Detailed Methods

The Kyoto Encyclopedia of Genes and Genomics (KEGG) metabolite ligand library (*1*) was used in this study. The two-dimensional (2D) structures of metabolites were converted into dockable three-dimensional (3D) form using the software LigPrep-v2.01 (Schrödinger LLC). The LigPrep software adds hydrogens, expands chiral and tautomeric forms, predicts possible protonation states, and minimizes all resulting structures. After removing duplicate entries, the final library contained 18,568 ligands.

Prior to docking, all enzyme structures were subjected to a protein preparation step using the protein-preparation module provided in Maestro (Schrödinger LLC). During the protein preparation step, hydrogens were added and optimized to make better hydrogen bonding interactions, protonation states of titratable residues such as His, Asp, and Glu were optimized, and side chain rotamers of Asn and Gln were examined and corrected. We also energy minimized all atoms of the protein and ligand such that heavy atoms do not move beyond 0.3 Å. We deleted water molecules in all but the glucokinase enzyme. In the glucokinase enzyme structure co-crystallized with β-D glucose (pdb id 1sz2), we noticed that one water molecule was hydrogen bonding to β-D glucose and residues Thr137 and Gly138 of the enzyme. An analogous water molecule was also present in the *apo* structure (PDB id 1q18), making hydrogen-bonding interactions with the same residues (*2*). We believe that this water molecule is structural and therefore it was retained during docking and rescoring calculations.

After the protein preparation step, we docked the KEGG library against the active site of all enzymes in the glycolysis pathway of *E. coli* using Glide-v5.02 (Schrödinger LLC) (*3*). We used the standard default parameter settings for all docking calculations. Docking scoring (i.e., ranking) was performed using the standard-precision mode. Docked poses were energy minimized before scoring.

After the docking calculation was finished, the pose of each docked ligand with the most favorable docking score was subjected to the MM-GBSA rescoring. Unlike the empirical scoring function used in the docking program, the physics-based MM-GBSA rescoring uses a molecular-mechanics based energy function that includes.van-der Waals and electrostatic interactions, and treats ligand and receptor desolvation at the level of generalized Born implicit solvation. We specifically use the OPLS all-atom force field for the solute and a generalized Born implicit solvent model, as described in detail elsewhere (*5*). During rescoring, the ligand was energy minimized in the active site so that it could make more favorable hydrogen bonding and van-der Waals interactions with the enzyme. Energy minimization and evaluations were performed using the Protein-Local-Optimization Program (PLOP, marketed commercially as Prime by Schrödinger LLC) (*4*). In these calculations, we held the protein rigid, although this does not need to be the case in general.

From the "ligand-minimized" ligand-protein complex structure, we separated structures of ligand and protein and evaluated their energies. Note that the protein structure is the same for all docked metabolite ligands and therefore its energy needs to be calculated only once. Finally, we calculated the relative binding energy of each ligand by subtracting energies of the ligand and the protein from that of the ligand-protein complex. The relative binding energies were then used to rank the ligands.

In the present work, we considered all available structures regardless of organism and used the native *E. coli* structure only if a ligand and/or cofactor bound *holo* structure was available, otherwise we built homology models using the structure of the closest homologous ligand or cofactor bound structure. Homology models for the *E. coli* sequences were built using PLOP, using the default protocol in which side chain rotamers of non-conserved amino acids were energy optimized and all atoms not present in the template structure were subjected to energy minimization (*4*). No other refinement steps were performed, other than the same 'protein preparation' protocol applied to crystal structures, as discussed above. Ligand and any cofactor or metal ions present in the active site of the homolog were also included during the model building. Cognate substrates were found in the following four enzyme structures: glucokinase (step I), fructose bisphosphate aldolase (step III), glyceraldehyde 3-phosphate dehydrogenase (step VI) and phosphoglycerate mutase (step VIII). The other structures either had product or an intermediate-like inhibitor present in the active site.

In Table S1 we list all enzymes involved in the glycolysis pathway, their Genbank identification number (gi), substrates, structures used in our calculation, and co-crystallized ligands. Note that the product of one enzyme is the substrate of the next enzyme in the metabolic pathway, with the exception of the fructose bisphosphate aldolase, which generates two products.

(1) Goto, S., Okuno, Y., Hattori, M., Nishioka, T., and Kanehisa, M. (2002) LIGAND: Database of chemical compounds and reactions in biological pathways, *Nucleic Acids Res*. **30**, 402-404.

(2) Lunin, V.V., Li, Y., Schrag, J.D., Iannuzzi, P., Cygler, M., and Matte, A. (2004) Crystal structures of Escherichia coli ATP-dependent glucokinase and its complex with glucose, *J. Bacterol*. **186**, 6915-6927.

(3) Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P., and Shenkin, P.S. (2004) Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J. Med. Chem*. **47**, 1739-1749.

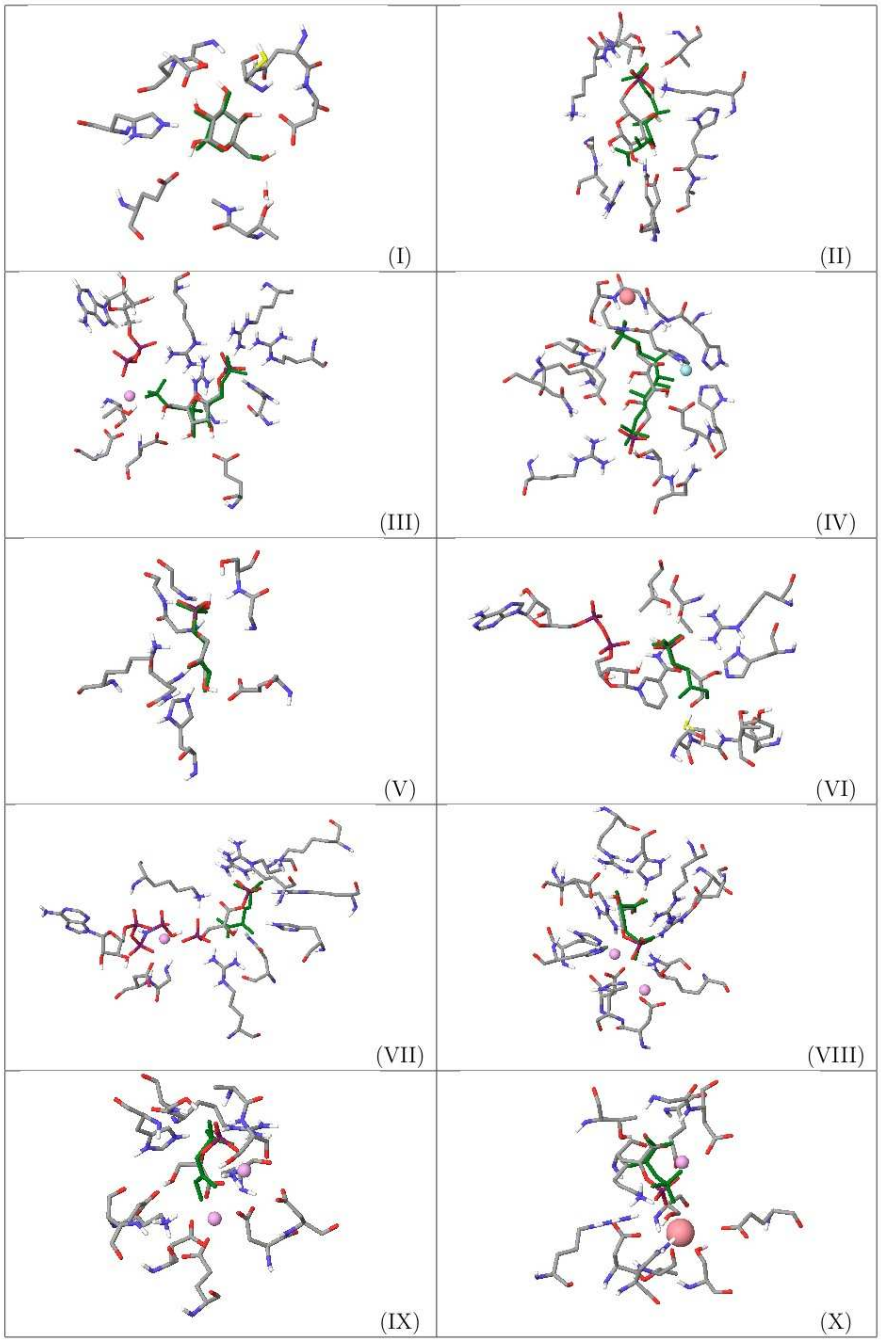
(4) Zhu, K., Shirts, M.R., Friesner, R.A., and Jacobson, M.P. (2007), Multiscale optimization of a truncated Newton minimization algorithm and application to proteins and protein-ligand complexes, *J. Chem. Theor. Comput.* **3**, 640-648.

(5) Jacobson, M.P., Kaminski, G.A., Friesner, R.A., and Rapp, C. S. (2002) Force field validation using protein side chain prediction, *J. Phys. Chem*. **B106**, 11673-11680.

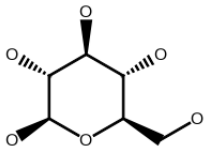
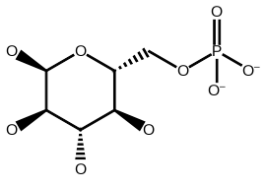
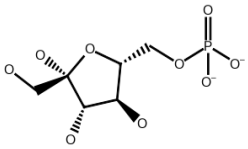
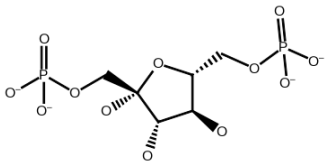
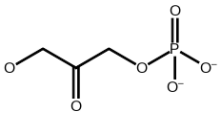
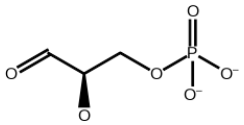
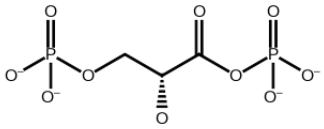
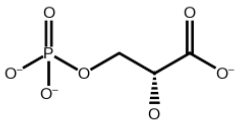
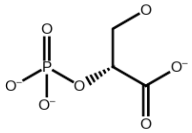
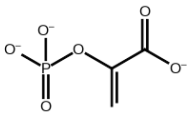
Supplementary Table S1. Details of test set.

Step	Enzyme	gi number	Substrate	Structure or template	SeqID	Co-crystallized Ligands[‡]	
						Ligand	Cofactors
I	Glucokinase	16130320	Glucose	1sz2	100%	Glucose	
II	Phosphoglucose isomerase	16131851	Glucose-6-phosphate	2cxr (*M. musculus*)	64%	6-Phospho gluconic acid	
III	Phosphofructokinase	16131754	β-D-Fructose-6-phosphate	1pfk	100%	β-D Fructose 1,6 bisphosphate	ADP, Mg^{2+}
IV	Fructose bisphosphate aldolase	16130826	β-D-Fructose-1,6-bisphosphate	3elf (*M. tuberculosis*)	40%	1,6 Fructose bisphosphate	Na^{+}, Zn^{2+}
V	Triosephosphate isomerase	16131757	Dihydroxyacetone phosphate	7tim (*S. cerevisiae*)	45%	Phospho glycolo hydroxamic acid	
VI	Glyceraldehyde 3-phosphate dehydrogenase	16129733	D-Glyceraldehyde 3-phosphate	1nqa (*G. sterothermophilus*)	58%	D-Glyceraldehyde 3-phosphate	NAD^{+}
VII	Phosphoglycerate kinase	16130827	1,3-Bisphospho-D-glycerate	1vpe (*T. maritima*)	45%	3-Phospho glycerate	ANP, Mg^{2+}
VIII	Phosphoglycerate mutase	16131483	3-Phospho-D-glycerate	1ejj (*G. stearothemophilus*)	49%	3-Phospho glycerate	Mg^{2+}
IX	Enolase	16130686	2-Phospho-D-glycerate	1ebg (*S. cerevisiae*)	51%	Phosphono aceto hydroxamic acid	2 Mg^{2+}
X	Pyruvate kinase	16139632	Phosphoenol-pyruvate	1aqf (*O.cuniculus*)	46%	L-Phospho lactate	K^{+}, Mg^{2+}

[‡] Co-crystallized ligands present in the template structure. Co-factors were retained in the docking calculations.



Supplementary Figure S1. Binding pose of the metabolite after rescoring (CPK color) superimposed on the co-crystallized ligand pose from the crystal structure of the enzyme, or the ligand pose from the template structure used for the homology model (green color). Roman numerals correspond to the glycolysis step shown in Table 1.

 I) Glucose	 II) Glucose 6-phosphate
 III) β-D Fructose 6-phosphate	 IV) β-D Fructose 1,6 bisphosphate
 V) Dihydroxy acetone phosphate	 VI) Glyceraldehyde 3-phosphate
 VII) 1,3 Bisphospho D-glycerate	 VIII) 3-Phospho D-glycerate
 IX) 2-Phospho D-glycerate	 X) Phosphoenol pyruvate

Supplementary Figure S2. Metabolites involved in the 10 major steps of the glycolysis pathway.