Lower glycolysis carries a higher flux than any biochemically possible alternative
Supporting Information

Steven J. Court, Bartłomiej Waclaw, and Rosalind J. Allen

I. THE TRUNK PATHWAY OF GLYCOLYSIS AND GLUCONEOGENESIS

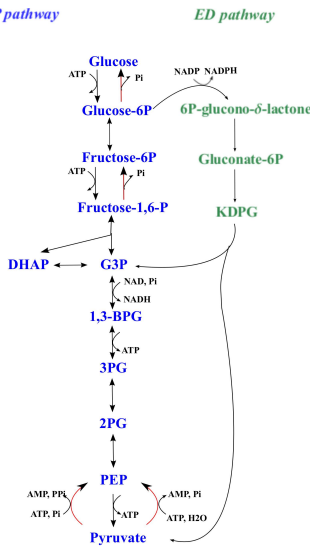

FIG. S 1: The Embden-Meyerhof-Parnas and Entner-Doudoroff glycolytic pathways. The red arrows indicate gluconeogenic reactions and KDPG denotes 2-keto-3-deoxy-6-phosphogluconate.

Fig. S1 shows the Embden-Meyerhof-Parnas and Entner-Doudoroff glycolytic pathways, which convert glucose to pyruvate. These pathways share a common set of reactions in "lower glycolysis"; these reactions are known as the *trunk pathway*[1]. The pentose phosphate pathway[2] also feeds into this same trunk pathway. In prokaryotes, the trunk pathways in the glycolytic (G3P to pyruvate) and gluconeogenic (pyruvate to G3P) directions both have 5 steps and share almost the same set of reactions. The exception is the highly exergonic final step in the glycolytic direction: in glycolysis phosphoenolpyruvate (PEP) is transformed to pyruvate with the conversion of an ADP molecule to ATP via pyruvate kinase (EC 2.7.1.40); in gluconeogenesis two different solutions exist to overcome this energetic barrier. The most prevalent[1] is the phosphoenolpyruvate synthetase reaction (pps, EC 2.7.9.2) which converts pyruvate to phosphoenolpyruvate and couples with the conversion of ATP and water to AMP and inorganic phosphate. Alternatively, some organisms use pyruvate, phosphate dikinase (ppdk, EC 2.7.9.1) which couples the reaction to the conversion of ATP and inorganic phosphate to AMP and pyrophosphate. Under the conditions assumed in this work (T=25^{0}C, pH=7.0, I=0.2) and assuming standard concentrations of 1M, the pps reaction has a free energy change of $\Delta G = -5.0$ kJmol^{-1} while ppdk is less thermodynamically favourable

with $\Delta G = +17.7$ kJmol^{-1}. However changes in the concentrations of the external metabolites, in particular inorganic phosphate [Pi] and pyrophosphate [PPi], can change the relative favourability of these reactions (see Section X).

II. CONSTRUCTION OF OUR NETWORK OF ALL POSSIBLE CHEMICAL REACTIONS

Metabolites

As discussed in the main text, we restrict our list of internal metabolites to unbranched, aliphatic 2, 3 and 4-carbon CHOP molecules (*i.e.* those containing only carbon, hydrogen, oxygen and phosphorus atoms, with the phosphorus being in the form of phosphate groups). We exclude hydrocarbons (*i.e.* molecules with no oxygen atoms), C≡C triple bonds and esters. We also neglect chirality. We consider only molecules which are charged, *i.e.* contain at least one carboxyl or phosphate group; this is because nonpolar molecules can have very high membrane permeability[3], making them prone to leaking out of the cell. In our computer program, each internal metabolite is constructed as a linear combination of the groups shown in Table SI. Note that we do not allow these groups (containing a single carbon atom) to contain more than one hydroxyl or phosphate group on the grounds of stability. The important metabolite pyruvate, for example, is produced by combining, in order, the three groups CH_3-CO-COOH. The full list of internal metabolites is generated systematically by combining these groups in all possible permutations, eliminating any molecules which do not satisfy the above restrictions.

Many of the reactions in our network involve cofactors such as ATP, ADP, NAD and NADH. These are classified as *external metabolites* in our analysis, and their concentrations define the intracellular conditions. The external metabolites included in our analysis are listed in Table SII. We assume that, in solution, these compounds exist as a mixture of different dissociated forms; for example ATP exists in cells as a mixture of ATP^{4-}, $HATP^{3-}$, H_2ATP^{2-} etc, while "CO_2(aq)" in our table represents H_2CO_3 and its various forms HCO_3^-, CO_3^{-2}, etc. This is taken into account in the calculation of the free energy as described in Section III.

-CH$_3$	**-CH$_2$(OH)**	**-COOH**	**-CHO**	**-CH$_2$p**	**-COp**
=CH$_2$	**=CH(OH)**	**=CO**	**=CHp**	-CH(OH)-	-CO-
-CHp-	-CH=	-C(OH)=	-Cp=	-CH$_2$-	

TABLE S I: The set of groups used to construct the internal metabolites in our network. Groups in bold indicate those which must be present at the ends of a molecule to prevent "dangling bonds". Note that the phosphate group "p" and carboxyl group "COOH" both denote a mixture of protonated and deprotonated forms. This is taken into account in the calculation of free energies of formation (see Section III).

Reactions

Our program systematically constructs all possible reactions connecting pairs of internal metabolites, consistent with the 12 EC classes listed in Table 1 of the main text. The EC classification[4]

ATP	ADP	AMP
NAD^+	NADH	$CO_2(aq)$
Pi	PPi	H_2O

TABLE S II: The set of external metabolites used in our analysis.

consists of four numbers that provide increasingly detailed information about the action of an enzyme. For instance EC 5 corresponds to isomerases, EC 5.4 to intramolecular transferases and EC 5.4.2 to phosphotransferases. The fourth number indicates the specific substrates and products. The reaction classes in our network are described by just the first 3 EC numbers, since they describe mechanistically equivalent chemical transformations, carried out on different chemical substrates. For each reaction class, we include all known couplings to the external metabolites. This means that a given pair of internal metabolites may be connected by more than one reaction (edge) in our network, e.g. the three different reactions coupling PEP and pyruvate in Fig. S1. Different couplings can involve different free energy changes and can be affected by the cellular environment in different ways. Table SIII provides a complete list of the reaction types included in our analysis. In total our network contains 828 metabolites and 7145 reactions.

III. ESTIMATING REACTION FREE ENERGIES

We calculate the *biological* standard free energy change $\Delta_r G'$ (denoted by a prime; defined under conditions with T=25^oC, pH=7.0, I=0.2, and with all concentrations 1M) associated with a given reaction in our network by subtracting the free energy of formation of the reactants from that of the products, including the external metabolites (cofactors):

$$\Delta_r G' = \sum_{i \in \{products\}} \Delta_f G'_i - \sum_{j \in \{reactants\}} \Delta_f G'_j \ . \tag{1}$$

This requires us to estimate the free energies of formation $\Delta_f G'$ for all the molecules in the network. However, for many of the compounds, no experimental values for $\Delta_f G'$ is available. In these cases we estimate $\Delta_f G'$ using a group contribution method based on Refs.[5–7]. The basic idea behind the group contribution method is that any molecule can be split into a number of functional groups, each making a contribution to the total free energy of formation of the molecule. Assuming that all molecules are built from a linear combination of the functional groups $\{g_i\}$ listed in Table SI, we write their $\Delta_f G'$ as

$$\Delta_f G' = E_0 + \sum_j E_1(g_j) + \sum_{<j,k>} E_2(g_j, g_k), \tag{2}$$

where E_0 is a constant (which is the same for all molecules), $E_1(g_j)$ is the contribution of group g_j and $E_2(g_j, g_k)$ is a small correction due to neighbouring group-group interactions. Alberty[8] has listed standard Gibbs free energies of formation for around 200 compounds of biological interest, 56 of which are linear CHOP molecules as in our network. We obtained the values of E_0, the vector E_1 and matrix E_2 by performing a least-squares fit to this set of 56 data points. We also account for the fact that chemicals in aqueous solution can exist as a mixture of dissociated species[9,10]. The

Reaction type	EC	Generalised reaction(s)
oxidation	1.1.1	R-CH_2(OH) + NAD ⇌ R′-CHO + NADH
	1.1.1	R-CH(OH)-R′ + NAD ⇌ R-CO-R′ + NADH
	1.2.1	R-CHO + NAD + H_2O ⇌ R-COOH + NADH
oxidation and phosphorylation	1.2.1	R-CHO + NAD + Pi ⇌ R-COp + NADH
phosphate transfer	2.7.1	R-OH + ATP ⇌ R-p + ADP
	2.7.1	R=Cp-R′ + ADP ⇌ RH-CO-R′ + ATP
	2.7.2	R-COOH + ATP ⇌ R-COp + ADP
	2.7.9	R-OH + ATP + H_2O ⇌ R-p + AMP + Pi
	2.7.9	R-OH + Pi + ATP ⇌ R-p + AMP + PPi
	2.7.9	R-COOH + ATP + H_2O ⇌ R-COp + AMP + Pi
	2.7.9	R-COOH + Pi + ATP ⇌ R-COp + AMP + PPi
	2.7.9	R=CP-R′ + AMP + Pi ⇌ R-CO-R′ + ATP + H_2O
	2.7.9	R=Cp-R′ + AMP + PPi ⇌ R-CO-R′ + ATP + Pi
hydrolysis	3.1.3	R-p + H_2O ⇌ R-OH + Pi
decarboxylation	4.1.1	R-COOH + H_2O ⇌ R-H + CO_2(aq)
	4.1.1	COOH-C(R′R″)-CO-R + Pi ⇌ C(R′R″)=Cp-R + CO_2(aq)
decarboxylation and phosphorylation	4.1.1	R-COOH + ATP ⇌ R-p + CO_2(aq) + ADP
dehydration	4.2.1	R(H)-(OH)R′ ⇌ R=R′ + H_2O
isomerisation	5.3.1	R-CO-CH_2(OH) ⇌ R-CH(OH)-CHO
	5.4.2	R-CH(OH)-CH_2p ⇌ R-CHp-CH_2(OH)
	5.4.2	R=C(OH)-CH_2p ⇌ R=Cp-CH_2(OH)
	5.4.2	R-CHp-COOH ⇌ R-CH(OH)-COp
	5.4.2	R=Cp-COOH ⇌ R=C(OH)-COp
tautomerism	5.3.2	R=C(OH)-R′ ⇌ RH-CO-R′
	5.3.2	R=CH(OH) ⇌ RH-CHO
ATP-driven carboxylation	6.4.1	R-H + CO_2(aq) + ATP ⇌ R-COOH + ADP + Pi

TABLE S III: The set of reaction types included in our network, with all external metabolite couplings. Here, R and R′ denote arbitary chemical groups (not necessarily linear chains), Pi denotes free inorganic phosphate; PPi denotes pyrophosphate, and p denotes the phosphate group of a phosphorylated substrate. All compounds are assumed to exist as an equilibrium mixture of protonated and deprotonated species in solution.

free energy of formation $\Delta_f G'$ of compound A at a given pH and ionic strength I can be written as[9]:

$$\Delta_f G' = \Delta_f G^{\mathrm{pH,I}}_{(\mathrm{A})} = -RT \ln \left[\sum_i \exp\left(-\frac{\Delta_f G^0_i}{RT} + N_{\mathrm{H}}(i) \ln 10^{-\mathrm{pH}} - \frac{2.91482(Z_i^2 - N_{\mathrm{H}}(i))I^{1/2}}{1 + 1.6 I^{1/2}} \right) \right], \tag{3}$$

where the index i runs over all differently protonated forms of A ($\mathrm{H}_n\mathrm{A}$, $\mathrm{H}_{n-1}\mathrm{A}^-$, $\mathrm{H}_{n-2}\mathrm{A}^{2-}$, etc.), Z_i is the charge of the form i, $N_{\mathrm{H}}(i)$ is the total number of hydrogen atoms of i, and $\Delta_f G^0_i$ is the

standard free energy of formation of that particular form (25^oC, pH=0, 1M concentration, and zero ionic strength $I = 0$). The sum over i comes from the entropy of mixing of different forms of A. The term proportional to $N_H(i)$ accounts for changes in concentrations of different forms for non-zero pH, and the last term gives a correction for non-zero ionic strength ($I > 0$, due to the presence of ions such as Na^+, Cl^- etc.). We use Eq. (3) to calculate free energies of the training molecules used to obtain parameters for the group contribution method (see Section III). Note that, since we use Eq.(3) to calculate the free energies of the training molecules, the parameters E_0, E_1, and E_2 as determined by us are valid only for the specific set of conditions, and hence the free energies of our CHOP molecules are also valid only for this set of conditions.

For our 56-compound training set, we find that the error – the square root of the variance of absolute differences between the experimental $\Delta_f G'$ values and those calculated using the group contribution method – to be 3.84kJmol^{-1}. Although these errors in $\Delta_f G'$ could, in principle, combine to produce large errors in the reaction free energies, we find that in practice this is not the case. Performing our full analysis using networks trained on various subsets of the 56 compounds demonstrates that our conclusions are robust. For instance, training the group contribution method on random samples of 80% of the compounds in the full training set can alter the free energies of formation by a few kJmol^{-1} but does not qualitatively change the results of our analysis.

A. E0, E1 and E2 values obtained from least-squares fitting

$E_0 = -227.22$ kJmol^{-1}.

E_1: vector of group contributions (kJmol^{-1}). The first column shows each molecular group, with the second column showing its contribution to the free energy of formation of a molecule.

$$E_1 = \begin{pmatrix} \text{-CH3} & 218.75 \\ \text{-CH2(OH)} & 67.712 \\ \text{-COOH} & -232.22 \\ \text{-CHO} & 31.93 \\ \text{-CH2p} & -825.42 \\ \text{-COp} & -1090.9 \\ \text{=CH2} & 236.12 \\ \text{-CH2-} & 82.90 \\ \text{=CH(OH)} & 51.21 \\ \text{-CH(OH)-} & -72.63 \\ \text{=CO} & 178.95 \\ \text{-CO-} & -110.57 \\ \text{=CHp} & -816.33 \\ \text{-CHp-} & -950.69 \\ \text{-CH=} & 83.713 \\ \text{-C(OH)=} & -99.314 \\ \text{-Cp=} & -966.85 \end{pmatrix}$$

E_2: matrix of corrections arising from group-group interactions (kJmol^{-1}). First row and first column indicate the molecular groups.

$E_2 =$

	-CH3	-CH2(OH)	-COOH	-CHO	-CH2p	-COp	=CH2	-CH2-	=CH(OH)	-CH(OH)-	=CO	-CO-	=CHp	-CHp-	-CH=	-C(OH)=	-Cp=
-CH3	16.99	3.41	−7.25	0.39	0	−7.8	0	8.76	0	−0.08	0	−7.26	0	0	0	0	0
-CH2(OH)	3.41	0	−17.84	0	0	0	0	−1.4	0	−1.34	0	−0.13	0	0.37	0	0	0
-COOH	−7.25	−17.84	14.71	−1.14	0	0	0	−1.58	0	5.46	0	9.75	0	0.37	0.25	0.37	0.37
-CHO	0.39	0	−1.14	0	0	0	0	2.08	0	5.49	0	0	0	0	0	0	0
-CH2p	0	0	0	0	0	0	0	0	0	−3.26	0	−1.49	0	0	0	0	0
-COp	−7.8	0	0	0	0	0	0	0	0	12.55	0	0	0	0	0	0	0
=CH2	0	0	0	0	0	0	0.75	0	0.75	0	0.75	0	0.75	0	0	0.37	0.37
-CH2-	8.76	−1.4	−1.58	2.08	0	0	0	−0.48	0	−5.18	0	−5.1	0	0	0	0	0
=CH(OH)	0	0	0	0	0	0	0.75	0	0	0	0	0	0	0	0	0	0
-CH(OH)-	−0.08	−1.34	5.46	5.49	−3.26	12.55	0	−5.18	0	−0.92	0	7.95	0	0	0	0	0
=CO	0	0	0	0	0	0	0.75	0	0	0	0	0	0	0	0	0	0
-CO-	−7.26	−0.13	9.75	0	−1.49	0	0	−5.1	0	7.95	0	0	0	0	0	0	0
=CHp	0	0	0	0	0	0	0.75	0	0	0	0	0	0	0	0	0	0
-CHp-	0	0.37	0.37	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-CH=	0	0	0.25	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0
-C(OH)=	0	0	0.37	0	0	0	0.37	0	0	0	0	0	0	0	0	0	0
-Cp=	0	0	0.37	0	0	0	0.37	0	0	0	0	0	0	0	0	0	0

IV. CALCULATING THE FLUX THROUGH A LINEAR PATHWAY AND THE INTERMEDIATE METABOLITE CONCENTRATIONS

To compute the flux carried by our candidate pathways we generalise the method of Heinrich et al.[11], extending it to the case where reactions include both internal and external metabolites. Let us consider an unbranched chain of n reactions with fixed concentrations of the initial substrate $[S_0]$ and final product $[S_n]$. For the simplest case, in which each reaction converts one substrate S_{i-1} into the next, S_i, and in which the flux v_i through reaction i obeys linear kinetics, we can write

$$v_i = k_i[S_{i-1}] - k_{-i}[S_i] = k_i\left([S_{i-1}] - \frac{[S_i]}{q_i}\right). \tag{4}$$

Here k_i and k_{i-1} are the first-order rate constants in the forward and backward directions and $q_i = e^{-\Delta_r G'/RT} = k_i/k_{-i}$ is the thermodynamic equilibrium constant, where $\Delta_r G'$ is the biological standard free energy change of the reaction as computed in our analysis using the group contribution method. We now generalise this to the case where external metabolites are involved. A generic reaction that converts internal metabolite S_{i-1} to S_i, with a coupled conversion of external metabolite e_{i-1} to e_i, can be expressed as

$$S_{i-1} + e_{i-1} \underset{\kappa_{-i}}{\overset{\kappa_i}{\rightleftharpoons}} S_i + e_i \tag{5}$$

By analogy with Eq. (4), we can write

$$v_i \;=\; \kappa_i[S_{i-1}][e_{i-1}] - \kappa_{-i}[S_i][e_i]. \tag{6}$$

Since we assume that all external metabolites are present at fixed concentrations (which determine the intracellular conditions), we can absorb the concentrations of the external metabolites into the rate constants, to obtain the pseudo first-order rate equation

$$v_i \;=\; K_i[S_{i-1}] - K_{-i}[S_i] = K_i\left([S_{i-1}] - \frac{[S_i]}{q_i'}\right), \tag{7}$$

where $K_i = \kappa_i[e_{i-1}]$, $K_{i-1} = \kappa_{-i}[e_i]$, and $q'_i = K_i/K_{-i} = \kappa_i[e_{i-1}]/\kappa_{-i}[e_i] = q_i\frac{[e_{i-1}]}{[e_i]} = \frac{[e_{i-1}]}{[e_i]}e^{-\Delta_r G'/RT}$. Note that the q' incorporates the shift in the equilibrium constant when taking into account the fixed external metabolite concentrations. We now consider a pathway consisting of a linear chain of such reactions. In steady state, the flux carried by each reaction is equal to the flux J through the pathway, *i.e.* $v_i = J$, for all i. Imposing this condition and rearranging Eq.(7) gives

$$[S_i] = q'_i\left([S_{i-1}] - \frac{J}{K_i}\right). \tag{8}$$

Iterating this equation, starting from $i = 1$, results in an expression for the steady state concentrations of the intermediate metabolites, in terms of the rate constants, $\Delta_r G'$ values, and the concentration $[S_0]$ of the starting substrate:

$$[S_j] = [S_0]\prod_{i=1}^{j} q'_i - J\sum_{i=1}^{j}\frac{1}{K_i}\prod_{m=i}^{j} q'_m. \tag{9}$$

Setting $j = n$ in Eq. 9 and rearranging leads to an expression for the pathway flux J in terms of the rate constants, $\Delta_r G'$ values, and the concentrations $[S_0]$ and $[S_n]$ of the initial and final substrates:

$$J = \frac{1}{D}\left([S_0]\prod_{i=1}^{n} q'_i - [S_n]\right), \tag{10}$$

where

$$D = \sum_{i=1}^{n}\frac{1}{K_i}\prod_{m=i}^{n} q'_m. \tag{11}$$

V. PERFECT CATALYST ASSUMPTION

Eqs (10) and (11) require knowledge of the rate constants for all enzymes in the pathway (via the parameters K_i). For most of the reactions in our network, these parameters are not available. We therefore assume that all enzymes in our network behave as perfect catalysts - *i.e.* their reaction rate is limited only by the rate of diffusion of substrate to the enzyme. Because of this assumption, the flux obtained in our calculations for a given pathway can be thought of as representing the maximum possible flux sustainable by that pathway. Under this assumption the rate equation becomes[11–14]

$$v_i = \frac{k_d[E_i]\left([S_{i-1}]q_i - [S_i]\right)}{1 + q_i}, \tag{12}$$

where $[E_i]$ is the concentration of the enzyme catalyzing step i and k_d denotes the diffusion-limited rate constant. Complementary derivations of Eq.(12) have been presented by Heinrich et al.[12] and Albery et al.[13]. Here we derive Eq. 12 following the approach of Pettersson[14]. We consider a special case of the simple enzyme mechanism describing the reversible binding of a substrate S and product P to an enzyme E:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} EP \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} E + P \tag{13}$$

The steady state of this process is given by[12]

$$v = \frac{E_0(k_1k_2k_3S - k_{-1}k_{-2}k_{-3}P)}{k_{-1}k_{-2} + k_{-1}k_3 + k_2k_3 + (k_1k_{-2} + k_1k_3 + k_1k_2)S + (k_{-1}k_{-3} + k_2k_{-3} + k_{-2}k_{-3})P}. \tag{14}$$

Following the approach of Pettersson[14], we consider an enzyme operating via scheme (13), in the absence of product ($P \to 0$ in Eq. 14). In this case, Eq. (14) reduces to a Michaelis-Menten rate equation:

$$v = \frac{E_0(k_1k_2k_3S)}{k_{-1}k_{-2} + k_{-1}k_3 + k_2k_3 + (k_1k_{-2} + k_1k_3 + k_1k_2)S} \quad \equiv \quad \frac{E_0k_{cat}S}{K_M + S}, \tag{15}$$

yielding the following relationships between the physiological parameters and the microscopic rate constants

$$k_{cat} = \frac{k_2k_3}{k_{-2} + k_3 + k_2} \quad \text{and} \quad K_M = \frac{k_{-1}k_{-2} + k_{-1}k_3 + k_2k_3}{k_1(k_{-2} + k_3 + k_2)}. \tag{16}$$

We now assume that for enzymes operating in scheme (13), the substrate and product association rate constants k_1 and k_{-3} take the value $k_d = 10^9 M^{-1}s^{-1}$, which is the rate constant for diffusion-controlled binding. We further assume, following Pettersson[14], that $k_2 \gg k_{-1}$ (Pettersson justifies this based on evolutionary considerations[14]). Writing

$$\frac{k_{cat}}{K_M} = \frac{k_1k_2k_3}{k_{-1}k_{-2} + k_{-1}k_3 + k_2k_3}, \tag{17}$$

and multiplying top and bottom by $1/k_1k_2k_3$, noting that the equilibrium constant of the reaction is $q = k_1k_2k_3/k_{-1}k_{-2}k_{-3}$, gives

$$\frac{k_{cat}}{K_M} = \frac{1}{\frac{1}{qk_{-3}} + \frac{k_{-1}}{k_1k_2} + \frac{1}{k_1}}. \tag{18}$$

Applying the diffusion limit $k_1 = k_{-3} = k_d$ and assuming $k_2 \gg k_{-1}$, leads finally to the expression

$$\frac{k_{cat}}{K_M} \approx \frac{k_d}{1 + q^{-1}} = \frac{k_dq}{1 + q}. \tag{19}$$

At low substrate concentrations, the Michaelis-Menten rate equation Eq. (15) takes the linear form $v \approx (\frac{k_{cat}E_0}{K_M})S$. If we assume the reaction is operating in this linear regime, we can use this approximation to write

$$v = k_{forw}S - k_{back}P = \left(\frac{k_dE_0}{1 + q^{-1}}\right)S - \left(\frac{k_dE_0}{1 + q}\right)P, \tag{20}$$

$$v = k_d E_0 \left(\frac{Sq - P}{1 + q} \right), \tag{21}$$

which is Eqn. 12 for a perfect catalyst.

Expressing this as a linear rate equation will allow us to use the form of Eq. (10) to calculate the flux. Comparing Eq. (21) with Eq. (10) gives the mapping

$$K_i = \frac{k_d [E_i] q_i}{1 + q_i} \qquad , \qquad K_{-i} = \frac{k_d [E_i]}{1 + q_i}, \tag{22}$$

where now the rates depend on the diffusion-limited rate k_d, the enzyme concentration and the equilibrium constant of the reaction. Finally, substituting Eq. (22) into the expression for the flux, Eq. (10), yields an expression for flux of a pathway composed of these "perfect enzymes", where q is replaced by q' for any reactions coupled to the conversion of the external metabolites:

$$J = \frac{1}{D} \left([S_0] \prod_{i=1}^{n} q'_i - [S_n] \right), \tag{23}$$

where

$$D = \sum_{i=1}^{n} \frac{1}{K_i(E_i, q'_i)} \prod_{m=i}^{n} q'_m. \tag{24}$$

VI. OPTIMIZING THE ENZYME CONCENTRATIONS

Expression (23) for the pathway flux depends on the concentrations of the enzymes $[E_i]$, which are unknown. Since we aim to compare the maximal flux that can be carried by a given pathway, we optimize the enzyme concentrations $[E_i]$ so as to maximize the flux J, satisfying the constraints that (a) the total enzyme concentration is fixed, $\sum_j [E_j] = [E]_t$, and (b) the steady state metabolite concentrations, as computed using Eq. (9), fall within the prescribed acceptable range, $[S]_{min} \leq [S_k] \leq [S]_{max}$ for all k. This optimization was performed using Powell's method[15].

VII. ENUMERATING PATHWAYS

We enumerate pathways between G3P and pyruvate in our network using the depth-first search algorithm. The number of pathways found increases approximately exponentially with pathway length (black line in Fig. S2). As we impose thermodynamic and biophysical constraints, the number of pathways is reduced, but still increases with pathway length. In the glycolytic direction, the red line in Fig. S2 shows the number of pathways that produce at least 2 ATP molecules, the green line shows the number of pathways that produce at least 2 ATP molecules and carry a positive flux under typical physiological conditions (as defined by the set of parameters described in the figure caption), and the blue line shows the number of pathways that produce at least 2 ATP molecules, carry a positive flux, and have intermediate concentrations between 1nM and 0.5M. The dashed blue line shows the number of gluconeogenic paths which carry a positive flux under these typical physiological conditions and also have intermediate concentrations within this range.

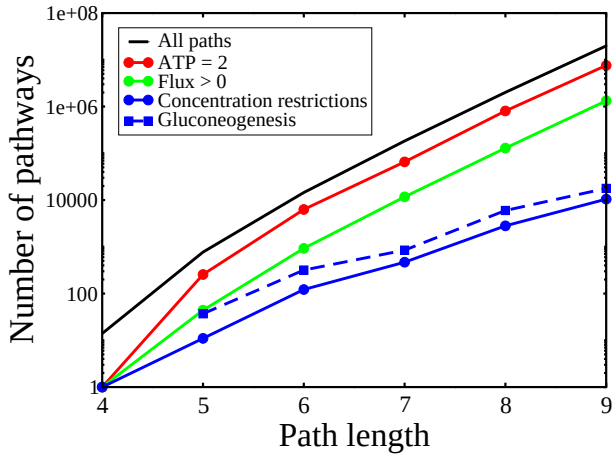

FIG. S 2: The number of pathways connecting G3P and pyruvate in our network increases exponentially as a function of the path length (number of reactions). See text for details. The "typical physiological conditions" discussed, correspond to [G3P]=0.019mM, [PYR]=0.051mM , [ATP]=9.6mM, [ADP]=0.56mM, [AMP]=0.28mM, [NAD]=2.6mM, [NADH]=0.083mM, [Pi]=1.0mM, [PPi]=0.001mM, and $[CO_2]$=1.0mM[16,17])

VIII. SAMPLING THE PARAMETER SPACE

To evaluate the relative performance of our candidate pathways (Fig. 2 in the main text), we sampled the space of external metabolite concentrations, together with the concentrations of G3P and pyruvate in the following way. The parameter set was defined as in Table SIV; the ratios [ATP]/[ADP] and [NAD]/[NADH], for example, are treated as single parameters since they occur in these combinations in the flux expression. Each parameter was sampled logarithmically over a wide range, covering several orders of magnitude above and below its typical physiological concentration; the ranges sampled are given in Table SIV. We used a wider range for the concentrations of inorganic phosphates and pyrophosphate since available literature data for these values show much greater variability than the other parameters. The results of our work are robust to the choice of parameter range. Indeed, if we restrict these ranges to a tighter region around the physiological values we find that the real pathways are ranked higher relative to the alternatives, as is clear from Fig. 2 in the main text.

IX. RELAXING THE ASSUMPTION OF OPTIMIZED ENZYME CONCENTRATIONS

In calculating the flux carried by our candidate pathways (e.g. in Fig. 2 of the main text), we assumed that the enzyme concentrations were optimized to maximise the flux carried by a given pathway (see Section IV). Our main results are, however, independent of this assumption. In Fig. S3 we show the relative performance of our candidate pathways in the glycolytic and gluconeogenic directions, this time calculating the flux assuming equal concentrations of all enzymes along the pathway - *i.e.* using Eq. (23) to compute the pathway flux, setting all the $[E_i]$ to an arbitrary constant (the constant is arbitrary because we only compute relative fluxes). The picture that

Parameter	Range sampled
[source]	1μM to 1mM
[source]/[product]	0.001 to 1000
[ATP]/[ADP]	0.001 to 1000
[NAD]/[NADH]	0.001 to 1000
[AMP]	1μM to 1mM
[Pi]	1μM to 100mM
[PPi]	0.1nM to 10mM
[CO_2]	1μM to 1mM

TABLE S IV: Parameters sampled independently in our analysis and the ranges over which they were sampled. The source and product concentrations correspond to G3P and PYR in the glycolytic direction and vice versa in the gluconeogenic direction.

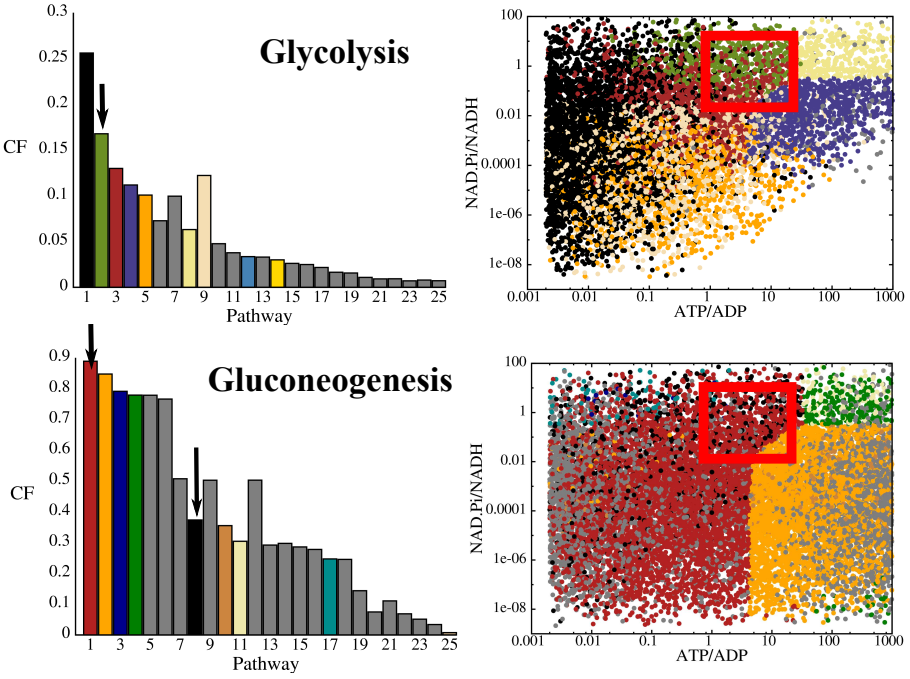

FIG. S 3: Comparative flux, CF, for each pathway averaged over all parameter space (left) and best performing pathway in different regions of parameter space (right) where now the enzyme distribution has not been optimized to maximize flux, but instead all enzymes of the pathway are present at the same concentration. We find a very similar outcome to the case of the optimized enzyme distribution (Fig. 2 in the main text): the real pathways (shown by arrows in the left panels) outperform the alternatives in the physiological region of parameter space (red boxes in right panels). The colour coding is as in Fig. 2 of the main text.

emerges is very similar to that from the optimized enzyme distribution, with the real glycolytic and

gluconeogenic pathways (arrows) performing best in the typical physiological region of parameter space (red box).

X. THE PPS AND PPDK GLUCONEOGENIC PATHWAYS PERFORM DIFFERENTLY IN DIFFERENT REGIONS OF PARAMETER SPACE

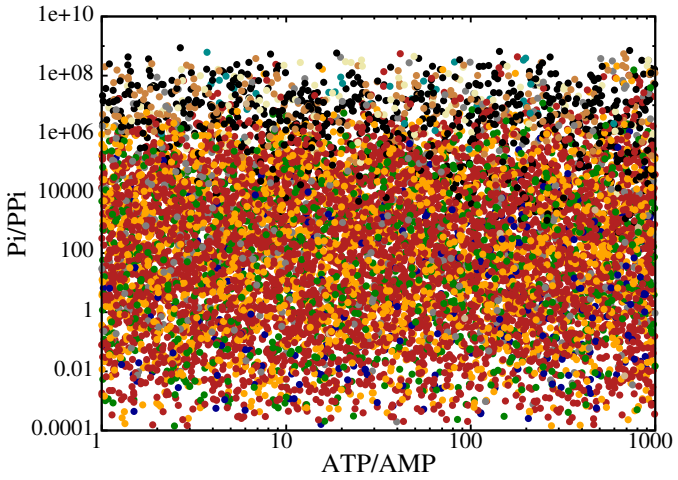

FIG. S 4: Scatter plot showing best performing gluconeogenic pathway in different regimes of parameter space. Colours as in Fig. 2 of the main text. The same data set used to generate Fig. 2 of the main text is used here, but now plotting the [ATP]/[AMP] and [Pi]/[PPi] ratios. We find that the dominant factor distinguishing between the pps (red) and ppdk (black) gluconeogenic pathways is the ratio of inorganic phosphate to pyrophosphate. Note that although the ppdk path only appears for [Pi]/[PPi] ratios of greater than 10^5, this does not mean that the pathway is infeasible for values lower than this, just that for these lower values other pathways outcompete it with respect to the size of their flux.

Fig. 2D of the main text shows that both the pps (red) and ppdk (black) natural variants of the gluconeogenic trunk pathway outperform the alternative pathways in the physiological region of parameter space, as defined by the cellular energy and redox states. In both the natural pathways, the transformation of pyruvate to PEP is coupled to the conversion of ATP to AMP. In the pps route, this releases inorganic phosphate (via the phosphoenolpyruvate synthetase enzyme, EC 2.7.9.2) ATP+H_2O → AMP+Pi; in the ppdk route, inorganic phosphate is consumed and pyrophosphate is released (by the pyruvate, phosphate dikinase enzyme, EC 2.7.9.1), ATP+Pi → AMP+PPi. Fig. S4 shows that the relative performance of these two natural variants in our analysis depends critically on the concentrations of inorganic phosphate and pyrophosphate. The reason for this is clear from consideration of the above external metabolite couplings. A high concentration of inorganic phosphate [Pi] will reduce the thermodynamic favourability of the pps reaction, while a high [Pi] (more specifically, a high [Pi]/[PPi] ratio) will in fact increase the favourability of the ppdk reaction.

XI. ALTERNATIVE TRUNK PATHWAYS FOUND IN OUR ANALYSIS

Glycolytic alternative pathways

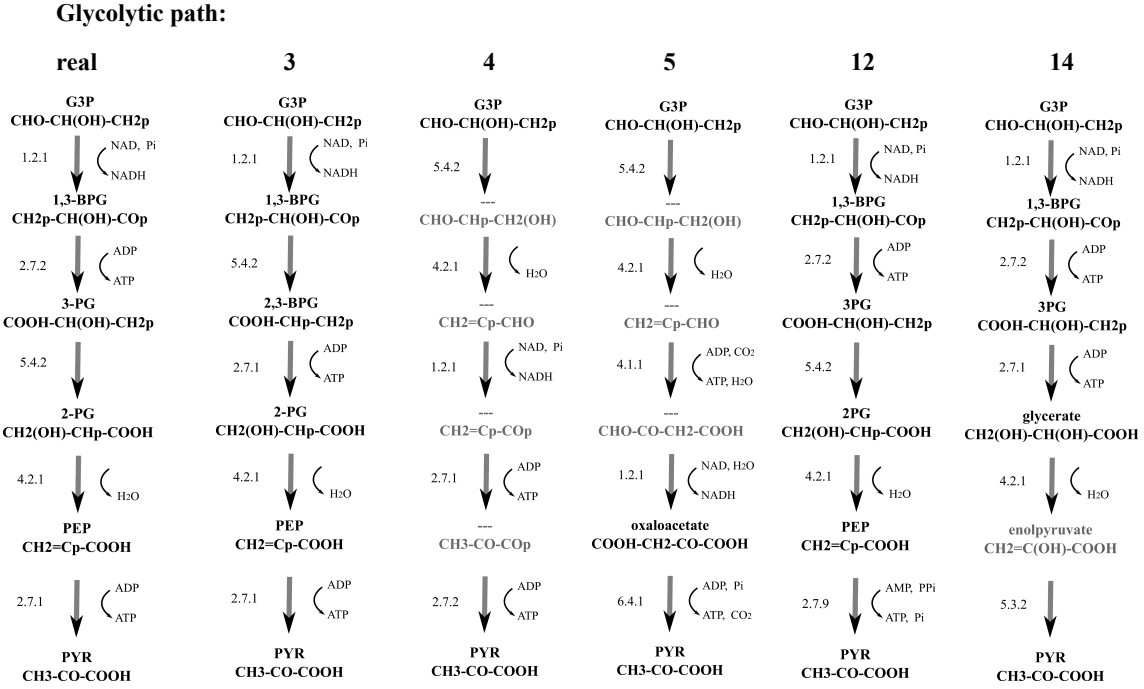

FIG. S 5: Alternative glycolytic pathways found in our analysis. Glycolytic pathways 3, 4, 5, 12 and 14 are shown, labelling all external metabolite couplings and first 3 EC numbers. Compound names and chemical formulae shown in bold are present in the KEGG database.

Fig. S5 shows alternative pathways 3-5 in the glycolytic direction, which all perform well in our analysis, in different regions of parameter space. These pathways are discussed in the main text. Pathway 1, which outperforms the real pathway when averaged over parameter space, is shown in Fig. 2E (right), main text.

Two other alternative glycolytic trunk pathways, 12 and 14, are also of interest and are shown in Fig. S5. Pathway 12 is the same as the real glycolytic trunk pathway, except that in the final step, ATP is generated from AMP and pyrophosphate (via pyruvate, phosphate dikinase) rather than from ADP – *i.e.* it is the exact reverse of the natural ppdk gluconeogenic pathway. This effectively results in a greater energetic yield than the natural glycolytic pathway since an ATP is being recovered from an AMP directly, rather than through the action of the adenylate kinase enzyme (EC 2.7.4.3) which involves the consumption of another ATP molecule. This pathway actually corresponds to a glycolytic variant observed in some anaerobic eukaryotes, in which glycolysis is the primary mode of ATP production[18]. Its flux and feasibility depend on the cellular [AMP] and [PPi] concentrations, which is not the case for the real pathway.

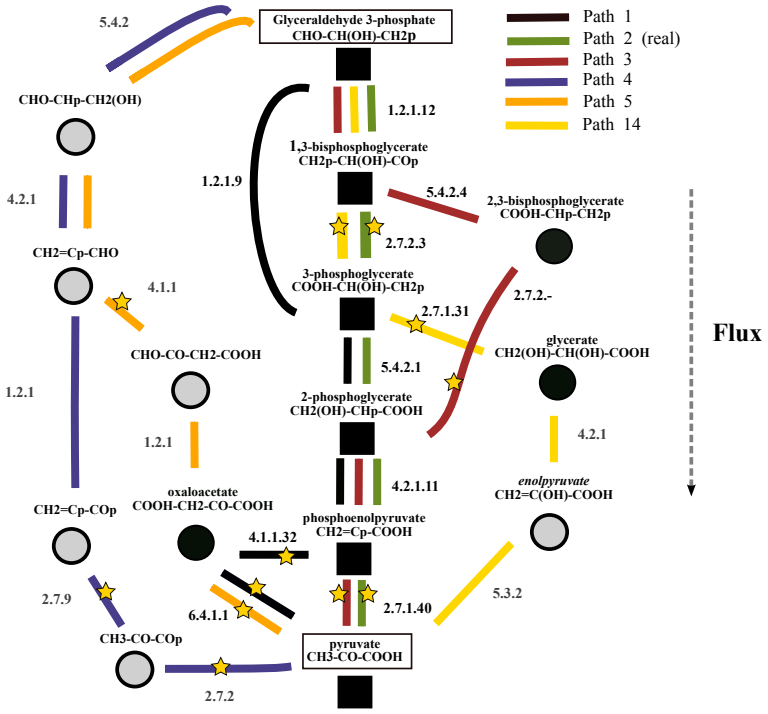

FIG. S 6: Expanded version of the schematic view of the glycolytic pathways found in Fig. 3 of the main text. Intermediate metabolites are shown by symbols; black squares indicate metabolites in the real trunk pathway, black circles indicate metabolites present in KEGG, and grey circles show metabolites generated by our program but not found in KEGG. Black EC numbers show reactions which exist in the KEGG database, grey numbers show those that do not. Yellow stars indicate ATP-producing reactions. Note that path 12 follows the same route as the real path (path 2, green) with the exception that the final reaction is catalyzed by ppdk, EC 2.7.9.1, instead of pyruvate kinase, EC 2.7.1.40.

Pathway 14 converts 3-PG to glycerate, and then to enolpyruvate. Enolpyruvate is actually unstable in solution and converts spontaneously to pyruvate in a tautomerization reaction. It is likely therefore that in reality this pathway would only require 4 enzymes (instead of 5 as in the real pathway), while still producing 2 ATP molecules. This pathway might also generate a higher flux of ATP than suggested by our analysis, due to the rapid spontaneous depletion of the enolpyruvate. Moreover, almost all the enzymes needed for this pathway exist in nature. The only exception is the enzyme which catalyzes the dehydration reaction of glycerate, which is not present in KEGG (Fig. S6). It is not unreasonable to suppose that such a simple dehydration might result from the uncharacterized, non-primary action of an existing enzyme.

Fig. S6 shows a schematic view of glycolytic pathways 1-5, 12 and 14, together with the EC numbers of the corresponding reactions. Full EC numbers shown in black indicate enzymes which appear in the KEGG database having been found in natural organisms; grey indicates their absence in KEGG. As discussed in the main text, notice that all of the enzymes in pathway 1 are known to exist in biology, with the final two (EC 4.1.1.32 and 6.4.1.1) usually associated with a gluconeogenic

role in eukaryotic cells (as in the standard textbook picture of gluconeogenesis). Note the first step, involving the oxidation of glyceraldehyde 3-phosphate to 3-phosphoglycerate (EC 1.2.1.9), actually couples with NADP reduction, wherease in our network we have only included NAD. A possible explanation for why this set of reactions has not been observed in nature emerges from our analysis – the problematic intermediate metabolite concentrations that arise due to the highly exergonic initial oxidation.

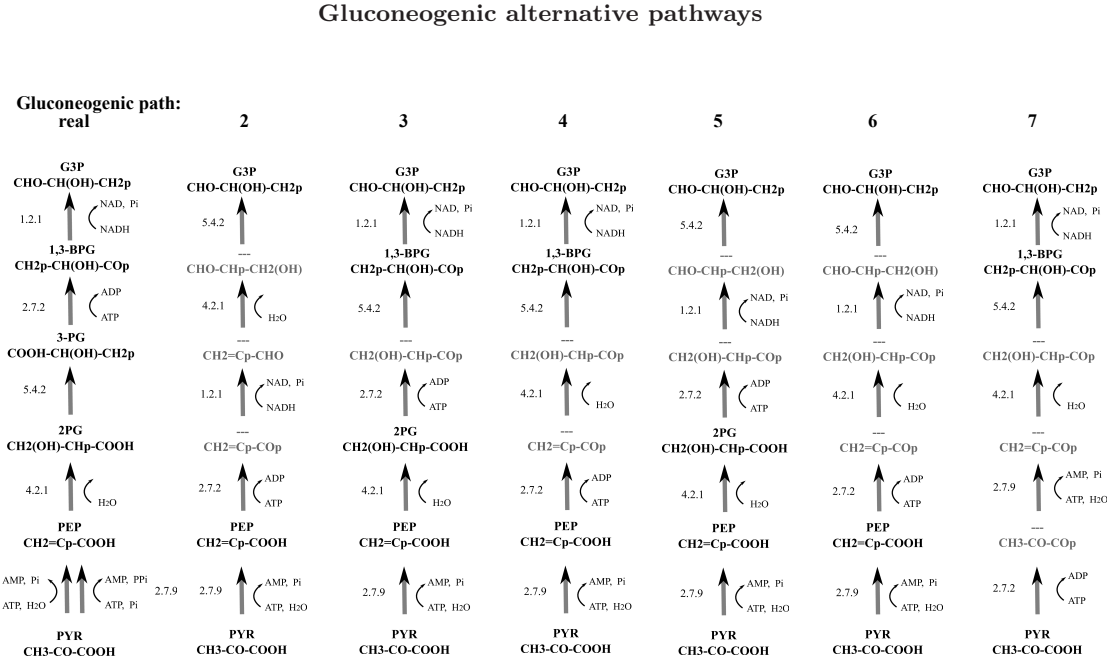

FIG. S 7: Gluconeogenic pathways 2-7,showing all external metabolite couplings and first 3 EC numbers . Compound names and chemical formulae in black are present in KEGG database, grey are not. Note that these top-performing gluconeogenic pathways all utilize the same set of reaction classes.

Fig. S7 shows alternative pathways 2-7 in the gluconeogenic direction, which all perform well in our analysis, in different regions of parameter space. These pathways are discussed in the main text. Fig. S8 shows a schematic view of the best-performing gluconeogenic pathways, labelled with the corresponding EC numbers. We note that there is less variation in the top-performing gluconeogenic pathways compared to the glycolytic ones. For example, the top 7 pathways in our analysis all share the pps reaction as their first step as well as making use of the same set of reaction classes.

XII. THE 6-REACTION GLUCONEOGENIC PATH VIA OXALOACETATE OUTPERFORMS ALTERNATIVES IN THE ABSENCE OF DIKINASE REACTIONS

In this paper, our analysis has focused on prokaryotic metabolism. In prokaryotes, the gluconeogenic trunk pathway has 5 steps, with the initial step, the transformation of pyruvate to PEP occurring in one step, coupled to the conversion of ATP to AMP (via the pps or ppdk

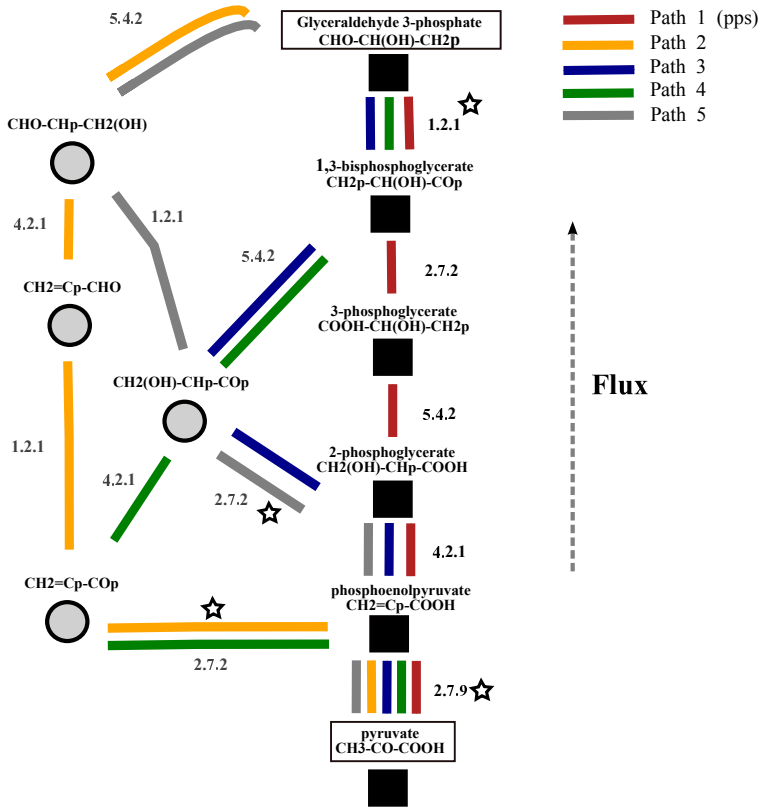

FIG. S 8: Schematic view of 5 best gluconeogenic paths where colours correspond to pathways as in Fig. 2 main text and Fig. S3. Empty stars indicate ATP consumption. Black boxes show trunk pathway metabolites, grey circles show metabolites not present in KEGG. Note that flux here flows in an upward direction, from pyruvate to G3P.

enzymes). The textbook picture of mammalian gluconeogenesis is slightly different. Here, pyruvate is converted to PEP in two steps: pyruvate carboxylase (EC 6.4.1.1) converts pyruvate to 4-carbon oxaloacetate, which is then converted to PEP by phosphoenolpyruvate carboxykinase (EC 4.1.1.32). The first step is driven by the dephosphorylation of ATP to ADP while the second step is driven by GTP which is converted to GDP. This pathway therefore has 6 steps and in total consumes 3 "energy units" (2 ATPs are converted to ADP while 1 GTP is converted to GDP).

Because this is a 6-step pathway, it was not considered in our main analysis. However, when we repeat our analysis in the absence of dikinase reactions (*i.e.* not allowing the conversion of ATP to AMP), we find that 5-step gluconeogenic pathways are no longer feasible under typical physiological conditions since they can now only consume 2ATP molecules (both coupling to ATP $\rightarrow$ ADP). Fig. S9 shows that in this case, the textbook mammalian pathway outperforms all alternatives under physiological conditions (note that since we do not include GTP in our analysis, the "natural" pathway here has 3 ATP $\rightarrow$ ADP conversion steps, however the energy released in the hydrolysis

of ATP and GTP are similar.)

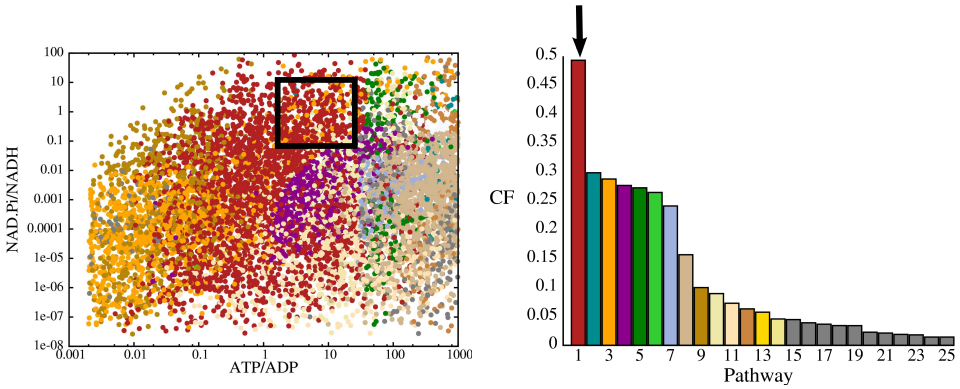

FIG. S 9: Repeating our analysis without dikinase reactions (*i.e.* not allowing the conversion of ATP to AMP). The textbook 6-step gluconeogenic pathway (with ATP substituted for GTP) is now shown in red (indicated by the arrow). This pathway outperforms the alternatives when its comparative flux is averaged across the whole parameter space, and also performs best in the physiological region of parameter space. Plots were produced using a flat enzyme distribution.

XIII. GLUCONEOGENESIS WITHOUT CONSTRAINTS ON METABOLITE CONCENTRATIONS

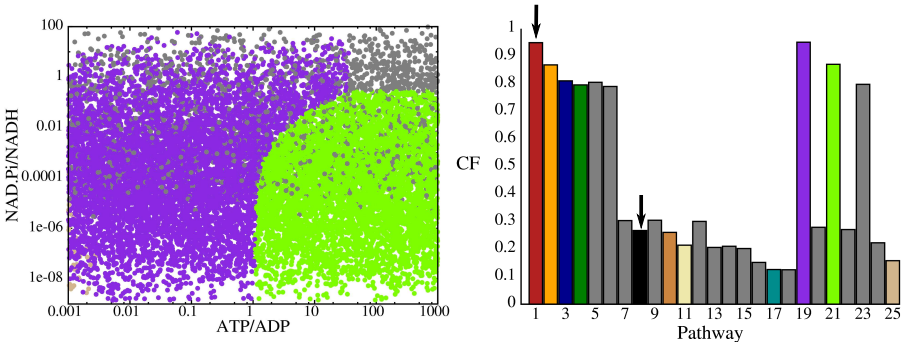

FIG. S 10: Comparison of gluconeogenic paths when no restrictions on the intermediate concentrations are imposed. Arrows indicate pps (red) and ppdk (black) pathways, with colours as in Fig. 2 main text and Fig. S3, except for now additionally colouring paths 19 and 21. Paths 19 and 21 now dominate the scatter plot but note that their relative performance, as shown by the plot of the cumulative flux (CF), does not greatly exceed that of the real pathway.

Fig. S10 shows the results of our analysis for paths in the gluconeogenic direction, when we do not include constraints on the intermediate metabolite concentrations. As for glycolysis (Fig. 4 main text), we see that removing these constraints has a dramatic effect on our results. In particular,

pathways 19, 21 and 23 perform well over the entire parameter space. All three of these pathways have very large kinks in their thermodynamic profiles. For instance, pathway 21 has a reaction with a free energy change of $\Delta G = -32\text{kJmol}^{-1}$ directly followed by a reaction with $\Delta G = +36\text{kJmol}^{-1}$.

We note that our analysis here and in the main text neglects several factors which might alter intermediate metabolite concentrations. In particular, we have not included branching reactions, which could reduce the concentration of problematic metabolites. In addition, factors like substrate channelling and spatial segregation may also be important in real cells.

XIV. REPEATING OUR ANALYSIS WITH TIGHTER METABOLITE CONCENTRATION CONSTRAINTS

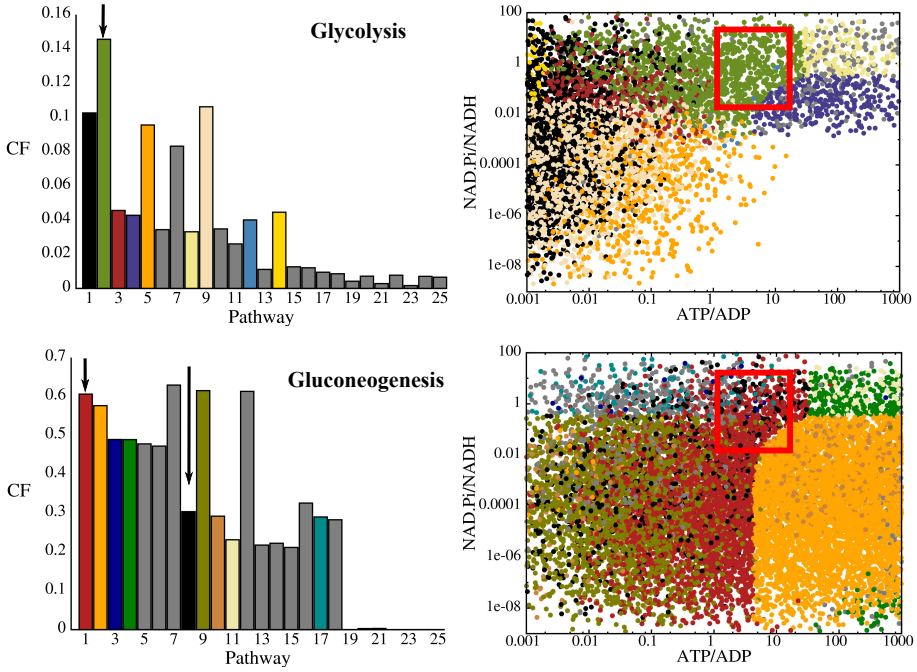

FIG. S 11: The results of the comparison, where now the lower and upper limits on the intermediate concentrations are 10^{-7}M and 10^{-2}M respectively. Note that a flat enzyme profile was used in generating these plots. Colours as in Fig. 2 main text. Arrows point at real glycolytic pathway (green) and gluconeogenic pps (red) and ppdk (black) pathways.

Fig. S11 shows the results of repeating our analysis with tighter restrictions on the intermediate metabolite concentrations; here we constrain these concentrations to lie in the range 10^{-7} to 10^{-2}M rather than 10^{-9} to 5×10^{-1}M as in the analysis in the main text. Under these more restrictive constraints, the real glycolytic pathway actually outperforms the alternatives over a wider region of parameter space (compare Fig. S11 to Fig. 2 of the main text). In the glycolytic direction, the real pathway now outperforms pathway 1 (which has the strongly exergonic first reaction, see main text) when its comparative flux is averaged across the whole parameter space. This seems to

suggest that not only does the real trunk pathway carry a high flux, but also prevents the buildup of large intermediate metabolite concentrations more effectively than alternative pathways. An ad hoc criterion for ranking paths based on the flatness of their thermodynamic profiles has been proposed in previous work[19]; our work suggests that this may indeed be a reasonable criterion, which can be addressed quantitatively with our analysis, and which may have played a role in evolutionary selection. It is interesting to note that glycolytic paths 12 and 14, discussed above, also perform relatively better under these stricter limits on the metabolite concentrations.

The gluconeogenic comparison shows three pathways (7, 9 and 12) now performing better relatively, when averaged over all parameter space, but note that these appear in the lower [ATP]/[ADP] regions of the scatter plot. In the region corresponding to typical physiological conditions (red box), the two pathways performing the best are again the real pps, and to a lesser extent, ppdk paths.

---

[1] Ronimus RS, Morgan HW (2002) Distribution and phylogenies of enzymes of the embden-meyerhof-parnas pathway from archaea and hyperthermophilic bacteria support a gluconeogenic origin of metabolism. *Archaea* 1:199–221.

[2] Berg J, Tymoczko J, Stryer L (2002) *Biochemistry* (W H Freeman and Company), 5th edition.

[3] Bar-Even A, Flamholz A, Noor E, Milo R (2012) Rethinking glycolysis: on the biochemical logic of metabolic pathways. *Nat Chem Biol* 8:509–517.

[4] Webb E (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* (Academic Press, San Diego), 6 edition.

[5] Mavrovouniotis ML (1991) Estimation of standard gibbs energy changes of biotransformations. *Journal of Biological Chemistry* 266:14440–14445.

[6] Mavrovouniotis ML (1990) Group contributions for estimating standard gibbs energies of formation of biochemical compounds in aqueous solution. *Biotechnology and Bioengineering* 36:1070–1082.

[7] Jankowski MD, Henry CS, Broadbelt LJ, Hatzimanikatis V (2008) Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophysical Journal* 95:1487 – 1499.

[8] Alberty RA (2006) *Biochemical Thermodynamics: Applications of Mathematica*, Methods of Biochemical Analysis.

[9] Alberty RA (1998) Calculation of standard transformed entropies of formation of biochemical reactants and group contributions at specified pH. *The Journal of Physical Chemistry A* 102:8460–8466.

[10] Alberty R (1983) Chemical thermodynamic properties of isomer groups. *Industrial & Engineering Chemistry Fundamentals* 22:318–321.

[11] Heinrich R, Montero F, Klipp E, Waddell TG, Meléndez-Hevia E (1997) Theoretical approaches to the evolutionary optimization of glycolysis: Thermodynamic and kinetic constraints. *European Journal of Biochemistry* 243:191–201.

[12] Heinrich R, Hoffmann E (1991) Kinetic parameters of enzymatic reactions in states of maximal activity; an evolutionary approach. *J. theor. Biol* 151:249–283.

[13] Albery JW, Knowles JR (1976) Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry* 15:5631–5640.

[14] Pettersson G (1989) Effect of evolution on the kinetic properties of enzymes. *European Journal of Biochemistry* 184:561–566.

[15] Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, New York, NY, USA), 3rd edition.

[16] Bennett BD, et al. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in escherichia coli. *Nat Chem Biol* 5:593–599.
[17] Garrett R, Grisham C (year?) *Biochemistry* (Belmont, CA: Thomson Brooks/Cole), 3rd edition, p 584.
[18] Feng XM, Cao LJ, Adam RD, Zhang XC, Lu SQ (2008) The catalyzing role of ppdk in giardia lamblia. *Biochemical and Biophysical Research Communications* 367:394 – 398.
[19] Cho A, Yun H, Park JH, Lee SY, Park S (2010) Prediction of novel synthetic pathways for the production of desired chemicals. *BMC Systems Biology* 4.