Glycolytic strategy as a tradeoff between energy yield and protein cost

Avi Flamholz[a,1], Elad Noor[a,1], Arren Bar-Even[a], Wolfram Liebermeister[a,b], and Ron Milo[a,2]

[a]Department of Plant Sciences, The Weizmann Institute of Science, Rehovot 76100, Israel; and [b]Institut für Biochemie, Charité–Universitätsmedizin Berlin, 10117 Berlin, Germany

Edited by Richard E. Lenski, Michigan State University, East Lansing, MI, and approved April 4, 2013 (received for review September 17, 2012)

Contrary to the textbook portrayal of glycolysis as a single pathway conserved across all domains of life, not all sugar-consuming organisms use the canonical Embden–Meyerhoff–Parnass (EMP) glycolytic pathway. Prokaryotic glucose metabolism is particularly diverse, including several alternative glycolytic pathways, the most common of which is the Entner–Doudoroff (ED) pathway. The prevalence of the ED pathway is puzzling as it produces only one ATP per glucose—half as much as the EMP pathway. We argue that the diversity of prokaryotic glucose metabolism may reflect a tradeoff between a pathway's energy (ATP) yield and the amount of enzymatic protein required to catalyze pathway flux. We introduce methods for analyzing pathways in terms of thermodynamics and kinetics and show that the ED pathway is expected to require several-fold less enzymatic protein to achieve the same glucose conversion rate as the EMP pathway. Through genomic analysis, we further show that prokaryotes use different glycolytic pathways depending on their energy supply. Specifically, energy-deprived anaerobes overwhelmingly rely upon the higher ATP yield of the EMP pathway, whereas the ED pathway is common among facultative anaerobes and even more common among aerobes. In addition to demonstrating how protein costs can explain the use of alternative metabolic strategies, this study illustrates a direct connection between an organism's environment and the thermodynamic and biochemical properties of the metabolic pathways it employs.

evolution | enzyme cost

Glycolysis is the process by which glucose is broken down anaerobically into incompletely oxidized compounds like pyruvate, a process which is usually coupled to the synthesis of ATP. Although the Embden–Meyerhof–Parnas pathway (EMP, often simply "glycolysis") is the nearly ubiquitous glycolytic route among eukaryotes (1, 2), it is not the only game in town. Prokaryotes display impressive diversity in glucose metabolism (2, 3) and natural glycolytic alternatives like the Entner–Doudoroff (ED), and phosphoketolase pathways attest to the fact that there are multiple biologically feasible routes for glucose metabolism (2, 4–8). Natural glycolytic pathways vary in their reaction sequence and in how much ATP they produce per glucose metabolized, ranging from zero to three ATP molecules in most cases (7).

The EMP and ED pathways (Fig. 1 *A* and *B* and Fig. S1) are the most common bacterial glycolytic pathways (2, 4, 9), and their general schemes are quite similar: glucose is phosphorylated and then cleaved into two three-carbon units, which are further metabolized to produce ATP (4). In some organisms, these pathways differ slightly in the specific redox cofactors they use (e.g., NAD^+ vs. $NADP^+$; Fig. 1*B*, Fig. S2, and Table S1), but here we focus on the prominent difference in ATP yield. If we take lactate as a representative final product, these two pathways have the same net reaction:

$$\text{D-glucose} + n\ \text{ADP} + n\ \text{P}_\text{i} \rightarrow 2\ \text{L-lactate} + n\ \text{ATP} + n\ \text{H}_2\text{O},$$

and differ primarily in n, the number of ATP produced, and the specific intermediate reaction steps (Fig. 1*B*, *SI Text*). As shown in Fig. 1 *A* and *B*, the pathways overlap, sharing the reaction sequence from glyceraldehyde 3-phosphate (G3P) through pyruvate known as "lower glycolysis." In the EMP pathway, glucose is phosphorylated twice and cleaved into two triose-phosphates (G3P and dihydroxyacetone phosphate), both of which are used to produce ATP through substrate-level phosphorylation in lower glycolysis (2, 7) (Fig. 1*B*). In the ED pathway, glucose is phosphorylated only once and oxidized to 2-keto-3-deoxy-6-phosphogluconate (KDPG), which is cleaved into one pyruvate and one G3P. Pyruvate does not support substrate-level phosphorylation (7) and so, in the ED pathway, only one of the cleavage products (G3P) is used to produce ATP through lower glycolysis. From a chemical perspective, the ED pathway represents a simple rearrangement of the EMP pathway wherein the reactions of lower glycolysis are performed on the six-carbon backbone in the upper portion of the ED pathway (Fig. 1*C*). While these transformations are quite exergonic and are coupled to ATP production in lower glycolysis (7), no ATP is produced in the upper portion of the ED pathway (Fig. 1*B*). As a result, the EMP pathway produces two ATP per glucose while the ED produces only one (Fig. 1 *A*–*C*).

Naïvely, it would seem that the EMP pathway is strictly superior, yielding twice as much ATP as the ED pathway. This begs the question: why is the ED pathway common? It has been argued that the ED pathway's primary function is not glucose metabolism, but rather the breakdown of sugar acids like gluconate that cannot be metabolized through the EMP pathway (10). Indeed, *Escherichia coli* strains lacking ED enzymes cannot grow on gluconate and are not capable of colonizing the mammalian large intestine (11). Moreover, because variants of the ED pathway appear in some archaea, it is often considered an example of "paleo-metabolism"—a historical artifact (1, 5). However, many bacteria are known to use the ED pathway to metabolize glucose, including *Pseudomonas saccharophila* (wherein the pathway was discovered), *Zymomonas mobilis*, and most pseudomonads (4, 6). In a striking study, Fuhrer et al. measured metabolic fluxes in seven diverse bacteria including aerobes and anaerobes, autotrophs and heterotrophs, and found that all seven rely on the ED pathway for glucose catabolism, even those harboring genes coding for EMP enzymes (9). If the EMP pathway were truly superior, why wouldn't these bacteria use it?

There is probably no such thing as an optimal glycolytic pathway (12). Rather, different pathways likely suit the demands of different environments or physiologies. The primary difference between these two pathways is energetic: the ED pathway yields

Author contributions: A.F., E.N., A.B.-E., and R.M. designed research; A.F., E.N., A.B.-E., and W.L. performed research; A.F., E.N., A.B.-E., and R.M. analyzed data; and A.F., E.N., W.L., and R.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

See Commentary on page 9629.

[1]A.F. and E.N. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: ron.milo@weizmann.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1215283110/-/DCSupplemental.

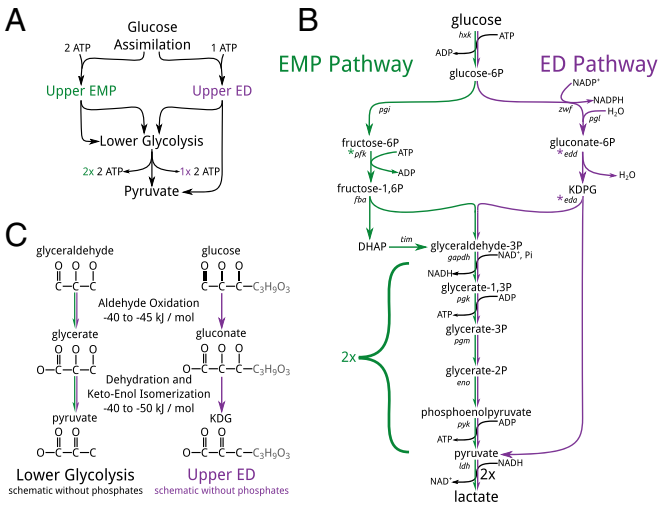

Fig. 1. Structural similarity and energetic differences between the ED and EMP pathways. (*A*) The ED (purple) and EMP pathways (green) overlap but differ in ATP yield. The EMP pathway hydrolyzes two ATP to phosphorylate glucose twice and recovers four ATP by metabolizing two triose-phosphates through lower glycolysis, yielding two ATP in total. In contrast, the ED pathway invests one ATP in phosphorylation and recovers two (glucose is cleaved into only one fermentable product) yielding one ATP per glucose. (*B*) A schematic of the ED and EMP pathways assuming that glucose is phosphorylated intracellularly by hexokinase and lactate is the final product. These pathways each contain unique enzymes (marked with *) but also share all of the reactions of lower glycolysis (from G3P through pyruvate). pfk is unique to the glycolytic direction of the EMP pathway, while the edd and eda enzymes are unique to the ED pathway. (*C*) Ignoring phosphorylation, lower glycolysis and the upper portion of the ED pathway are composed of the same, highly exergonic reaction sequence. Abbreviations: eda, kdpg aldolase; edd, phosphogluconate dehydratase; eno, enolase; fba, fructose bisphosphate aldolase; gapdh, glyceraldehyde 3-phosphate dehydrogenase; hxk, hexokinase; ldh, lactate dehydrogenase; pfk, 6-phosphofructokinase; pgi, phosphoglucose isomerase; pgk, phosphoglycerate kinase; pgl, phosphogluconolactonase; pgm, phosphoglycerate mutase; pyk, pyruvate kinase; tim, triosephosphate isomerase; zwf, glucose 6-phosphate dehydrogenase.

less ATP and, as such, contains more exergonic reactions than the EMP pathway (13, 14). Here we introduce methodologies for thermodynamic and kinetic analysis of metabolic pathways that allow us to investigate how these energetic differences affect pathways' operation (15). We find that the EMP pathway is much more thermodynamically constrained than the ED. Due to this energetic difference, our kinetic analysis predicts that the ED pathway requires several-fold less enzymatic protein than the EMP pathway to metabolize the same amount of glucose per second. Since expression of unnecessary protein can limit the growth of microbes (16–18), we expect that organisms that can use the ED pathway will do so to reduce glycolytic protein levels. Consequently, we hypothesize a tradeoff between glycolytic ATP yield and protein cost. Under this hypothesis, organisms that depend on glycolysis for ATP production (e.g., fermentative anaerobes) will tend to use the EMP pathway, while organisms with alternative sources of ATP (e.g., aerobes) will favor the ED pathway. Finally, we show that a genomic analysis of the metabolic capabilities of more than 500 prokaryotes supports this hypothesis. We therefore suggest that an organism's choice of glycolytic pathway reflects an economic calculation carried out through evolution that balances the production of ATP and the synthesis of protein to maximize overall fitness.

Results

ED and EMP Pathways Are Common Among Prokaryotes. We attempt to systematically quantify the prevalence of alternate glycolytic pathways among prokaryotes. Various biochemical assays can show whether an organism uses the ED or EMP pathway for glucose metabolism (4). Indeed, a literature survey indicates that diverse bacteria use the ED pathway for growth on glucose (*SI Text*). However, detailed biochemical testing has been performed for only a small number of microbes. We therefore leverage the hundreds of annotated prokaryotic genomes currently available to estimate the prevalence of different glycolytic pathways.

Although the ED and EMP pathways overlap heavily, several enzymes are unique to each. Specifically, 6-phosphofructokinase (pfk, EC 2.7.1.11) is unique to the EMP pathway, while 6-phosphogluconate dehydratase (edd, EC 4.2.1.12) and KDPG aldolase (eda, EC 4.1.2.14) are unique to the ED pathway (Fig. 1*B*). We therefore term an organism "genetically capable" of a particular glycolytic pathway when its genome contains the unique pathway enzymes and a pyruvate kinase (pyk) (*Materials and Methods*, *SI Text*, and Tables S2 and S3). *Z. mobilis*, for example, is not genetically capable of the EMP pathway because it has no pfk, whereas *Bacillus subtilis* is not capable of the ED pathway because it lacks a functional edd enzyme (9). Fig. 2 shows how ED and EMP capabilities are distributed among heterotrophic prokaryotes. After merging closely related organisms (*Materials and Methods*), 57% of prokaryotes considered are EMP capable, 27% are ED-capable, and 14% are genetically capable of both pathways. The observation that ED pathway is widespread holds when this analysis is repeated for various definitions of pathway capability (more or fewer genes) or a larger sample of microbes (*SI Text* and Fig. S3).

ED Pathway Is Less Thermodynamically Constrained than the EMP Pathway. Although the ED and EMP pathways produce different amounts of ATP, both pathways' net reactions are quite exergonic (13). The net reaction, however, paints an incomplete picture: the pathway may contain thermodynamic bottlenecks that make it difficult for all constituent reactions to be favorable in biological conditions (19–21).

When a reaction is in equilibrium ($\Delta_r G' = 0$), the forward flux (J^+) and reverse flux (J^-) are equal and there is no net flux. A reaction is favorable when its $\Delta_r G' < 0$, and we quantify this favorability by $-\Delta_r G'$ (i.e., how far it is from equilibrium). The ratio between the forward and reverse fluxes increases as the reaction becomes more favorable as determined by the flux-force relationship, $J^+/J^- = \exp(-\Delta_r G'/RT)$, where R is the gas constant and T temperature (22). In principle, any net flux $J = J^+ - J^- > 0$ is achievable for any $\Delta_r G' < 0$. However, achieving a given net flux when near equilibrium requires higher forward and reverse fluxes, demanding larger amounts of enzyme. We quantify this effect using the *Net Flux Ratio* J/J^+—that is, the ratio between the net flux and the forward flux through a reaction. The flux–force relationship relates the Net Flux Ratio to $\Delta_r G'$:

$$\frac{J}{J^+} = 1 - \exp\left(\frac{\Delta_r G'}{RT}\right).$$

This relationship will serve as the basis for analyzing the influence of thermodynamics on the protein investment due to the ED and EMP pathways.

The EMP pathway conserves more of the energy in glucose as ATP than the ED pathway and so must dissipate less Gibbs energy overall, all other factors being equal. To illustrate the thermodynamic differences between these pathways, we apply mathematical optimization tools to investigate pathway favorability given biological constraints on metabolite concentrations, pH, and ionic strength. In brief, we find metabolite concentrations that make the least favorable pathway reaction as favorable as possible, thereby maximizing its Net Flux Ratio (*Materials and Methods*). After optimization, we find that the least favorable EMP reactions have a Net Flux Ratio of ~0.7 (–2.9 kJ/mol), while the least

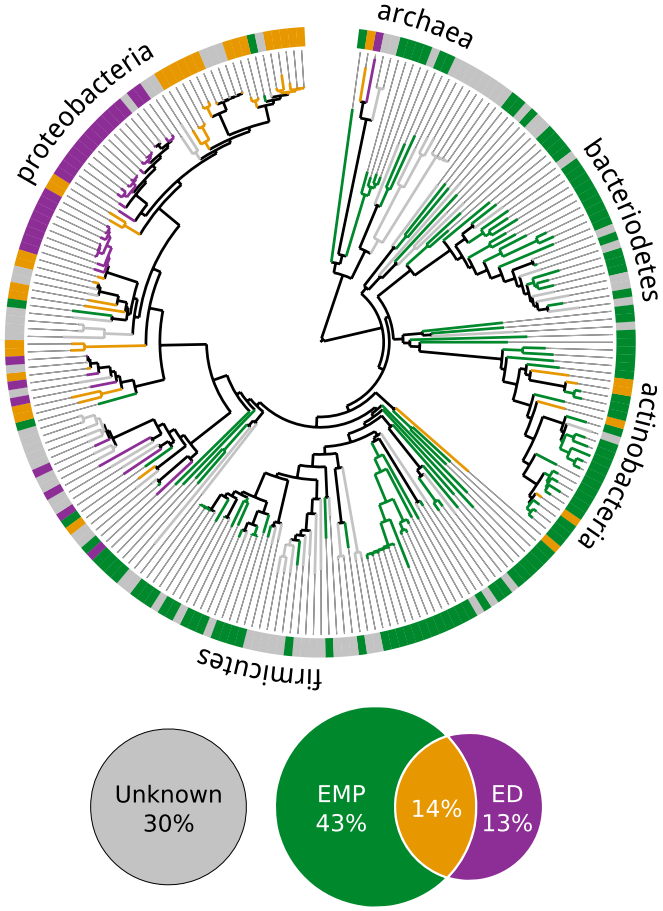

Fig. 2. Phylogenetic analysis of the ED and EMP pathways. Organisms are considered to be ED or EMP capable if their genome is annotated as containing the pathway's unique genes and a pyk (*Materials and Methods*, *SI Text*). Closely related organisms were merged to avoid double-counting, and the resulting distribution of pathway capability is shown on a phylogenetic tree of heterotrophic bacteria and archaea (*Materials and Methods*). Of these microbes, 57% are EMP-capable, while 27% are ED-capable; 14% are genetically capable of both pathways, and we were unable to annotate the remaining 30%. The pathways are largely independent: most organisms were annotated as containing only one pathway (80% of those annotated at all). We identified 10 branching points on the phylogenetic tree where all descendants of one branch uniformly contain the same pathway and the second branch contains descendants with the other pathway. This might suggest that a microbe's choice of glycolytic pathway is due to some selection process.

favorable ED reactions are much further from equilibrium and have a Net Flux Ratio of ~0.9 (–6.4 kJ/mol; Fig. 3). Indeed, the EMP pathway contains a thermodynamic bottleneck, comprising several central reactions [fructose 1,6-bisphosphate aldolase (fba), triose-phosphate isomerase (tim), glyceraldehyde 3-phosphate dehydrogenase (gapdh), phosphoglycerate kinase (pgk), and phosphoglycerate mutase (pgm)] as shown in Fig. 3. The ED pathway avoids this bottleneck by omitting some unfavorable reactions (fba and tim) and using very exergonic reactions instead (at the expense of ATP yield; Fig. 1*C*). The thermodynamic bottleneck in the EMP pathway might be the reason that fructose 1,6-bisphosphate, the substrate of fba, is one of the most abundant metabolites in *E. coli* (~10 mM) and that tim evolved to be an extremely efficient catalyst (k_{cat} ~10,000 s^{-1}) (22, 23).

EMP Pathway Requires Substantially More Enzymatic Protein than the ED Pathway. Thermodynamic considerations push substrate concentrations up and product concentrations down to make reactions more favorable. Thermodynamic favorability is not the only factor affecting catalytic rates, however. Enzymes have limited affinity toward their substrates and limited catalytic rates (24). In particular, substrate concentrations below an enzyme's Michaelis constant (K_M) will severely limit the reaction rate (23). Moreover, the physical constraints on pathway flux are interdependent and sometimes conflicting. In living cells, metabolite concentrations are constrained from above by osmotic pressure and from below by limited enzyme affinity (23, 25). From a kinetic perspective, it is preferable for substrate concentrations to be as high as possible (i.e., above the K_M). If a reaction is sufficiently unfavorable ($\Delta_r G^0 >> 0$), however, it is impossible to satisfy thermodynamic constraints while ensuring that the subsequent enzyme is near saturation. When these constraints conflict, catalytic efficiency is compromised. Cells can synthesize more enzyme to compensate for inefficient catalysis, but at what cost? Increased protein production is thought to impose a significant burden on cells by occupying ribosomes, consuming cellular building blocks, and increasing misfolding (16, 17, 26). We assume that this cost is proportional to the mass of enzyme expressed and calculate the expected enzyme mass associated with a pathway to approximate its cost.

Our model of pathway protein cost uses a rate law derived from the reversible Michaelis–Menten kinetic model (*SI Text*) (25, 27). The level (λ_E) of an enzyme E is modeled as the product of three factors: the baseline enzyme level (J/k_{cat}), a thermodynamic penalty $(1 - \exp(\Delta_r G'/RT))^{-1}$, and a kinetic penalty $(1 + \Pi (K_{M,i}/C_i)^{mi})$. Given metabolite concentrations C_i and the net reaction flux $J = J^+ - J^-$, the expected enzyme level is:

$$\lambda_E = \frac{J^+}{k_{cat}}\left(1+\prod\left(\frac{K_{M,i}}{C_i}\right)^{m_i}\right)$$
$$= \frac{J}{k_{cat}}\left(1-\exp\left(\frac{\Delta_r G'}{RT}\right)\right)^{-1}\left(1+\prod_{i=1}^{N_C}\left(\frac{K_{M,i}}{C_i}\right)^{m_i}\right),$$

where k_{cat} is the maximum catalytic rate per active site, N_C is the number of substrates in the reaction, $K_{M,i}$ is the Michaelis constant for the i^{th} substrate, and m_i is its stoichiometric coefficient.

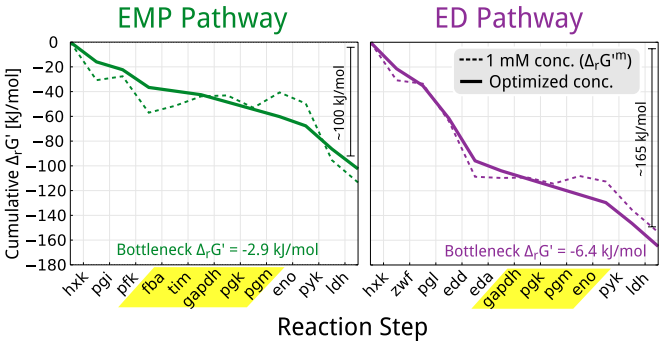

Fig. 3. The ED pathway is substantially more thermodynamically favorable than the EMP pathway. The pathway thermodynamic profile is represented as the cumulative sum of the intermediate reaction Gibbs energies ($\Delta_r G$) in various conditions. All $\Delta_r G$ values are transformed to pH 7.5 and ionic strength 0.2 M and are given per mole of the pathway net reaction so that the final sum equals the $\Delta_r G'$ of the net reaction. The dashed line represents the profile given 1 mM reactant concentrations ($\Delta_r G'^m$), and the solid line represents reaction energies when concentrations are optimized to maximize the Net Flux Ratio (*Materials and Methods*). Optimization was performed presuming metabolite concentrations range between 1 μM and 10 mM. The least favorable reactions are highlighted in yellow. Under optimized concentrations, the least favorable ED reactions (gapdh, pgk, pgm, and eno with $\Delta_r G' \sim -6.4$ kJ/mol) are nearly twice as exergonic as the least favorable EMP reactions (fba, tim, gapdh, pgk, and gpm with $\Delta_r G' \sim -2.9$ kJ/mol).

λ_E approaches the baseline enzyme level when the enzyme is substrate-saturated ($C_i >> K_{M,i}$) and the backward flux is negligible ($\Delta_r G' << 0$). If a one-to-one reaction is irreversible, the thermodynamic penalty disappears, leaving the familiar irreversible Michaelis–Menten relationship. If a reaction is nearer to equilibrium, the thermodynamic penalty amplifies λ_E to account for backward flux.

For a pathway sustaining a net flux J_{path}, we calculate the protein cost of a pathway $\Lambda([C]) = J_{path}^{-1} \Sigma_E (M_E \lambda_E)$ as the protein mass per unit pathway flux, where M_E is the molecular mass of each enzyme in the pathway. Since Λ is convex within the feasible region of metabolite concentrations (*SI Text*), we can find its minimum Λ^* (*Materials and Methods*). We consider Λ^* to be a proxy for the pathway protein cost as it represents the smallest investment of enzyme mass per flux unit in grams of protein per mol pathway reaction per second [g/(mol s^{-1})].

Computing Λ^* requires kinetic parameters for all pathway enzymes. However, these data are usually unavailable and often inconsistent between studies or organisms (24). As the ED and EMP pathways are equally central and chemically similar (Fig. 1 *A*–*C*), we assume, as a first approximation, that all enzymes are equally efficient, faster-than-average central metabolic enzymes (24) (*Materials and Methods*). Under these assumptions, the value of Λ^* is roughly 3.5-fold greater for the EMP pathway than for the ED pathway (Figs. 4 and 5, *SI Text*), suggesting that the EMP pathway requires much more protein mass than the ED to achieve the same glycolytic flux. However, some EMP enzymes (e.g., triose-phosphate isomerase) are much-better-than-average catalysts (24). Perhaps fast or high-affinity enzymes ameliorate the protein cost of the ED pathway? To address this question, we repeated our protein cost analysis using the most reliable kinetic data from the literature (*Materials and Methods*). Fig. 5 shows that measured kinetic parameters do not rescue the EMP pathway, but rather increase the difference in Λ^* between the ED and EMP pathways to roughly fivefold.

To examine the extent to which our computational analysis matches the observed behavior of glycolysis, we compare our protein cost estimates to the measured abundances of glycolytic enzymes. When *E. coli* is grown aerobically on glucose, it consumes about 5 mmol glucose per gram dry weight per hour through the EMP pathway (9). Since 50–55% of *E. coli* dry weight is protein (28), we can estimate that between 3% and 7% of *E. coli*'s proteome should be EMP enzymes (Fig. 5, *Materials and Methods*). Recent proteomic surveys show that EMP enzymes are consistently among the most highly expressed proteins in many model systems, making up 4–6% of the *E. coli* proteome and 14–15% of the *Saccharomyces cerevisiae* (protein units; *SI Text* and Fig. S4). Altogether, our analysis suggests that the energetic difference between the ED and EMP pathways leads to a large difference in the protein cost associated with each pathway. As protein investment in EMP glycolysis has been measured and is quite large, this difference is probably of selective importance.

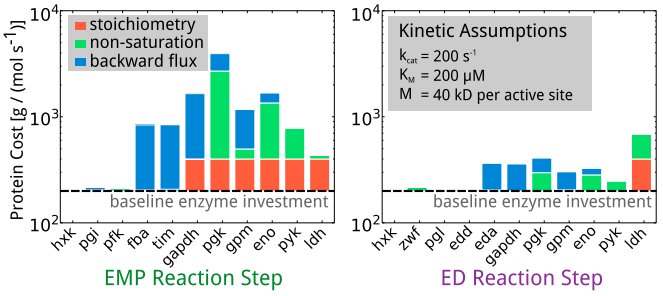

Fig. 4. Protein cost of the ED and EMP pathways. The protein cost of each pathway was calculated as the minimum total enzyme mass per unit pathway flux required for a pathway's operation (*Materials and Methods*). All enzymes were initially assumed to be equally fast and high-affinity. Predicted enzyme levels are shown in log-scale so that each multiplicative term in the equation for λ_E contributes additively. The baseline enzyme investment (M_E/k_{cat}) is the minimum enzyme mass required to convert a mole of substrate per second if there is no thermodynamic or kinetic constraints. Due to pathway stoichiometry, some reactions must occur twice for each glucose metabolized. The effect of stoichiometry on λ_E is shown in red. If the enzyme is not saturated or the reaction is near equilibrium, then the saturation (green) or thermodynamic (blue) terms increase enzyme levels further. Under these assumptions, several EMP enzymes must be expressed at levels five- to sevenfold higher than the theoretical minimum and the pathway as a whole is expected to require 3.5-fold more protein mass than the ED pathway to catalyze the same flux.

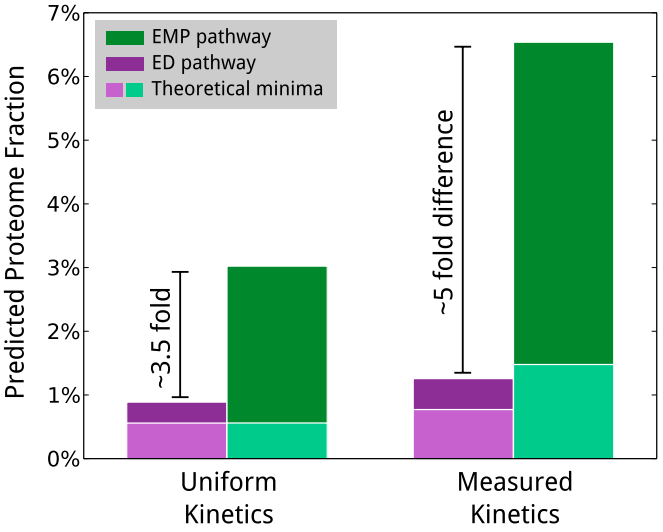

Fig. 5. The EMP pathway is expected to occupy 3–7% of the *E. coli* proteome, three- to fivefold more than the ED. Aerobically grown *E. coli* consumes ~5 mmol glucose $gDW^{-1} \cdot h^{-1}$ through the EMP pathway and 50–55% of *E. coli* dry weight is protein. Transparent gray bars represent the mass fraction of the proteome that would be required in the hypothetical case that all pathway reactions were irreversible and all enzymes could be substrate-saturated. If all enzymes were kinetically identical (as in Fig. 4), then we predict that 3% of *E. coli*'s protein mass would be EMP enzymes. Applying measured kinetic parameters for each enzyme only increases this estimate, predicting that EMP enzymes would occupy 7% of the proteome. This prediction is due to the interplay of thermodynamic and kinetic factors, which prevents several EMP enzymes from being saturated and requires some reactions to operate near equilibrium. The ED pathway, in contrast, can achieve an equivalent steady-state flux with much less enzymatic protein mass because of its more favorable thermodynamic profile.

Genetic Distribution of Glycolytic Strategies Follows Energy Supply. At first glance, the EMP pathway seems superior to the ED pathway as it yields twice as much ATP per glucose. However, protein cost analysis reverses this argument, showing that the ED pathway likely requires several-fold less enzymatic protein than the EMP pathway to sustain the same flux precisely because of its lower ATP yield. If protein synthesis is growth-limiting (16, 29), why would a bacterium incur sizable additional protein cost for only one more ATP? For fermentative anaerobes, glycolysis is the only source of ATP production. Other organisms, however, have ample nonglycolytic means of ATP production. For aerobes, for example, glycolytic ATP yield represents only a small fraction of the 25–30 ATP produced through the full oxidation of glucose (2). We therefore predict that organisms with a nonglycolytic source of ATP will tend to use the ED pathway due to its lower protein cost while organisms relying on glycolytic ATP production will tend to use the EMP pathway.

Indeed, nearly all prokaryotes known to use the ED pathway for glucose catabolism have ample nonglycolytic sources of ATP like oxidative phosphorylation or photosystems (*SI Text*) (9). A notable exception is *Z. mobilis*, an obligate anaerobe used to brew African palm wine and Mexican pulque from sugar-rich saps (4). Although *Z. mobilis* grows very slowly in standard laboratory glucose concentrations, it grows quickly in 0.5 M glucose, indicating that it is likely adapted to very high sugar concentrations—an energy-rich environment of another sort (30).

We applied annotated genomic data to extend this anecdotal analysis. By overlaying oxygen-requirement annotations on a phylogenetic tree along with pathway capabilities, we ask whether particular types of organisms (aerobes, anaerobes, or facultative anaerobes) tend to be capable of one pathway or another. We find that there is a clear and statistically significant correlation between a microbe's glycolytic capabilities and its oxygen requirement (Fig. 6, *Materials and Methods*). Specifically, the prevalence of the ED pathway rises with an organism's exposure to molecular oxygen: anaerobes use the EMP pathway almost exclusively ($P < 0.003$), while the ED pathway is overrepresented among aerobes ($P < 0.001$) and facultative organisms tend to contain genes for both pathways ($P < 0.05$).

Discussion

Several models of bacterial growth have suggested that protein synthesis can be growth-limiting in various circumstances (16, 29, 31, 32). Increased protein expression is known to have a detrimental effect on growth in other ways as well, by increasing misfolding (26) and through toxic promiscuous activity (33), for example. As glycolytic enzymes constitute a large fraction of microbial proteomes (Fig. S4), lowering their levels (i.e., by using the ED pathway) could allow cells to increase their growth rate. We therefore predict a tradeoff between a glycolytic pathway's ATP yield and the growth rate it can support. As some organisms (e.g., fermentative anaerobes) rely completely on glycolytic ATP production, the ATP yield of the ED pathway may not suffice. Other organisms live in environments where the availability of terminal electron acceptors like molecular oxygen enables respiration (2) and makes glycolytic ATP production negligible, allowing them to use the ED pathway without material drawbacks.

Testing our protein cost analysis experimentally requires proteomic data. To date, proteomic surveys have yet to be conducted in organisms relying on the ED pathway. Tandem measurements of metabolic fluxes and protein levels in organisms like *Pseudomonas fluorescens* and *Z. mobilis*—ED-using heterotrophs—would allow comparison of our predictions with measured ED enzyme levels. Furthermore, as many bacteria (including *E. coli*) are capable of both the ED and EMP pathways (Fig. 2), it is possible to execute controlled comparisons of the growth of a single organism using each pathway. A fair comparison, however, requires that the bacteria are well-adapted to both pathways—inefficient ED metabolism cannot be compared with evolutionarily tuned EMP metabolism—and so this approach may necessitate long-term laboratory evolution experiments. Such explorations might also elucidate why *E. coli*, which is capable of both pathways and grows quickly in aerobic environments, tends to use the EMP pathway.

Through the example of the EMP pathway, we showed that a careful consideration of biophysical constraints, thermodynamic, and kinetic factors can estimate the overall protein investment due to a pathway. These same considerations predict that the ED pathway requires three- to fivefold less protein than the EMP pathway to achieve the same glycolytic flux. This difference in protein cost helps rationalize an otherwise confusing reality: that many modern bacteria use the ED pathway even though it yields less ATP than the EMP (9). Indeed, we find evidence in the genomic record that prokaryotes with greater access to nonglycolytic energy sources (i.e., oxidative phosphorylation) tend to contain ED enzymes (Fig. 6).

However, although the genomic tendencies discussed are statistically significant, they are only tendencies. It is not the case that all aerobes rely solely on the ED pathway in the way that nearly all anaerobes rely on the EMP (Fig. 6), which raises a number of fascinating questions. For example, what factors other than protein cost determine the glycolytic pathway (1, 5)? If protein cost is a primary determinant of glycolytic strategy, is there an "exchange rate" between ATP production and protein investment (18)? Is this tradeoff constant, or does it vary greatly between organisms and conditions? More sharply, how does additional protein production affect cell growth (16, 33–35), and how does metabolism evolve to cope with high protein cost (17, 34)? Many researchers have begun to address these questions, but they are by no means resolved. We hope that future work will elucidate the degree to which such tradeoffs explain the structure and regulation of natural metabolic systems.

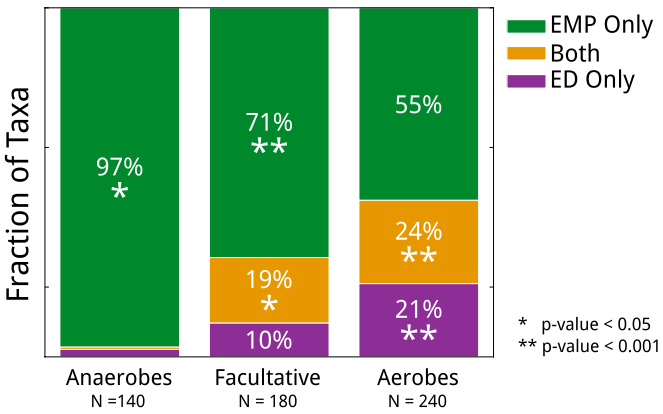

Fig. 6. Prokaryotic glycolytic strategy correlates with the availability of nonglycolytic energy sources. Each organism was marked as ED-capable, EMP-capable, or capable of both pathways as in Fig. 2 (*Materials and Methods*). Organisms were further categorized as aerobes, anaerobes, or facultative anaerobes according to the IMG database (38). Closely related organisms were merged (*Materials and Methods*). The distribution of pathway capabilities is not uniform. Rather, the fraction of organisms capable only of the EMP pathway decreases with increasing exposure to oxygen: 97% of anaerobes are solely capable of the EMP pathway ($P < 0.003$), while only 55% of aerobes are similarly categorized. Facultative prokaryotes tend to contain the EMP pathway alone ($P < 0.001$) or genes for both pathways ($P < 0.05$), while aerobes are highly enriched with the ED pathway ($P < 0.001$). This trend agrees with our prediction that organisms with significant nonglycolytic sources of ATP (i.e., oxidative phosphorylation) will tend to use the ED pathway due to its lower protein cost.

Materials and Methods

Phylogenetic Analysis. Gene annotations were downloaded from the KEGG database (36) and placed on a phylogenetic tree of bacteria and archaea (37). Organisms were considered genetically capable of a particular pathway if their genome contained a pyk (EC 2.7.1.40) and all enzymes unique to that pathway (Fig. 1). Microbes with a phylogenetic distance less than 0.008 units were merged to avoid double-counting closely related strains. This threshold was chosen so that all sequenced *E. coli* strains collapse to a single record. Merged microbes were taken to contain all genes present in either original record. Organisms were marked as heterotrophs according to annotations from the IMG database (38).

Energy Supply Analysis. Oxygen requirement annotations were downloaded from the IMG database (38) and matched to the phylogenetic tree. Closely related strains were collapsed as in the phylogenetic analysis. Merged records were taken to represent a distribution of oxygen requirements—one of each category for each original record. For example, if two aerobes were merged with an anaerobe, then the merged record was treated as two-thirds aerobe and one-third anaerobe. The contracted phylogenetic tree was used to calculate the co-occurrence of each oxygen requirement with the genes

of the ED and EMP pathway. *P* values for each pair of oxygen requirement and pathway capability were calculated assuming the null hypothesis that oxygen requirements and pathways are drawn independently from the overall distribution.

Thermodynamic, Kinetic, and Protein Data. When available, experimentally measured standard reaction Gibbs energies ($\Delta_r G'^\circ$) were calculated from the National Institute of Science and Technology Thermodynamics of Enzyme-Catalyzed Reactions DataBase (39) and adjusted to pH 7.5 and ionic strength 0.2 M (14). When experimental data were not available, a novel group contribution approximation was then used to calculate $\Delta_r G'^\circ$ in a manner consistent with experimental measurements. K_M and k_{cat} values were manually chosen from the BRENDA database (40). We gave preference to studies measuring K_M and k_{cat} values of native *E. coli* enzymes, and we used data from other microbes when *E. coli* data were not available (Datasets S1 and S2). Measured glycolytic protein levels were retrieved from PaxDB (41).

Thermodynamic Analysis. As the Net Flux Ratio $J/J^+ = 1 - \exp(\Delta_r G'/RT)$ is monotonic in $-\Delta_r G'$, minimizing $\Delta_r G'$ of a reaction maximizes the ratio. Constraint-based optimization was used to maximize the Net Flux Ratio of a pathway's least favorable reaction:

$$\begin{aligned}&\text{maximize } B = \min_r(-\Delta_r G')\\&\text{where } \Delta_r G' = \Delta_r G'^\circ + RT \cdot S \cdot \ln(C)\\&\text{such that } \ln(C_{min}) \le \ln(C) \le \ln(C_{max})\\&\text{and } \Delta_r G' \le 0 \text{ for all r.}\end{aligned}$$

Here, $\Delta_r G'^\circ$ is a vector of standard reaction Gibbs energies for pathway reactions and C is a vector of reactant concentrations (Table S4). S is the MxN pathway stoichiometric matrix where $S_{i,j} = m_{i,j}$, the stoichiometric coefficient of compound i in reaction j. This optimization is linear in ln(C) and therefore solvable using standard optimization techniques. Our implementation uses the cvxpy package for the Python programming language.

Protein Optimization. The minimum possible value of Λ (Λ^*), taken as a proxy for the pathway protein cost, is defined by the following constrained optimization:

$$\begin{aligned}&\text{minimize } \Lambda^* = \Sigma_E(M_E\lambda_E/V_{total})\\&\text{where } \lambda_E/V_{total} = (\gamma_i/k_{cat,i}) \cdot \left(1 + \Pi_j \left(K_{Mi,j}/C_j\right)^{mi,j^+}\right) \Big/ (1 - \exp(\Delta_r G'_i/RT))\\&\text{and } \Delta_r G' = \Delta_r G'^\circ + RT \cdot S \cdot \ln(C)\\&\text{such that } \ln(C_{min}) \le \ln(C) \le \ln(C_{max})\\&\text{and } \Delta_r G' \le 0 \text{ for all r,}\end{aligned}$$

where γ_i is the stoichiometric multiplicity of reaction i in the pathway and V_{total} is the overall pathway flux. Λ^* has units of g/(mol s^{-1}). Multiplying Λ^* by the pathway flux in units of mol s^{-1}·gDW^{-1}, we get the predicted mass ratio of pathway protein to dry weight (Tables S5–S8). The equation for Λ is convex within the feasible region of logarithmic metabolite concentrations—that is, when $\Delta_r G' < 0$ (*SI Text* and Fig. S5). Therefore, Λ^* can be calculated using general constrained optimization methods. We use the SciPy implementation of sequential least squares programming for the Python programming language. We tested several (>10) feasible starting points and verified that all optima found are consistent (*SI Text*). All source code is available at http://code.google.com/p/milo-lab/.

ACKNOWLEDGMENTS. We thank Leeat Keren, Niv Antonovsky, Ayelet Levin-Karp, Oren Yishai, Yaniv Lubling, Dan Tawfik, Uri Barenholz, Dan Fraenkel, Naama Tepper, and Shira Amram for thought-provoking discussions that enriched this work enormously. We also thank Ben Herzberg and Dan Arlow for their help with the manuscript and Rakevet Yisrael for hours of comfortable, uninterrupted work. This work was funded by the European Research Council (260392 – Project SYMPAC); Israel Science Foundation (Grant 750/09); Helmsley Charitable Foundation; the Larson Charitable Foundation; Estate of David Arthur Barton; Anthony Stalbow Charitable Trust, and Stella Gelerman (Canada); and the German Research Foundation (LI 1676/2-1). E.N. is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship. R.M. is the incumbent of the Anna and Maurice Boukstein Career Development Chair in Perpetuity.

1. Romano AH, Conway T (1996) Evolution of carbohydrate metabolic pathways. *Res Microbiol* 147(6-7):448–455.
2. Kim BH, Gadd GM (2008) *Bacterial Physiology and Metabolism* (Cambridge Univ Press, Cambridge, UK).
3. Fraenkel D (1996) Escherichia coli and Salmonella. *Cellular and Molecular Biology*, ed Neidhardt FC (American Society for Microbiology, Washington, DC), Vol. 1, pp 189–199.
4. Conway T (1992) The Entner-Doudoroff pathway: History, physiology and molecular biology. *FEMS Microbiol Rev* 9(1):1–27.
5. Meléndez-Hevia E, Waddell TG, Heinrich R, Montero F (1997) Theoretical approaches to the evolutionary optimization of glycolysis—chemical analysis. *Eur J Biochem* 244(2):527–543.
6. Entner N, Doudoroff M (1952) Glucose and gluconic acid oxidation of Pseudomonas saccharophila. *J Biol Chem* 196(2):853–862.
7. Bar-Even A, Flamholz A, Noor E, Milo R (2012) Rethinking glycolysis: On the biochemical logic of metabolic pathways. *Nat Chem Biol* 8(6):509–517.
8. Sung S-JS, Xu D-P, Galloway CM, Black CC (1988) A reassessment of glycolysis and gluconeogenesis in higher plants. *Physiol Plant* 72(3):650–654.
9. Fuhrer T, Fischer E, Sauer U (2005) Experimental identification and quantification of glucose metabolism in seven bacterial species. *J Bacteriol* 187(5):1581–1590.
10. Peekhaus N, Conway T (1998) What's for dinner? Entner-Doudoroff metabolism in Escherichia coli. *J Bacteriol* 180(14):3495–3502.
11. Sweeney NJ, Laux DC, Cohen PS (1996) Escherichia coli F-18 and E. coli K-12 eda mutants do not colonize the streptomycin-treated mouse large intestine. *Infect Immun* 64(9):3504–3511.
12. Schuster S, Pfeiffer T, Fell DA (2008) Is maximization of molar yield in metabolic networks favoured by evolution? *J Theor Biol* 252(3):497–504.
13. Flamholz A, Noor E, Bar-Even A, Milo R (2011) eQuilibrator–The biochemical thermodynamics calculator. *Nucleic Acids Res* 40:D770–D775.
14. Noor E, et al. (2012) An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. *Bioinformatics* 28(15):2037–2044.
15. Bar-Even A, Noor E, Lewis NE, Milo R (2010) Design and analysis of synthetic carbon fixation pathways. *Proc Natl Acad Sci USA* 107(19):8889–8894.
16. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of cell growth and gene expression: Origins and consequences. *Science* 330(6007):1099–1102.
17. Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436(7050):588–592.
18. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U (2012) Multidimensional optimality of microbial metabolism. *Science* 336(6081):601–604.
19. Alberty RA (2003) *Thermodynamics of Biochemical Reactions* (Wiley, Hoboken, NJ), 1st Ed.
20. Vojinović V, von Stockar U (2009) Influence of uncertainties in pH, pMg, activity coefficients, metabolite concentrations, and other factors on the analysis of the thermodynamic feasibility of metabolic pathways. *Biotechnol Bioeng* 103(4):780–795.
21. Mavrovouniotis ML (1993) Identification of localized and distributed bottlenecks in metabolic pathways. *Proc Int Conf Intell Syst Mol Biol* 1:275–283.
22. Beard DA, Qian H (2007) Relationship between thermodynamic driving force and one-way fluxes in reversible processes. *PLoS ONE* 2(1):e144.
23. Bennett BD, et al. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat Chem Biol* 5(8):593–599.
24. Bar-Even A, et al. (2011) The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50(21):4402–4410.
25. Heijnen J (2010) Impact of thermodynamic principles in systems biology. *Advances in Biochemical Engineering/Biotechnology* (Springer, Berlin), pp 140–153.
26. Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134(2):341–352.
27. Liebermeister W, Klipp E (2006) Bringing metabolic networks to life: Convenience rate law and thermodynamic constraints. *Theor Biol Med Model* 3:41.
28. Umbarger HE, Neidhard FC (1996) Escherichia coli and Salmonella. Cellular and Molecular Biology, ed Neidhardt FC (American Society for Microbiology, Washington, DC), Vol. 1, pp 13–17.
29. Molenaar D, van Berlo R, de Ridder D, Teusink B (2009) Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol* 5:323.
30. Seki M, Furusaki S, Shigematsu K (1990) Cell growth and reaction characteristics of immobilized Zymomonas mobilis. *Ann N Y Acad Sci* 613(1):290–300.
31. Marr AG (1991) Growth rate of Escherichia coli. *Microbiol Rev* 55(2):316–333.
32. Schaechter M, Maaloe O, Kjeldgaard NO (1958) Dependency on medium and temperature of cell size and chemical composition during balanced grown of Salmonella typhimurium. *J Gen Microbiol* 19(3):592–606.
33. Eames M, Kortemme T (2012) Cost-benefit tradeoffs in engineered lac operons. *Science* 336(6083):911–915.
34. Shachrai I, Zaslaver A, Alon U, Dekel E (2010) Cost of unneeded proteins in E. coli is reduced after several generations in exponential growth. *Mol Cell* 38(5):758–767.
35. Zhuang K, Vemuri GN, Mahadevan R (2011) Economics of membrane occupancy and respiro-fermentation. *Mol Syst Biol* 7:500.
36. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28(1):27–30.
37. Dehal PS, et al. (2010) MicrobesOnline: An integrated portal for comparative and functional genomics. *Nucleic Acids Res* 38(Database issue):D396–D400.
38. Markowitz VM, et al. (2010) The integrated microbial genomes system: An expanding comparative analysis resource. *Nucleic Acids Res* 38(Database issue):D382–D390.
39. Goldberg RN, Tewari YB, Bhat TN (2004) Thermodynamics of enzyme-catalyzed reactions—A database for quantitative biochemistry. *Bioinformatics* 20(16):2874–2877.
40. Schomburg I, et al. (2002) BRENDA: a resource for enzyme data and metabolic information. *Trends Biochem Sci* 27(1):54–56.
41. Wang M, et al. (2012) PaxDb, a database of protein abundance averages across all three domains of life. *Mol Cell Proteomics* 11(8):492–500.