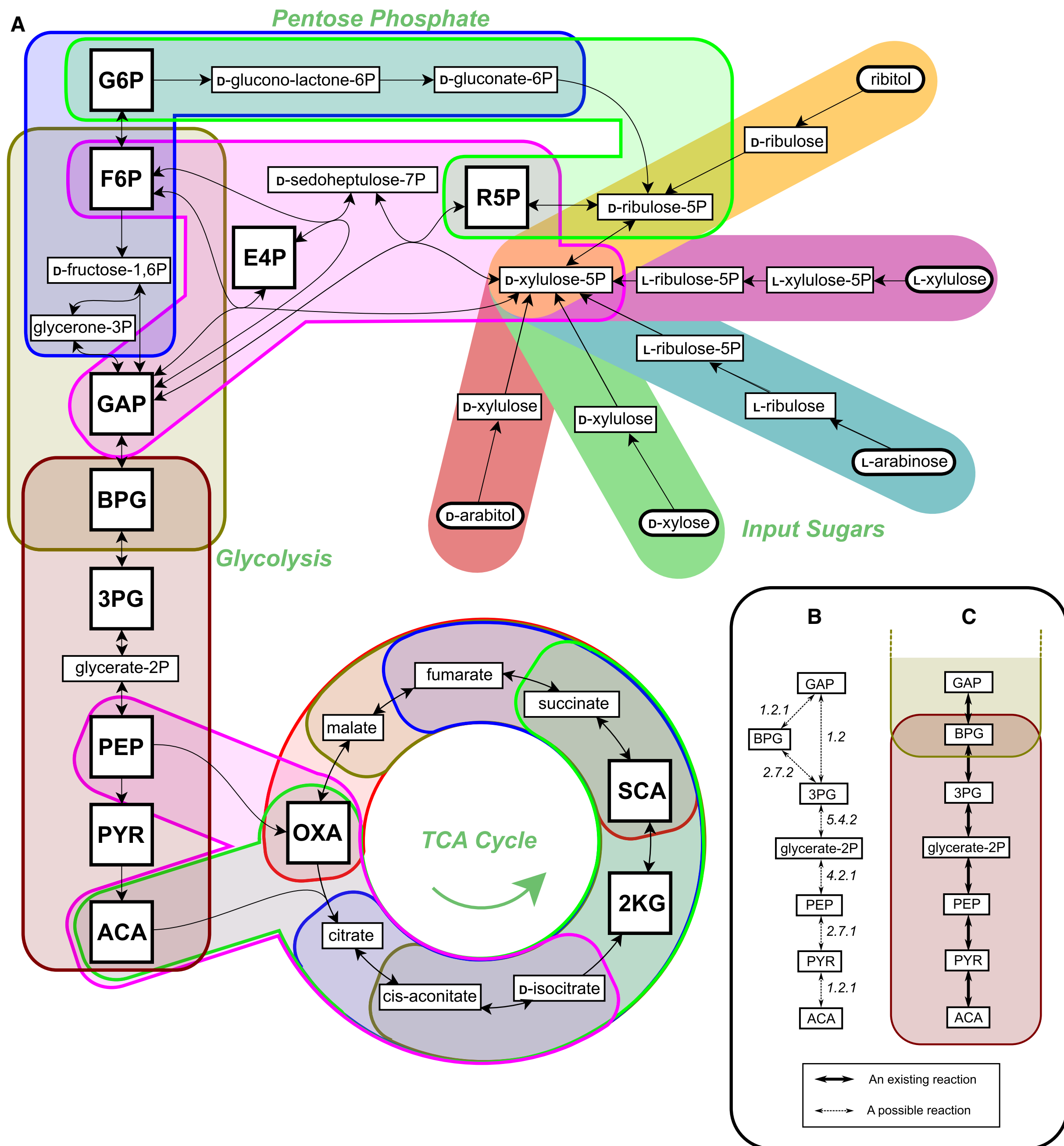

Figure 5. Minimality Modules Are Pieces of Metabolism within which All Paths Are Minimal

(A) Minimality modules in the central carbohydrate metabolic network. Minimality modules are shown in colored shapes. Glycolysis, for example, is composed of three overlapping modules: G6P to glycerone-3P (blue), F6P to BPG (brown), and BPG to PYR (red). The first and last modules include other compounds as well. The TCA cycle is composed of five modules. Other modules correspond to the pentose phosphate pathway (violet), the pentose phosphate shunt (green), and systems for sugar catabolism (solid colors).

(B and C) Glycolysis displays several minimality modules. (B) Compounds from the glycolysis pathway including some of the possible reactions connecting them. (C) Two of the minimality modules (one ending at BPG and one starting from BPG) in glycolysis. 3PG is not included in the first module, because a shorter pathway (using enzyme class 1.2—aldehyde dehydrogenase) exists between GAP and 3PG, a pathway found in plant metabolism, but not in *E. coli*.

steps are catalyzed by distinct enzymes, expressed from different genes with different regulation (in this example *araD* in the L-arabinose module and *yiaS* in the L-xylulose module). This is presumably due to the need to differentially regulate the enzymes based on the availability of their cognate sugar. It thus seems that cells do not economize by sharing the genes that encode such enzymes, but rather devote a distinct gene per pathway separately.

DISCUSSION

This study suggests that the structure of central carbon metabolism of *E. coli* can be viewed by means of an optimality principle: it is a minimal walk between key precursor compounds, and between input sugars and precursor compounds. Since the precursors are essential for building biomass (and one is essential for positive net ATP in glycolysis), they cannot be bypassed. All other compounds, called nonprecursor compounds, can be bypassed without losing the ability to generate biomass. The nonprecursor compounds in the network form the shortest possible bridges between the precursors, when considering all of the possible combinations of allowed EC class transformations.

Why should steps between input sugars and central carbon metabolism, and steps between precursors along central carbon metabolism, be minimized? This minimization seems to align with the constraints of the organism. *E. coli* growth rate is often limited by protein synthesis (Kurland and Dong, 1996; Marr, 1991). Experiments have shown that cells that make unneeded protein grow slower (Andrews and Hegeman, 1976; Koch, 1983; Dong et al., 1995; Dekel and Alon, 2005; Stoebel et al., 2008; Shachrai et al., 2010) and that the limit on the concentration of protein in the cell bounds the achievable metabolic states (Beg et al., 2007). Thus, cells with shorter pathways may have a competitive advantage due to their economy in proteins. Furthermore, short pathways have fewer intermediates and generate higher flux than long pathways of equally effective enzymes (Meléndez-Hevia et al., 1994).

One caveat should be noted: a long pathway can in principle exist that uses, due to the nature of its reactions, highly efficient enzymes. One can imagine a situation in which the enzymes are so efficient that the long pathway requires less total enzyme mass than a short pathway to achieve the same flux (Heinrich and Klipp, 1996). However, no such example has been found in the present study, where a putative pathway is shorter than an existing pathway between precursors. Further studies can test the hypothesis that this minimality had evolved under the selection pressure to shorten paths between essential metabolites (Cornish-Bowden, 2004; Pal et al., 2006).

One possible prediction, if this principle holds more generally, is that in organisms where one of the precursors is no longer essential (e.g., it is always supplied from the environment, or its end products are), a shortcut would evolve that bypasses that precursor compound. One case where this may have occurred is in plants in which a shortcut that bypasses BPG exists. This shortcut is predicted by the present algorithm (shortcut 2 in Figure 4) using EC class 1.2. The plant indeed employs this EC class, specifically the enzyme nonphosphorylating GAP dehydrogenase (EC 1.2.1.9). The metabolite bypassed in this shortcut, BPG, is an energy precursor metabolite essential for glycolysis to have positive net ATP production. Plants are able to produce ATP using photophosphorylation, and thus the need for BPG as an energy precursor may be less crucial than in nonphotosynthetic organisms such as *E. coli*. Other cases where shortcuts might be expected are in obligate parasites that are provided with biomass building blocks. This prediction can be tested in future research.

The Entner-Doudoroff bypass is another good example for a shortcut that exists in many organisms but is not generally used in the core carbon metabolism of *E. coli.* This pathway occurs in diverse prokaryotes and converts G6P to GAP and PYR in only four steps: an alcohol dehydrogenase (**1.1**.1.49), a carboxyl-ester hydrolyase (**3.1.1**.31), a hydro lyase (**4.2.1**.12), and an aldehyde lyase (**4.1.2**.14). Although this pathway is minimal and shorter than glycolysis (which requires nine steps to do the same), it skips the precursors between G6P and PYR (such as F6P). This might be the reason why the Entner-Doudoroff bypass is not used under most conditions by *E. coli* (Zablotny and Fraenkel, 1967), whereas glycolysis is ubiquitous.

The present study also addresses potential cases of longer-than-minimal pathways. If a longer-than-minimal path is found between two compounds in a certain cell type, the present approach predicts that an essential metabolite for that particular cell type lies on that path. This can suggest a search for new branching pathways that begin at that essential metabolite, or for new biological functions of that metabolite.

What further principles can be found in the structure of central carbon metabolism? One question is what defines the particular sequence of reactions of each minimal pathway. Our study suggests that most pairs of precursors separated by more than one step could, in principle, have been connected by several other alternative possible paths of the same length (but not shorter). Many of these alternatives use the same steps but in different order, and some of the alternatives use different steps altogether. It would be interesting to ask why the particular minimal path that occurs in the organism was selected out of these multiple, equally short alternatives. Possibilities to explore include effects that can differentiate between paths of equal length, such as energy and reduction potential production, toxicity effects of intermediate compounds, and differential enzyme efficiency in each possible path.

An additional question relevant to this study is whether there exist additional enzyme classes that have yet to be discovered. One can in principle expand the present set of EC enzymes by adding thermodynamically feasible reactions beyond those catalyzed by known enzymes. It would be interesting to study how many additional reaction classes can be added before the observed minimality feature fails to hold.

In summary, this study suggests that central carbon metabolism in *E. coli* uses the minimal number of enzymes to traverse between input sugars and key precursor metabolites essential for biomass and energy production. We provide an algorithm that employs a comprehensive set of enzyme classes to generate a large number of possible biochemical paths between compounds. The present study may be extended to test for minimal paths in other networks and perhaps to design new optimal pathways in bacteria to produce or break down desired compounds. It would be interesting to see whether the optimality game might capture the structure of other parts of metabolism in bacteria and other organisms.

SUPPLEMENTAL INFORMATION

Supplemental Information includes supplemental text and can be found with this article at doi:10.1016/j.molcel.2010.08.031.

ACKNOWLEDGMENTS

We thank M. Kirschner, M. Springer, A. Bar-Even, A. Mayo, N. Kashtan, D. Koster, E. Dekel, I. Shachari, and Y. Hart for comments and discussions. We thank the Kahn Family Foundation, the Radcliffe fellowship, and the Meil de Botton Aynsley Foundation for support; and the Department of Systems Biology at Harvard Medical School for hospitality.

Received: September 30, 2009
Revised: February 10, 2010
Accepted: August 20, 2010
Published: September 9, 2010

REFERENCES

Alberty, R.A. (2005). Thermodynamics of Biochemical Reactions (New York: Wiley-Interscience).

Alon, U. (2007). An Introduction to Systems Biology: Design Principles of Biological Circuits (Boca Raton, FL: Chapman & Hall/CRC).

Andrews, K.J., and Hegeman, G.D. (1976). Selective disadvantage of non-functional protein synthesis in Escherichia coli. J. Mol. Evol. *8*, 317–328.

Beg, Q.K., Vazquez, A., Ernst, J., de Menezes, M.A., Bar-Joseph, Z., Barabási, A., and Oltvai, Z.N. (2007). Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. Proc. Natl. Acad. Sci. USA *104*, 12663–12668.

Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J., and Rabinowitz, J.D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat. Chem. Biol. *5*, 593–599.

Cornish-Bowden, A. (2004). The Pursuit of Perfection: Aspects of Biochemical Evolution (New York: Oxford University Press).

Dekel, E., and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. Nature *436*, 588–592.

Dong, H., Nilsson, L., and Kurland, C.G. (1995). Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. J. Bacteriol. *177*, 1497–1504.

Ebenhöh, O., and Heinrich, R. (2001). Evolutionary optimization of metabolic pathways. Theoretical reconstruction of the stoichiometry of ATP and NADH producing systems. Bull. Math. Biol. *63*, 21–55.

Ebenhöh, O., and Heinrich, R. (2003). Stoichiometric design of metabolic networks: multifunctionality, clusters, optimization, weak and strong robustness. Bull. Math. Biol. *65*, 323–357.

Fell, D. (1996). Understanding the Control of Metabolism, First Edition (London: Portland Press).

Fell, D.A., and Wagner, A. (2000). The small world of metabolism. Nat. Biotechnol. *18*, 1121–1122.

Finley, S.D., Broadbelt, L.J., and Hatzimanikatis, V. (2009). Thermodynamic analysis of biodegradation pathways. Biotechnol. Bioeng. *103*, 532–541.

Handorf, T., and Ebenhöh, O. (2007). MetaPath Online: a web server implementation of the network expansion algorithm. Nucleic Acids Res. *35*, W613–W618. Published online May 5, 2007. 10.1093/nar/gkm287.

Handorf, T., Ebenhöh, O., and Heinrich, R. (2005). Expanding metabolic networks: scopes of compounds, robustness, and evolution. J. Mol. Evol. *61*, 498–512.

Hatzimanikatis, V., Li, C., Ionita, J.A., Henry, C.S., Jankowski, M.D., and Broadbelt, L.J. (2005). Exploring the diversity of complex metabolic networks. Bioinformatics *21*, 1603–1609.

Heinrich, R., and Klipp, E. (1996). Control analysis of unbranched enzymatic chains in states of maximal activity. J. Theor. Biol. *182*, 243–252.

Heinrich, R., and Schuster, S. (1996). The Regulation of Cellular Systems (New York: Chapman & Hall).

Heinrich, R., and Schuster, S. (1998). The modeling of metabolic systems. Structure, control and optimality. Biosystems *47*, 61–77.

Heinrich, R., Montero, F., Klipp, E., Waddell, T.G., and Melendez-Hevia, E. (1997). Theoretical approaches to the evolutionary optimization of glycolysis. Eur. J. Biochem. *243*, 191–201.

Janga, S.C., and Babu, M.M. (2008). Network-based approaches for linking metabolism with environment. Genome Biol. *9*, 239.

Jankowski, M.D., Henry, C.S., Broadbelt, L.J., and Hatzimanikatis, V. (2008). Group contribution method for thermodynamic analysis of complex metabolic networks. Biophys. J. *95*, 1487–1499.

Karp, P.D., Keseler, I.M., Shearer, A., Latendresse, M., Krummenacker, M., Paley, S.M., Paulsen, I., Collado-Vides, J., Gama-Castro, S., Peralta-Gil, M., et al. (2007). Multidimensional annotation of the Escherichia coli K-12 genome. Nucleic Acids Res. *35*, 7577–7590.

Koch, A.L. (1983). The protein burden of lac operon products. J. Mol. Evol. *19*, 455–462.

Kurland, C.G., and Dong, H. (1996). Bacterial growth inhibition by overproduction of protein. Mol. Microbiol. *21*, 1–4.

Marr, A.G. (1991). Growth rate of Escherichia coli. Microbiol. Mol. Biol. Rev. *55*, 316–333.

Meléndez-Hevia, E., and Isodoro, A. (1985). The game of the pentose phosphate cycle. J. Theor. Biol. *117*, 251–263.

Meléndez-Hevia, E., Waddell, T.G., and 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*, 201–219.

Meléndez-Hevia, E., Waddell, T.G., and 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.

Mittenthal, J.E., Yuan, A., Clarke, B., and Scheeline, A. (1998). Designing metabolism: alternative connectivities for the pentose phosphate pathway. Bull. Math. Biol. *60*, 815–856.

Mittenthal, J.E., Clarke, B., Waddell, T.G., and Fawcett, G. (2001). A new method for assembling metabolic networks, with application to the Krebs citric acid cycle. J. Theor. Biol. *208*, 361–382.

Neidhardt, F.C., Ingraham, J.L., and Schaechter, M. (1987). Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology (Washington, D.C.: American Society for Microbiology).

Neidhardt, F.C., Ingraham, J.L., and Schaechter, M. (1990). Physiology of the Bacterial Cell: A Molecular Approach (Sunderland, MA: Sinauer Associates).

Pal, C., Papp, B., Lercher, M.J., Csermely, P., Oliver, S.G., and Hurst, L.D. (2006). Chance and necessity in the evolution of minimal metabolic networks. Nature *440*, 667–670.

Palsson, B.O. (2006). Systems Biology: Properties of Reconstructed Networks (New York: Cambridge University Press).

Papin, J.A., Stelling, J., Price, N.D., Klamt, S., Schuster, S., and Palsson, B.O. (2004). Comparison of network-based pathway analysis methods. Trends Biotechnol. *22*, 400–405.

Price, N.D., Reed, J.L., Papin, J.A., Wiback, S.J., and Palsson, B.O. (2003). Network-based analysis of metabolic regulation in the human red blood cell. J. Theor. Biol. *225*, 185–194.

Raymond, J., and Segre, D. (2006). The effect of oxygen on biochemical networks and the evolution of complex life. Science *311*, 1764–1767.

Savageau, M.A. (1976). Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology (Reading, MA: Addison Wesley Publishing Company).

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

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

Shachrai, I., Zaslaver, A., Alon, U., and Dekel, E. (2010). Cost of unneeded proteins in *E. coli* is reduced after several generations in exponential growth. Mol. Cell *38*, 758–767.

Stephani, A., and 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., and Heinrich, R. (1999). Optimal stoichiometric designs of ATP-producing systems as determined by an evolutionary algorithm. J. Theor. Biol. *199*, 45–61.

Stoebel, D.M., Dean, A.M., and Dykhuizen, D.E. (2008). The cost of expression of Escherichia coli lac operon proteins is in the process, not in the products. Genetics *178*, 1653–1660.

Stryer, L. (1995). Biochemistry, Fourth Edition (New York: W.H. Freeman & Company).

Webb, E.C. (1992). Enzyme Nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes (San Diego: Academic Press).

Zablotny, R., and Fraenkel, D.G. (1967). Glucose and gluconate metabolism in a mutant of Escherichia coli lacking gluconate-6-phosphate dehydrase. J. Bacteriol. *93*, 1579–1581.