Thermodynamics drives coenzyme redundancy in metabolism

Joshua E. Goldford[1*], Ashish B. George[2], Avi Flamholz[3] and Daniel Segrè[4*]

[1]Physics of Living Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[2]Carl R. Woese Institute for Genomic Biology and Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[3]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA

[4]Bioinformatics Program, Biological Design Center, Department of Biomedical Engineering, Department of Physics, Department of Biology, Boston University, Boston MA, 02215, USA

[*]Corresponding author

Email: goldford@mit.edu, dsegre@bu.edu

Running Title: Thermodynamics drives coenzymes redundancy in metabolism

Keywords: Metabolism, evolution, coenzymes, constraint-based modeling

Abstract

Coenzymes redistribute a variety of resources (e.g., electrons, phosphate groups, methyl groups) throughout cellular metabolism. For a variety of reactions requiring acceptors or donors of specific resources, there often exist degenerate sets of molecules (e.g., NAD(H) and NADP(H)) that carry out similar functions. Several hypotheses can explain the persistence of coenzyme degeneracy, but none have been tested quantitatively. Here, we use genome-wide metabolic modeling approaches to decompose the selective pressures driving enzymatic specificity for coenzymes in the metabolic network of *Escherichia coli*. Flux balance modeling predicts that only two enzymes (encoded by *leuB* and *pdxB*) are thermodynamically constrained to use NAD(H) over NADP(H). In contrast, structural and sequence analyses reveal widespread conservation of residues that retain selectivity for NAD(H), suggesting that additional forces drive enzyme specificity. Using a model accounting for the cost of oxidoreductase enzyme expression, we find that coenzyme redundancy universally reduces the minimal amount of protein required to catalyze coenzyme-coupled reactions, inducing individual reactions to strongly prefer one coenzyme over another when reactions are near thermodynamic equilibrium. We propose that partitioning of flux across multiple coenzyme pools could be a generic phenomenon of cellular metabolism, and hypothesize that coenzymes typically thought to exist in a single pool (e.g., CoA) may exist in more than one form (e.g., dephospho-CoA).

Introduction

Electron transfer is a process critical for all living systems, where electron donors and acceptors are coupled to concomitantly facilitate oxidative energy transduction and reductive macromolecule biosynthesis (1). The ability to simultaneously achieve these functions is achieved by partitioning reaction flux through multiple diffusible coenzymes capable of transferring electrons between common functional groups (e.g. alcohols, aldehydes, and activated carboxylic acids) widely distributed throughout cellular metabolism, where such partitioning is assumed to be essential for all living systems (2–4). These key functions are carried out by a pair of redundant coenzymes capable of electron transfer at identical standard thermodynamic potential: nicotinamide adenine dinucleotide (NAD^+) and the phosphorylated derivative $NADP^+$ (or the reduced forms, NADH and NADPH). NAD(H) and NADP(H) are coupled to distinct sets of reactions, where NAD(H) is typically used in the breakdown of nutrients to generate precursors and ATP (catabolism), while NADP(H) is used in the reductive synthesis of macromolecules (anabolism) (1, 2). These functional roles have been attributed to differences in the *in vivo* redox potentials of NAD(H) and NADP(H), where NAD(H) and NADP(H) are typically poised in the oxidized (NAD^+) and reduced state (NADPH), respectively(5–9). This strategy is commonly assumed to be a universal way for cells to be able to simultaneously oxidize and reduce metabolites that might be challenging or impossible with just a single coenzyme (2, 3).

The assumption that coenzyme redundancy is essential in living systems leaves unanswered a set of fundamental questions about the nature and early evolution of metabolism: is the coupling of each reaction to a specific coenzyme a strong constraint for the proper functioning of metabolism, or can some coenzymes be switched with limited consequences? More generally, is the presence of two coenzymes poised in different redox states an absolute necessity for the proper functioning of a metabolic network, or could metabolism in principle operate with only one coenzyme? If one coenzyme is sufficient, is it possible to quantitatively explain how evolution could gradually lead to the selective advantage of two coenzymes? And is the rise of

multiple coenzymes an idiosyncratic step along a complex historical process, or a predictable universal feature of enzyme-driven biochemical networks?

Prior results from a variety of experiments and molecular analyses highlight the strong sensitivity of biological systems to the balance of different redox currency coenzymes, but do not rule out the possibility of a simpler precursor to the NAD(H)/NADP(H) system. NAD(P)-dependent oxidoreductases can display strong selectivity for either NAD(H) or NADP(H), and perturbations to this specificity can strongly affect cellular phenotypes and fitness(10, 11). For example, it has been proposed that the emergence of an NADP(H)-dependent isocitrate dehydrogenase (IDH) from an ancestral NAD(H)-dependent enzyme in prokaryotes was driven by insufficient production of NADPH during growth on acetate(12). More generally, specificity for NAD(H) or NADP(H) for some enzymes can strongly influence intracellular reaction flux distributions (10), serving as the basis for metabolic engineering strategies to dramatically improve yields of valuable metabolic byproducts(10, 13–15). Importantly, recent work has demonstrated that specificity for NAD(P)H can emerge in laboratory evolution studies, suggesting that coenzyme preference might be a highly adaptable feature of metabolic networks (16). However, it is unclear if certain enzymes are typically constrained to use only NAD(H) or NADP(H), and whether a single coenzyme could support the functioning of metabolism.

While an experimental assessment of the landscape of possible variants of metabolism with different coenzyme systems would be extremely challenging, one can leverage theoretical and computational approaches to help address these questions. The recent development and application of metabolic modeling has brought a quantitative and model-driven approach to the study of metabolic evolution (17–21), and has led to uncovering optimality criteria that may have shaped universal features of core metabolism, such as the emergence of both NAD(H) and NADP(H)(22). Although recent quantitative work has explored the logic for why NAD(P)H emerged as a prominent coenzyme in biochemistry(23), the forces that drove the emergence of both NAD(H) and NADP(H), rather than just a single coenzyme, remain unclear.

Here we use computational systems biology and various network-level models of metabolism to systematically probe the possible advantages of maintaining multiple pools of biochemically redundant redox cofactors. By taking advantage of computational experiments that would be extremely challenging experimentally, we show that the *Escherichia coli (E. coli)* metabolic network is predicted to be insensitive to coenzyme preference for the majority of enzymes, and that a metabolism with just a single coenzyme (rather than both NAD(H) and NADP(H)) is capable of producing biomass under a variety of environments. We found that switching coenzyme preferences may lead to significant metabolic impairment for only a small fraction of oxidoreductases, despite evidence of widespread sequence conservation of NAD(H)/NADP(H) selective residues. By comparing results of standard versus thermodynamically informed stoichiometric models we find that most of these constraints are caused by thermodynamic effects rather than stoichiometric balancing, suggesting that thermodynamics is the dominant selective pressure shaping the specificity of oxidoreductases. Given the very limited number of enzymes strongly coupled with a specific coenzyme based on the above analysis, we hypothesize that additional evolutionary forces may favor redundancy. In particular, we developed a model of enzyme cost as a function of enzyme-coenzyme specificity and found that coenzyme redundancy enables efficient usage of protein resources when reactions are near thermodynamic equilibrium. These results support the hypothesis that coenzyme redundancy is a universal feature of evolved metabolic networks. Our results highlight the role of thermodynamics in shaping the specificity of enzymes at both local (e.g., single enzyme) and whole-cell scales.

Results

Thermodynamics constrain only a small number of NAD-coupled oxidoreductases

We first explored the robustness of metabolism to altering enzyme specificity of individual oxidoreductases to coenzymes as a way of investigating whether coenzyme coupling constrains metabolic networks. To simulate the consequences of altering the NAD(P)-specificity of individual oxidoreductases, we used flux balance analysis (FBA)(24) to simulate the growth rates of *in silico* mutants, with altered coenzyme specificity under 117,180 different environments. Specifically, we altered the iJO1366 (25) *E. coli* genome-scale metabolic model as follows: first, we changed stoichiometric coefficients to replace, one by one, the original coenzyme stoichiometric coefficients with the alternative coenzyme. In other words, if a reaction chosen for a mutation was originally coupled with NAD(H), it will now become coupled with NADP(H) instead. Next, for every switched coenzyme coupling, we modified the directionality constraints for the reaction under consideration using knowledge of reaction free energies and bounds on intracellular metabolite concentrations, constraining NAD^+, NADH, $NADP^+$ and NADPH concentrations to *in vivo* measurements (Fig. 1a-b, Methods). We compiled a list of 76 genes encoding NAD(P)-coupled oxidoreductases (Supplementary Table 1) and modeled "mutations" that switch the coenzyme preference of an enzyme by changing the coenzyme stoichiometry for each gene independently and used the gene-to-reaction rules to update the model constraints. For each model with a "mutant" coenzyme preference, we computed maximal growth rates in 117,180 media conditions, spanning fermentation, aerobic respiration and anaerobic respiration using nitrate, nitrite, dimethyl sulfoxide (DMSO), trimethylamine *N*-oxide (TMAO) or fumarate, on combinations of 180 carbon sources and 93 nitrogen sources (Methods, Supplementary Table 2).

We computed the frequency of lethal *in silico* mutations across all media conditions (Methods) (Fig. 1c). To our surprise, the vast majority (97.3%, 8,103,217 of 8,323,596) of environment and

coenzyme mutation pairs were non-lethal, suggesting that metabolism is overall robust relative to specific coenzyme assignments to different reactions. However, of the 219,092 lethal mutant-environment pairs, all came from forcing one of six NAD(H)-coupled oxidoreductases to use NADP(H), rather than NAD(H). These genes included *leuB*, *pdxB*, *paaH*, *gatD*, *lgoD,* and *fucO*. We also decreased the range of the feasible concentration ratios to see if new dependencies emerged, and found environments where sugar alcohol oxidoreductases encoded by *srlD* and *mltD* require coupling to NAD(H) rather than NADP(H) (Supplementary Fig. 1). Notably, only two oxidoreductases encoded by the two genes *leuB* and *pdxB*, were predicted to universally require NAD(H) as a coenzyme rather than NADP(H). This is in agreement with extensive work by Dean and colleagues, who demonstrated that the gene product of *leuB*, isopropylmalate dehydrogenase, is universally coupled to NAD(H), and that introducing mutations to alter specificity towards NADP(H) always reduced fitness, consistent with our model predictions (26). Notably, all the observed lethal gene-environment pairs are alleviated upon relaxing the updated thermodynamic constraints in the network (Supplementary Fig. 2), demonstrating that the thermodynamic driving force is a dominant factor in constraining the coenzyme preference in our model.

Sequenced-inferred specificity of enzymes show widespread selectivity of endogenously coupled coenzymes

Flux modeling predictions suggest that while the maximum organismal growth rate is insensitive to the coenzyme specificity for the majority of enzymes, a small number of enzymes are expected to be highly constrained to use a single coenzyme, limited to NAD(H)-coupled oxidoreductases. To determine if these model predictions are consistent with measures of enzyme specificity, we used a combination of protein structure and bioinformatic analysis to assess the coenzyme binding preferences of oxidoreductases both in *E. coli*, and across phylogenetically divergent species.

We first investigated the two genes predicted to be universally constrained to NAD(H): *leuB* and *pdxB*. While *leuB* has been previously shown to only use NAD(H)(26), *pdxB,* which encodes the enzyme erythronate-4-phosphate dehydrogenase involved in *de novo* pyridoxal 5'-phosphate biosynthesis, has received less attention. Analysis of the protein structure of erythronate-4-phosphate dehydrogenase bound to NAD(H) in *Salmonella Enterica* indicates that an aspartic acid residue residue (Asp146) coordinates the 2′ and 3′ hydroxyl groups in the ribosyl portion of NAD(H), conferring selectively for NAD(H) over NADP(H) (Fig. 2a)(27–29). We reasoned that if erythronate-4-phosphate dehydrogenase was constrained to use NAD(H) in all environments, then other orthologues from diverse species living in largely heterogeneous environments may display similar specificity. We thus obtained all *pdxB* orthologs from the KEGG database (orthogroup K03473, n=1087), and performed multiple sequence alignment on the coenzyme-binding Rossman fold (Methods). We found that the previously identified aspartic acid residue was conserved in all *pdxB* orthologs, consistent with predictions that *pdxB* is universally required to use NAD(H) like *leuB* (Fig. 2b).

We next sought to determine whether the observed degree of conservation of Asp146 in *pdxB* orthologs is higher than other NAD(H)-dependent oxidoreductases that are not predicted to be constrained to use NAD(H). We determined whether there was structural evidence in the Protein Data Bank (PDB) for each NAD(P)-dependent oxidoreductase in the *E. coli* (or a homolog with greater than 30% homology) bound to NAD(P), and annotated residues that confer selectivity to NAD(H) or NADP(H). We found 11 oxidoreductases with NAD(H)-bound structures in the PDB with clear evidence of hydrogen bonding between an aspartic acid or glutamic acid residue and the 2′ and 3′ hydroxyl groups in the ribosyl portion of NAD(H) (Supplementary Table 3). We performed multiple sequence alignment for orthogroups for each oxidoreductase, and quantified the conservation of the aspartic acid or glutamic acid residue across orthologs (Methods, Fig. 2c). While the universally constrained oxidoreductase encoding genes *leuB* and *pdxB* show extremely high conservation of the coordinating aspartic acid residue (94 and 100%, respectively), several other oxidoreductase not predicted to be constrained to using NAD(H) show comparably high conservation of the coordinating aspartic acid or glutamic acid residue

(Fig. 2c). More broadly, analysis of residues that confer selectivity for NAD(H) or NADP(H) showed that these residues were significantly less variable that the remaining portion of the protein sequences, and the degree of conservation is not associated with flux balance predictions (Wilcoxon sign rank test: $P < 10^{-5}$, Supplementary Fig. 3). Altogether, these results suggest that flux balance analysis is insufficient for predicting the vast degree to which oxidoreductases selectively use one coenzyme over another.

Altogether, the results from genome-scale simulations suggest that metabolism, in principle, can operate in a variety of environments when coenzyme specificity for individual oxidoreductases are altered. On the other hand, sequence conservation patterns indicate clear conserved preferences for specific coenzyme pairings for a number of enzymes. One possible explanation for this apparent discrepancy is that coenzyme specificity may be selectively advantageous without being essential to carry out metabolic functions. Prompted by this possibility, we reasoned that the existence of two distinct coenzymes, and the assignment of each to a subset of reactions, may be the outcome of selection processes that favored organisms with a more efficient two-coenzyme system relative to a primitive single-coenzyme system. If this was the case, we would expect to be able to uncover two distinct previously unexplored features of metabolism: the first is the feasibility of a metabolism that operates using a single redox currency coenzyme. It is possible, in principle, that metabolism today may still be able to run (albeit less efficiently) with a single coenzyme? The second feature is a quantifiable fitness advantage that a two-coenzyme system may have over a metabolic network with a single-coenzyme. In the following two sections we explore each of these two features.

Metabolic networks can feasibly operate using only a single mid-potential coenzyme

Motivated by the modeling results presented above suggesting that thermodynamic constraints induce a small number of enzymes to retain specificity for NAD(H), one can pose a more extreme question, and hypothesize that perhaps metabolism could in principle operate using only a single NAD-coenzyme system, rather than both NAD(H) and NADP(H). This question is

substantially more challenging to address, due to the computational complexity of performing genome-scale thermodynamic FBA calculations under many different conditions. To explore the possibility that metabolism could feasibly operate using a single coenzyme, we thus restricted our analysis to a reduced model of *E. coli* core metabolism (irJO1366)(30) , and modified it to use NAD(H) only by converting NADP(H)-dependent reactions to use NAD(H) instead and removing the transhydrogenase reaction (Supplementary Fig. 4a). We simulated the consequences of these modifications using thermodynamic flux balance analysis (TFBA, originally called thermodynamic metabolic flux analysis)(6, 31) which maximizes cellular growth rate, subject to mass balance and thermodynamic constraints. This modeling approach, in contrast to the method presented above, assumes coenzyme concentrations are not fixed, thus allowing for a wider exploration of $NAD(P)^+$/NAD(P)H ratios that could enable cellular growth. We first simulated growth using TFBA with glucose as the sole carbon source in aerobic conditions, and found that the maximum growth rate achievable was identical to that of the unmodified, two-coenzyme model. NADP(H)-dependent reactions, like NADP(H)-dependent glutamate dehydrogenase, remained thermodynamically feasible by altering co-substrate metabolite concentrations accordingly (Supplementary Fig. 4b). We next constrained the growth rate to the maximum achievable growth rate for each model, and maximized/minimized the $NAD(P)^+$/NAD(P)H ratio for each model, and found that the single coenzyme model was restricted to a similar lower bound as NAD^+/NADH in the unmodified, two-coenzyme model, suggesting that NAD(H)-coupled reactions require a minimum free energy to drive NAD^+ coupled oxidations (Supplementary Fig. 4c). We next extended our analysis to model the growth rates for a variety of carbon and nitrogen sources, and found that generically, single coenzyme models are capable of producing biomass at nearly the same rate as the two-coenzyme model (Supplementary Fig. 4d). Taken together, these results motivate the hypothesis that NAD(P)H-redundancy might not be an essential features of metabolism, suggesting that ancient living systems may have originally operated using a single redox currency coenzyme and that NAD(P)H-redundancy might have emerged as a consequence of adaptive evolution. However, it is unclear what environments and selective advantages may have driven the emergence of multi-coenzyme systems.

Enzyme cost minimization drives separation of coenzyme preference globally

Thus far, the fitness consequences of modifying oxidoreductase specificity for different coenzymes has been assessed from the perspective of maximum specific growth rate using stoichiometric models of metabolism. These models suggest that thermodynamics may drive specificity for a few individual enzymes, giving rise to "local"' constraints for enzyme specificity that are also recapitulated in protein sequence-level variation. However, these results do not explain the widespread specificity for NAD(H) or NADP(H) by oxidoreductases.

Beyond inducing local constraints in single reactions, thermodynamics may also induce "global" fitness consequences through the control of the total protein demand to carry biosynthetic flux. This is due to the fact that strong thermodynamic driving forces in enzyme-catalyzed reactions reduce the minimum protein abundance necessary to maintain a fixed reaction flux (32–35). We hypothesized that coenzyme redundancy might allow the cell to lower the total protein required to maintain intracellular reaction fluxes. Building off prior work modeling the protein cost as a function of metabolic flux (33), we constructed a model of enzyme cost as a function of coenzyme specificity, and aimed to study how the number of coenzymes affects the amount of protein enzymes needed to run metabolism (Fig. 3a, Methods)

Given the uncertainty about many of the parameters involved in the model, we generated our predictions based on an extensive random sampling of key parameters, including the reaction network topology, the effective equilibrium constant for reaction r ($K^{\dagger}_{r}$), the maximal turnover rates for reaction r (κ_r), and the flux for reaction r (v_r). In this formulation, the effective equilibrium constant aggregates the effects of the free energy change at standard molar conditions, and the concentration differences between products and substrates that are not coenzymes, capturing the coenzyme-independent forces driving reaction r (see Supplementary Text). Our model also includes the ratio of charged to uncharged species for various coenzyme

pairs ($\Gamma_\alpha = \frac{[NADH]}{[NAD]}$, $\Gamma_\beta = \frac{[NADPH]}{[NADP]}$) and the fraction of flux for reaction *r* through the NAD(H) coenzyme pool ($v_{r\alpha}$)and NADP(H) pool ($v_{r\beta}$).

We next minimized the total protein cost of coenzyme-coupled reactions by varying the coenzyme concentration ratios (Γ_α,Γ_β) and the specific reaction fluxes ($v_{r\alpha}$, $v_{r\beta}$) using a grid search, fixing the coenzyme concentration ratios and solving the ensuing linear program to find the optimal flux distribution (see Supplementary text). We then obtained the coenzyme concentration ratios that minimized the total protein abundance (Fig. 3b).

In simulations where a single coenzyme can feasibly drive all flux (Fig. 3a, Supplementary Text), we found that flux partitioning through two coenzyme pools led to a decrease in the minimum protein abundance required to maintain metabolic flux, relative to the single coenzyme scenario (Fig. 3b). Notably, at the optimal point, each reaction flux was driven through primarily one coenzyme, indicating that enzyme specificity for a single coenzyme could emerge from the cellular-level objective of optimizing proteome allocation (Supplementary Fig. 5a). The emergence of enzyme specificity for one of the two coenzymes was driven by the combination of both kinetic and thermodynamic factors (Supplementary Fig. 5b-c), where reactions that required less protein to maintain the given flux demand were used to re-balance coenzyme pools, operating against the coenzyme concentration gradient (Supplementary Fig. 5b-c).

To see if these results were generic for different realizations of parameters, we computed optimal flux distributions and coenzyme ratios for 100 different random instances, and found that the two-coenzyme model always could achieve lower enzyme cost than the one coenzyme model (Fig. 3c). For each simulation, we plotted the coenzyme ratios found to minimize the total protein abundance, and we found that all simulations resulted in optimal ratio sets with opposing thermodynamic drive, where one coenzyme pool was primarily in the charged state, and the other was primarily in the uncharged state (Fig. 3d). This is strikingly consistent with the observation that *in vivo* measured ratios of oxidized to reduced NAD(H) and NADP(H)

coenzymes are greater than and less than one, respectively (5, 36). Interestingly, our model also predicts that the largest reduction in enzyme cost comes when the mean effective equilibrium constants are closer to unity (Fig. 3e), suggesting that two-coenzyme systems may be particularly beneficial if the coupled reactions are held closer to equilibrium, as typically found in anaerobic biochemical reactions.

Importantly, this model requires no detailed knowledge of the underlying organization of metabolic flux, or groups transferred via the coenzyme (e.g., electrons, methyl groups, etc.), and can in principle be used to analyze generic trends that may explain the emergence of multi-coenzyme systems versus single coenzyme systems in biochemistry.

NAD(P)-dependent oxidoreductase coenzyme-specificity is associated with reaction thermodynamics

From the models presented above, we found that enzyme preference for specific coenzymes is strongly associated with thermodynamic drive, where enzymes catalyzing reactions near equilibrium are predicted to be more specific for one coenzyme over another. To test this prediction, we computed free energies for all NAD(P)-coupled oxidoreductases in the KEGG database using eQuilibrator(37, 38), and computed binding preferences of protein sequences for NAD(H) and NADP(H) using an artificial neural network (ANN)(39), which estimates the likelihood of FAD(H2), NAD(H) and NADP(H) binding in protein sequences containing Rossman folds. We noticed that the distribution of free energies for NAD(P)-coupled oxidoreductases were bimodal, where oxidoreductases using oxygen as a co-substrate were driven far from equilibrium (421.8 +/- 155 kJ/mol) compared non-oxygen utilizing oxidoreductases (12.4 +/- 48.5 kJ/mol) (Fig. 4a-b), providing a simple way to categorize oxidoreductases by thermodynamic potential. We computed binding preferences for NAD(H) and NADP(H) (Fig. 4c), and computed the proportion of Rossman folds that were single coenzyme binders (e.g., NAD(H) or NADP(H)) or multi-coenzyme binders (e.g., NAD(H) and NADP(H)). Consistent with predictions from the theory, we found that 61% (2139/3497) of

folds from oxygen-coupled NAD(P)-dependent oxidoreductases are expected to be multi-coenzyme binders, compared to just 23% (28992/126718) from non-oxygen coupled NAD(P)dependent oxidoreductases (Fisher's exact test: $P < 10^{-5}$).

Ribosyl phosphorylation may be a generic strategy to partition coenzyme pools

The simple phenomenological model presented above demonstrates that coenzyme redundancy (i.e. multiple coenzymes for the same group transfer reaction) may be a generic mechanism to reduce intracellular protein cost, potentially explaining the maintenance of both NAD(H) and NADP(H) as intracellular redox coenzymes. However, this simple model is agnostic to the chemical details of the group or electron transfer (apart from thermodynamic potentials), suggesting that protein cost can be minimized by simply adding additional coenzymes to the intracellular repertoire. This begs the question of why several coenzymes appear to exist as only one variant in the cell.

One interesting example is Coenzyme A, a molecule involved in the redistribution of intracellular acyl groups. We computed free energies for various group transfer reactions (Fig. 5a), and found that the distribution of free energies for acyl transfer reactions were similar to NAD(P)-coupled reactions, suggesting multiple coenzyme systems may be also advantageous for Coenzyme A coupled acetyl-transfer reactions. Analysis of the biosynthesis of Coenzyme A revealed a striking similarity to the biosynthesis of NAD(P), where a kinase acts on the ribosyl moiety of the coenzyme structure (Fig. 5b-c). For the case of NAD(P) biosynthesis, NAD kinase phosphorylates NAD at the 3′ hydroxyl of the ribosyl moiety to create NADP. For the case of Coenzyme A, dephospho-CoA kinase phosphorylates the 2′ hydroxyl of the ribosyl moiety of dephospho-CoA to produce CoA. The strikingly similar reaction motif may indicate that the dephospho-CoA is an active and alternative coenzyme capable of performing similar functions as CoA, analogous to NAD^+ and $NADP^+$. Interestingly, recent mass spectrometry experiments have shown that dephospho-CoA can be acylated in rat liver(40). We thus propose that

dephospho-CoA may be a redundant coenzyme for acyl-transfer reactions rather than simply serve as an intermediate during CoA biosynthesis.

Discussion

The rise and evolution of cofactor coupling in redox biochemistry constitutes a fascinating puzzle central to the emergence and complexification of life itself. In order to understand the stoichiometric and thermodynamic constraints and the selection pressure that shaped specific enzyme-cofactor couplings observed today, one would ideally shuffle all of these couplings (e.g., by reassigning cofactors to different enzymes), and determine fitness changes under many different environments. While such an endeavor would be very challenging experimentally, it can be pursued efficiently using genome scale stoichiometric models of metabolism. In particular, we used flux balance modeling to survey dozens of *in silico* mutants of oxidoreductase enzymes in thousands of environmental conditions, explicitly modeling the consequences of changing an enzyme's coenzyme preference on growth rate. Our analysis suggests that few enzymes are predicted to be universally constrained to use NAD(H) or NADP(H), and that several dependencies are only observable in specific environments. Future experimental work could test these predictions by altering coenzyme specificity for some of these oxidoreductases and measuring growth rate in various media conditions.

Our results suggest that thermodynamic forces, rather than stoichiometric balancing, play a critical role in shaping the preferences of oxidoreductases for specific coenzymes. Structural and bioinformatics analyses of *pdxB* suggests that these thermodynamic constraints create strong selective pressures shaping protein sequence evolution, evidenced by our observation that all *pdxB* orthologs contain a conserved residue that specifically coordinates NAD(H), and potentially sterically prevents the binding of NADP(H). Similar findings have been demonstrated with isopropylmalate dehydrogenase (26), although prior experimental work suggested that kinetic factors shape the coenzyme specificity. Interestingly, our results suggest

that thermodynamic forces can alone predict isopropylmalate dehydrogenase's dependency on NAD(H), without necessarily invoking kinetic constraints.

The flux modeling results presented in this study only consider stoichiometric and thermodynamic factors shaping maximum organismal growth rate. However, many other constraints or metabolic demands may exist that are important for constraining a particular oxidoreductase coenzyme preference. For example, our flux models may not accurately capture metabolic demands induced by oxidative stress, for which NADPH is the primary electron source for cellular detoxification and antioxidants(2). Specificity may be shaped through mechanisms not directly linked to stoichiometric balancing or thermodynamics, including epistatic mechanisms, where the sequence variation in the coenzyme binding site indirectly influences co-substrate specificity and kinetics (41) or thermostability(42). Nonetheless, our results suggest that network stoichiometry and thermodynamics are important factors that can constrain the oxidoreductase specificity for NAD(H) or NADP(H).

The model presented in Fig. 3 suggests that the emergence of widespread enzyme specificity towards one of many coenzymes could be induced by a selective pressure to minimize the total abundance of enzymes in the cell. Importantly, our model is generic and does not rely on chemical details specific to nicotinamide coenzymes or electron transfer, and could potentially explain the ubiquity of multi-coenzyme systems (e.g. ferredoxins, thioredoxins, glutaredoxins, flavins, quinones, hemes, pterins, nucleotide phosphates) in cellular metabolism. Notably, our results suggest that enzyme specificity towards one of the two coenzymes may be a generic strategy to reduce the total protein cost of oxidoreductase enzymes, and that coenzyme choice is shaped by both kinetic and thermodynamic factors.

Our model can be used for the targeted search for redundant coenzymes, and could in principle explain the occurrence of acyl-dephospho-CoAs(40). Specifically, our model predicts there are enzymes that selectively bind (acyl)-dephospho-CoA rather than (acyl)-CoA, and that the

steady-state gradient of acylated to non-acylated dephospho-CoA should be in the opposing direction of acylated to non-acylated CoA.

Acknowledgements

We thank Arren Bar-Even, Adrian Jinich, Igor Libourel and Pankaj Mehta for helpful discussions. We acknowledge the support provided by the Directorates for Biological Sciences (BIO) and Geosciences (GEO) at the NSF and NASA under Agreements No. 80NSSC17K0295, 80NSSC17K0296 and 1724150 issued through the Astrobiology Program of the Science Mission Directorate. J.E.G. is supported by the Gordon and Betty Moore Foundation as Physics of Living Systems Fellows through grant number GBMF4513.

Contributions

J.E.G.. and D.S. designed the research. J.E.G. wrote code and ran simulations. J.E.G. and A.F. performed analysis. A.B.G. contributed to protein cost modeling. J.E.G., A.F., and D.S. wrote the manuscript. All authors read and approved the final manuscript.

Competing Financial Interests

The authors declare no competing financial interests

Corresponding Authors

Correspondence to: goldford@mit.edu, dsegre@bu.edu

Figures

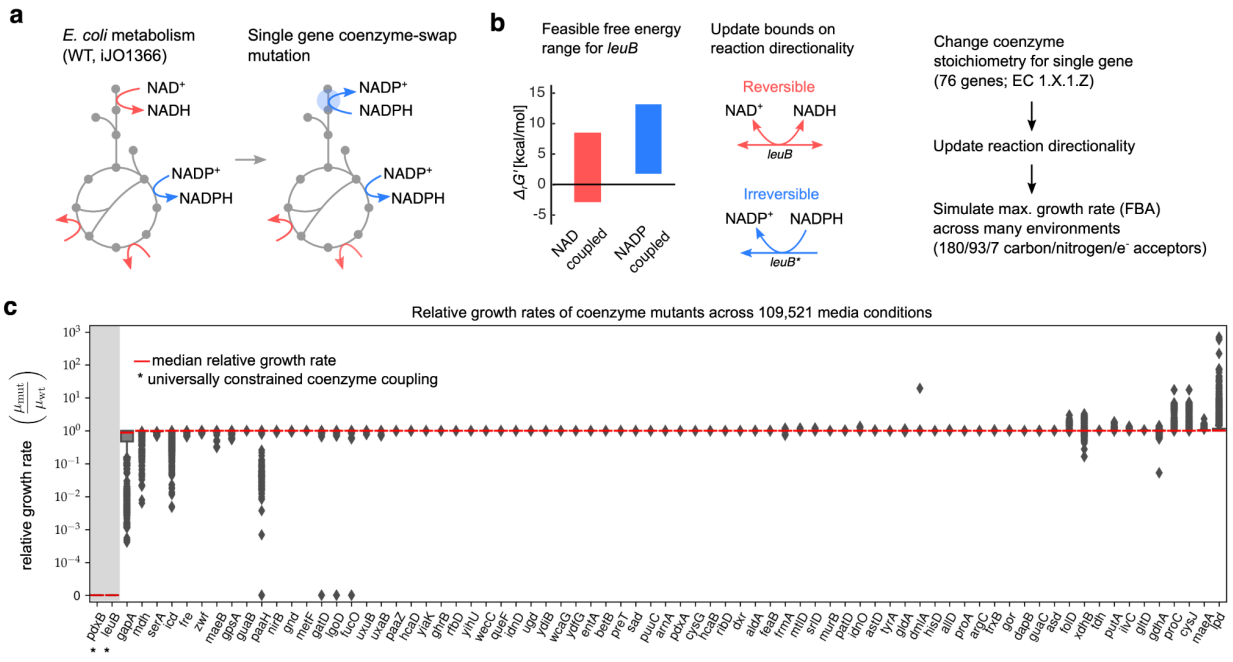


Figure 1: *In silico* analysis of coenzyme rewiring within the *E. coli* metabolic network highlights a small number of constrained reactions. We determined the consequences of "rewiring" the *E. coli* metabolic network by changing the coenzyme specificity of each oxidoreductase reaction separately. In the *E. coli* metabolic model iJO1366, there are 76 genes that encode oxidoreductases coupled to either NAD(H) or NADP(H). For each oxidoreductase annotated to catalyze a reaction with a single coenzyme, we swapped the coenzyme preference with the alternative coenzyme (**a**). (**b**) We computed feasible free energy ranges for each coenzyme-coupled reaction using bounds on intracellular metabolite concentrations, as well as the following measured *in vivo* concentrations of nicotinamide coenzymes: [NAD] = 2.6×10^{-3}, [NADH] = 8.3×10^{-5}, [NADP] = 2.1×10^{-6}, and [NADPH] = 1.2×10^{-4}.

We then set reaction bounds such that a reaction was irreversible if the free energy range was restricted to solely positive (irreversible in the backward direction) or negative values (irreversible in the forward direction) (see Methods). (**c**) We computed maximum growth rates (μ_{mut}) for each rewired model across 117,180 growth media spanning different combinations of 180 carbon sources, 93 nitrogen sources and 6 electron acceptors (oxygen, nitrite, nitrate, TMAO, DMSO, fumarate) as well as fermentative growth using flux balance analysis (FBA) enabling us to compute the relative fitness in each environment (y-axis, μ_{mut}/μ_{wt}). For each gene encoding an oxidoreductase, we modeled mutations that change the coenzyme specificity for that oxidoreductase, and computed the relative growth rate across all environments that supported wild-type growth (n=109,521), and plotted the distribution of relative growth rates per gene (x-axis). Only two oxidoreductases are constrained to the endogenous coupling in all environments (denoted by *, grey shaded region): the NAD(H)-coupled isopropylmalate dehydrogenase *leuB* and erythronate 4-phosphate dehydrogenase *pdxB*.

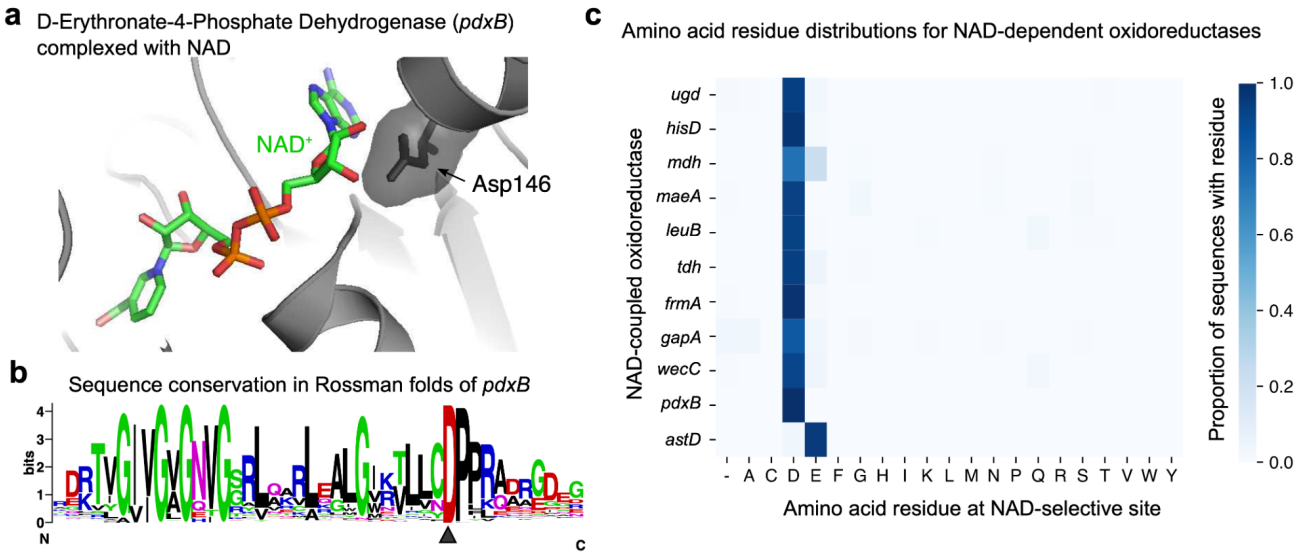


Figure 2: Flux balance models underpredict the widespread conservation of NAD-stabilizing Asp/Glu residues in NAD-coupled oxidoreductases. The crystal structure of *S. enterica* D-Erythronate-4-Phosphate Dehydrogenase (encoded by *pdxB*) complexed with NAD^+ (PDB 3OET) is shown in **a** (27), demonstrating that Asp146 interacts with the 2′ and 3′ hydroxyl groups on the ribose moiety of NAD(H), preventing NADP(H) binding (29, 43). **b**. Orthologs of *pdxB* were retrieved from the KEGG database (K03473). The Rossman fold was identified using HMMR 3.1(44) to generate a sequence logo for the NAD(H) binding domain of these enzymes (Methods). The triangle denotes the 100% conservation of the aspartic acid residue responsible for discriminating between NAD(H) and NADP(H). **c.** Conservation of residues coordinating hydroxyl groups in the ribosyl moiety of NAD(H) across 11 NAD(H)-coupled enzymes in the *E. coli* metabolic network. We performed multiple sequence alignment for sequences within each orthogroup, and computed the proportion of sequences that contained one of twenty amino acids at the site conferring selectivity for NAD(H). The heatmap shows the proportion of sequences containing the specific amino acid residue at the NAD-selective site (*x*-axis) for several NAD(H)-binding orthgroups (*y*-axis).

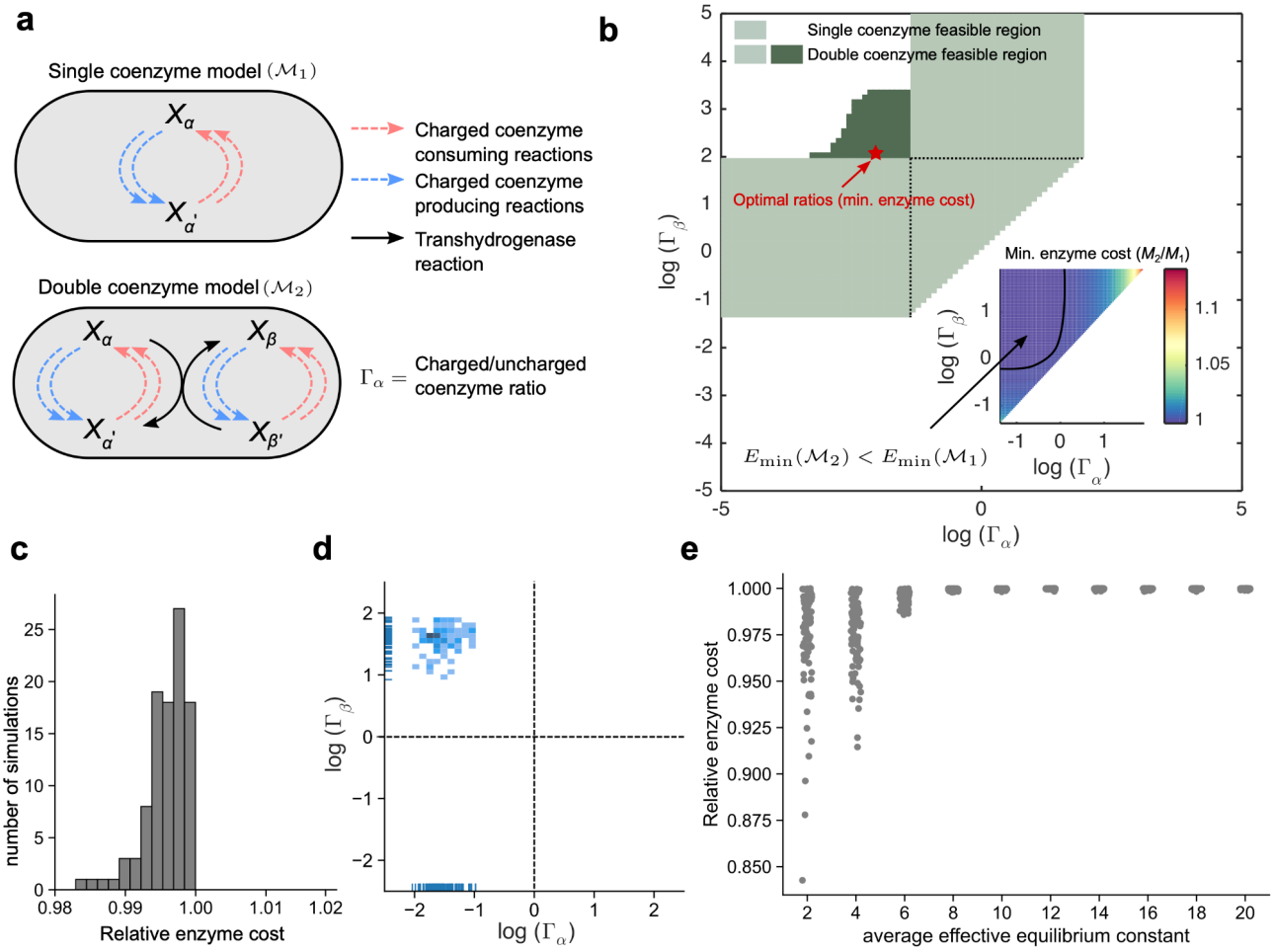


Figure 3: A minimal model of enzyme cost predicts the emergence of coenzyme redundancy. a. Model of a reaction network where multiple reactions (dashed arrows) produce and consume a charged coenzyme, where reaction flux is feasible with a single coenzyme (top), compared to a model where reaction flux is partitioned between two coenzyme pools connected by a transhydrogenase reaction (bottom). **b**. The feasible regions for coenzyme ratios (*x* and *y* axis, $\mathbf{\Gamma}_\alpha$ and $\mathbf{\Gamma}_\beta$) for a representative flux distribution, comparing a single-coenzyme model to a two-coenzyme model. For a simulation where we randomly sampled reaction fluxes, thermodynamic and kinetic parameters (using an mean effective equilibrium constant $\mu_K = 6$, see Methods), the feasible space for a single coenzyme (light green region) is plotted for various concentration ratios (*x* and *y* axis) of the coenzyme, compared to a two-coenzyme system (light green and dark green region). For a two-coenzyme system, we used a grid search to compute the enzyme cost at various coenzyme ratios (inset), and searched for coenzyme ratios that minimized

the total protein cost (red star). **c.** We simulated 100 random realizations, and plotted the distribution of the minimal enzyme cost relative to the cost using a single coenzyme (*x*-axis), and found that all simulations were lower than unity. **d**. The coenzyme ratio for coenzymeα(Γ_α, *x*-axis) versus coenzyme β(Γ_β, *y*-axis) that minimizes total protein abundance were plotted for each realization as a 2-D histogram. **e**. We varied the mean effective equilibrium constant (μ_K, *x*-axis), and plotted the fractional reduction in minimal protein required (*y*-axis) relative to a single coenzyme model (see Supplementary text for definitions and derivations). Note that as the coenzyme-independent thermodynamic drive increases, the benefit of multiple coenzyme pools decreases.

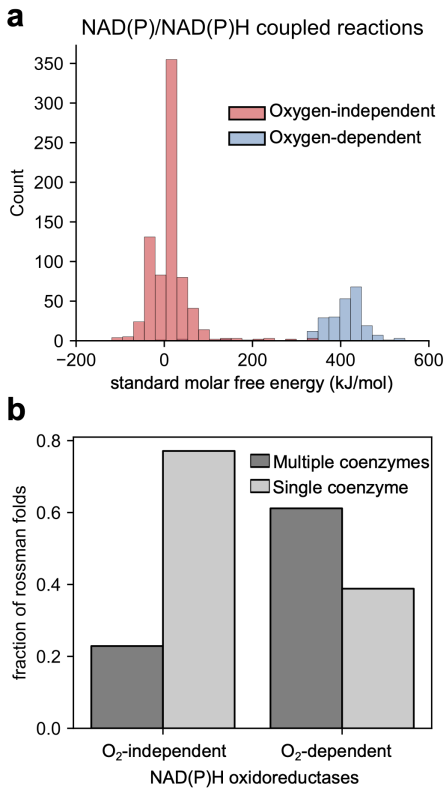


Figure 4: Oxygen-coupled NAD(P)-oxidoreductases are predicted to be more coenzyme promiscuous. **a.** We computed standard molar free energies for all NAD(P)-coupled reactions in the KEGG database using eQuilibrator (38) , and plotted the distribution of free energies for (a) NAD(P)-coupled reactions in the NAD(P)H producing direction (n=990 reactions). Distributions of free energies from oxidoreductase reactions that use molecular oxygen as a co-substrate are plotted in blue, while distributions of free energies for other NAD(P)-coupled reactions are plotted in red. **b.** The proportion of Rossman folds that were predicted using an artificial neural network model (39) to bind a single coenzyme (light grey bars, e.g., NAD(H) or NADP(H)) or multiple coenzymes (dark grey bars, e.g., NAD(H) and NADP(H)) are plotted for oxygen-dependent or independent NAD(H)/NADP(H) oxidoreductases (x-axis). We found 61% of the folds from oxygen-dependent NAD(H)/NADP(H) oxidoreductases (2139/3497) were multi-coenzyme binders, compared to just 23% amongst oxygen-dependent NAD(H)/NADP(H) oxidoreductases (28992/126718) (Fisher's exact test: $P < 10^{-5}$).

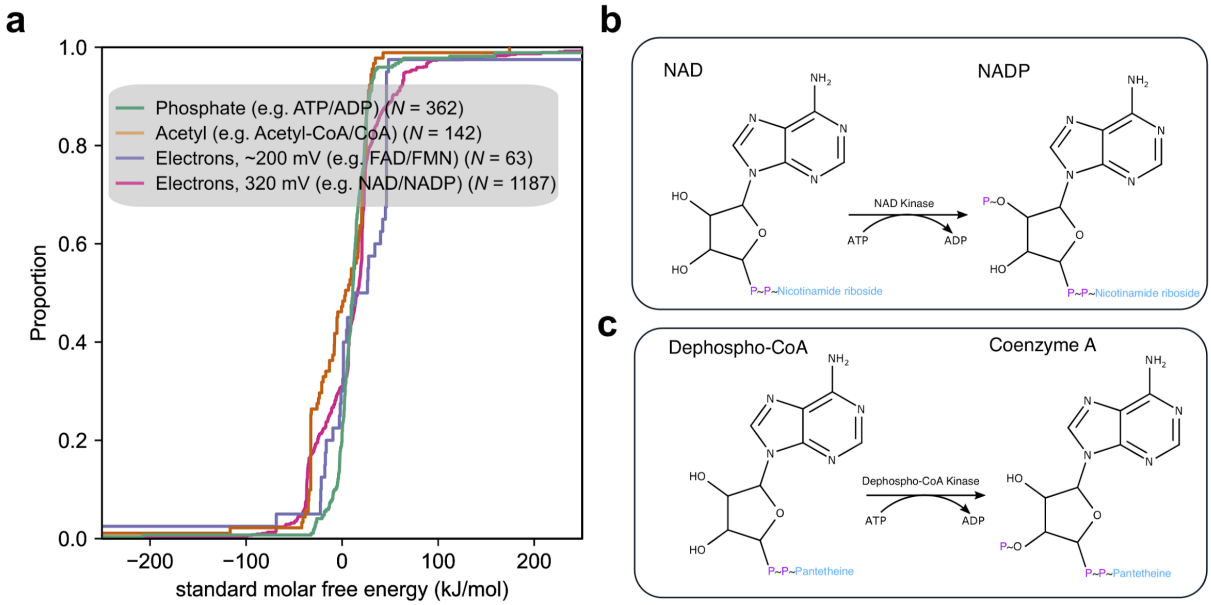


Figure 5: Comparison between NAD(P) and Coenzyme A biosynthesis reveals potential redundancy for acyl-CoA's. a. Free energies for group transfer reactions of phosphate groups, acetyl groups and electron transfer reactions in KEGG were computed using eQuilibrator (37, 38), and the empirical cumulative distributions across reactions were plotted. Free energies for all group and electron transfer reactions are similarly distributed near equilibrium, suggesting that Coenzyme A-coupled reactions may also depend on redundant coenzyme systems similar to NAD(H)/NADP(H) . Free energies for redox and group transfer reactions were computed for the following coenzyme pairs. Phosphate-coupled reactions: ADP/ATP, CDP/CTP, GDP/GTP; Acetyl-coupled reactions: Coenzyme A/Acetyl-CoA; Electrons (220 mV): FAD/FADH2, FMN/FMNH2, Glutathione disulfide/Glutathione; Electrons (320 mV): NAD/NADH, NADP/NADPH. **b**. The last biosynthetic step of $NADP^+$ synthesis is the phosphorylation of the 3′ hydroxyl group on the ribose moiety of NAD^+, compared to the biosynthesis of Coenzyme A (**c**), where the 2′ hydroxyl of the ribose moiety is phosphorylated by dephospho-CoA kinase (EC 2.7.1.24)

References

1. D. L. Nelson, M. M. Cox, *Lehninger Principles of Biochemistry* (W. H. Freeman; 4th Edition, 2005).

2. L. Agledal, M. Niere, M. Ziegler, The phosphate makes a difference: cellular functions of NADP. *Redox Rep.* **15**, 2–10 (2010).

3. J. H. Grose, L. Joss, S. F. Velick, J. R. Roth, Evidence that feedback inhibition of NAD kinase controls responses to oxidative stress. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7601–7606 (2006).

4. S. K. Spaans, R. A. Weusthuis, J. van der Oost, S. W. M. Kengen, NADPH-generating systems in bacteria and archaea. *Front. Microbiol.* **6**, 742 (2015).

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

6. C. S. Henry, L. J. Broadbelt, V. Hatzimanikatis, Thermodynamics-based metabolic flux analysis. *Biophys. J.* **92**, 1792–1805 (2007).

7. M. R. de Graef, S. Alexeeva, J. L. Snoep, M. J. Teixeira de Mattos, The steady-state internal redox state (NADH/NAD) reflects the external redox state and is correlated with catabolic adaptation in Escherichia coli. *J. Bacteriol.* **181**, 2351–2357 (1999).

8. Y. Zhou, *et al.*, Determining the extremes of the cellular NAD(H) level by using an Escherichia coli NAD(+)-auxotrophic mutant. *Appl. Environ. Microbiol.* **77**, 6133–6140 (2011).

9. A. e. B. Canelas, W. M. van Gulik, J. J. Heijnen, Determination of the cytosolic free NAD/NADH ratio in Saccharomyces cerevisiae under steady-state and highly dynamic conditions. *Biotechnol. Bioeng.* **100**, 734–743 (2008).

10. A. M. Chánique, L. P. Parra, Protein Engineering for Nicotinamide Coenzyme Specificity in Oxidoreductases: Attempts and Challenges. *Front. Microbiol.* **9**, 194 (2018).

11. S. P. Miller, M. Lunzer, A. M. Dean, Direct demonstration of an adaptive constraint. *Science* **314**, 458–461 (2006).

12. G. Zhu, G. B. Golding, A. M. Dean, The Selective Cause of an Ancient Adaptation. *Science* **307**, 1279–1282 (2005).

13. N. S. Scrutton, A. Berry, R. N. Perham, Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature* **343**, 38–43 (1990).

14. J. K. B. Cahn, *et al.*, A General Tool for Engineering the NAD/NADP Cofactor Preference of Oxidoreductases. *ACS Synth. Biol.* **6**, 326–333 (2017).

15. L. Calzadiaz-Ramirez, *et al.*, In Vivo Selection for Formate Dehydrogenases with High Efficiency and Specificity toward NADP. *ACS Catal.* **10**, 7512–7525 (2020).

16. M. Bouzon, *et al.*, Emergence of NADP+-reducing enzymes in Escherichia coli central metabolism

via adaptive evolution. *Cold Spring Harbor Laboratory*, 2021.02.08.430375 (2021).

17. R. U. Ibarra, J. S. Edwards, B. O. Palsson, Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420**, 186–189 (2002).

18. D. Heckmann, *et al.*, Predicting C4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell* **153**, 1579–1588 (2013).

19. S.-R. Hosseini, O. Martin, A. Wagner, Phenotypic innovation through recombination in genome-scale metabolic networks. *Proceedings of Royal Society B* **283**, 20161536 (2016).

20. J. E. Goldford, H. Hartman, R. Marsland 3rd, D. Segrè, Environmental boundary conditions for the origin of life converge to an organo-sulfur metabolism. *Nat Ecol Evol* **3**, 1715–1724 (2019).

21. J. E. Goldford, D. Segrè, Modern views of ancient metabolic networks. *Current Opinion in Systems Biology* (2018).

22. R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, U. Sauer, Multidimensional optimality of microbial metabolism. *Science* **336**, 601–604 (2012).

23. A. Jinich, *et al.*, Quantum chemistry reveals thermodynamic principles of redox biochemistry. *PLoS Comput. Biol.* **14**, e1006471 (2018).

24. J. D. Orth, I. Thiele, B. Ø. Palsson, What is flux balance analysis? *Nat. Biotechnol.* **28**, 245 (2010).

25. J. D. Orth, *et al.*, A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Mol. Syst. Biol.* **7**, 535 (2011).

26. M. Lunzer, S. P. Miller, R. Felsheim, A. M. Dean, The Biochemical Architecture of an Ancient Adaptive Landscape. *Science* **310**, 499–501 (2005).

27. J. Y. Ha, *et al.*, Crystal Structure of d-Erythronate-4-phosphate Dehydrogenase Complexed with NAD. *J. Mol. Biol.* **366**, 1294–1304 (2007).

28. Y. Kallberg, B. Persson, Prediction of coenzyme specificity in dehydrogenases/reductases. A hidden Markov model-based method and its application on complete genomes. *FEBS J.* **273**, 1177–1184 (2006).

29. P. Laurino, *et al.*, An Ancient Fingerprint Indicates the Common Ancestry of Rossmann-Fold Enzymes Utilizing Different Ribose-Based Cofactors. *PLoS Biol.* **14**, 1–23 (2016).

30. M. Ataman, D. F. Hernandez Gardiol, G. Fengos, V. Hatzimanikatis, redGEM: Systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models. *PLoS Comput. Biol.* **13**, e1005444 (2017).

31. P. Salvy, *et al.*, pyTFA and matTFA: a Python package and a Matlab toolbox for Thermodynamics-based Flux Analysis. *Bioinformatics* **35**, 167–169 (2019).

32. A. Flamholz, E. Noor, A. Bar-Even, W. Liebermeister, R. Milo, Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 10039–10044 (2013).

33. E. Noor, A. Flamholz, W. Liebermeister, A. Bar-Even, R. Milo, A note on the kinetics of enzyme action: a decomposition that highlights thermodynamic effects. *FEBS Lett.* **587**, 2772–2777 (2013).

34. E. Noor, *et al.*, Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput. Biol.* **10**, e1003483 (2014).

35. E. Noor, *et al.*, The Protein Cost of Metabolic Fluxes: Prediction from Enzymatic Rate Laws and Cost Minimization. *PLoS Comput. Biol.* **12**, e1005167 (2016).

36. J. O. Park, *et al.*, Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat. Chem. Biol.* **12**, 482–489 (2016).

37. A. Flamholz, E. Noor, A. Bar-Even, R. Milo, EQuilibrator - The biochemical thermodynamics calculator. *Nucleic Acids Res.* **40**, 770–775 (2012).

38. M. E. Beber, M. G. Gollub, D. Mozaffari, K. M. Shebek, E. Noor, eQuilibrator 3.0 -- a platform for the estimation of thermodynamic constants. *arXiv [q-bio.MN]* (2021).

39. H. M. Geertz-Hansen, N. Blom, A. M. Feist, S. Brunak, T. N. Petersen, Cofactory: Sequence-based prediction of cofactor specificity of Rossmann folds. *Proteins: Struct. Funct. Bioinf.* **82**, 1819–1828 (2014).

40. Q. Li, S. Zhang, J. M. Berthiaume, B. Simons, G.-F. Zhang, Novel approach in LC-MS/MS using MRM to generate a full profile of acyl-CoAs: discovery of acyl-dephospho-CoAs. *J. Lipid Res.* **55**, 592–602 (2014).

41. D. J. 1. Maddock D. J. Patrick W. M. & Gerth M. L. Substitutions at the cofactor phosphate-binding site of a clostridial alcohol dehydrogenase lead to unexpected changes in substrate specificity. Protein Eng. Des. Sel. 28 251–8 (2015).Maddock, W. M. Patrick, M. L. Gerth, Substitutions at the cofactor phosphate-binding site of a clostridial alcohol dehydrogenase lead to unexpected changes in substrate specificity. *Protein Eng. Des. Sel.* **28**, 251–258 (2015).

42. S. Watanabe, T. Kodaki, K. Makino, Complete reversal of coenzyme specificity of xylitol dehydrogenase and increase of thermostability by the introduction of structural zinc. *J. Biol. Chem.* **280**, 10340–10349 (2005).

43. A. Goncearenco, I. N. Berezovsky, Protein function from its emergence to diversity in contemporary proteins. *Phys. Biol.* **12**, 045002 (2015).

44. S. R. Eddy, Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).

45. Z. A. King, *et al.*, BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.*, gkv1049– (2015).

46. E. Noor, H. S. Haraldsdóttir, R. Milo, R. M. T. Fleming, Consistent estimation of Gibbs energy using component contributions. *PLoS Comput. Biol.* **9**, e1003098 (2013).

47. J. Schellenberger, *et al.*, Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* **6**, 1290–1307 (2011).

48. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**,

27–30 (2000).

Materials and Methods

Software availability

Code and data are available on the following github repository:

https://github.com/jgoldford/coenzymes

Metabolic modeling of oxidoreductase specificity

We computationally accessed the fitness consequences of rewiring the NAD(P)H specificity of individual metabolic genes in *E. coli* by altering stoichiometric and thermodynamic constraints in a genome-scale metabolic model (GEMM), and by simulating the growth rate under various media conditions. We used the GEMM iJO1366(25) downloaded from the BIGG database (45), with Gibbs Free Energies derived using the component contribution method (46), using the eQuilibraitor API at pH 7 and ionic strength 0.1 M (37, 38). For each oxidoreductase gene that mapped to a reaction in iJO1366, we altered the coenzyme preference by first changing the stoichiometric coefficient in that reaction to the alternative coenzyme. For genes encoding reactions that utilized both coenzymes, we altered specificity by forcing all reactions to use only NAD(H) or NADP(H). To model the thermodynamic effect of changing coenzymes specificity, we used the computed reaction free energies at standard molar conditions, and bounds on intracellular metabolite concentrations to estimate the maximum and minimum driving force obtainable for each reaction using the following equation:

$$\begin{aligned} \underset{\ln(x)}{\text{minimize/maximize}} \quad & z_r = \Delta G'^{o}_r + RT s_r^T \ln(x) \\ \text{subject to} \quad & \ln(10^{-6}) \leq \ln(x) \leq \ln(10^{-2}) \end{aligned}$$

where s_r is the vector of stoichiometric coefficients for reaction r. Note that we fixed the concentration of coenzymes based on previous reported measurements (5, 36), where [NAD] =

2.6×10^{-3}, [NADH] = 8.3×10^{-5}, [NADP] = 2.1×10^{-6}, and [NADPH] = 1.2×10^{-4}. We then allowed all other metabolite concentrations to vary when predicting the maximum and minimum driving forces for each reaction. After each *in silico* mutation that changed coenzyme preferences, we re-computed the minimum and maximum driving force, and used these values to set the upper and lower bounds of reactions, such that:

$$l'_r = \begin{cases} 0 & \text{if } \max(z_r) < 0 \\ -1000 & \text{otherwise} \end{cases}$$

$$u'_r = \begin{cases} 0 & \text{if } \min(z_r) > 0 \\ 1000 & \text{otherwise} \end{cases}$$

Where l'_r and u'_r are the updated lower and upper bounds for reaction *r*, respectively. We simulated the consequences of changing oxidoreductase enzyme specificity for NAD(P)H in *E. coli* by then simulating the maximum growth rate using flux balance analysis (24) under 117,180 conditions, consisting of various carbon sources, nitrogen sources and electron acceptors (Oxygen, Nitrite, Nitrate, TMAO, DMSO, fumarate) as well as fermentative growth, using the following linear program:

$$\begin{aligned} \underset{v}{\text{maximize}} \quad & v_{\text{biomass}} \\ \text{subject to} \quad & S'v = 0 \\ & l' < v < u' \end{aligned}$$

Where S', l', u' is the altered stoichiometric matrix, flux lower bound, and flux upper bound after updating stoichiometric and thermodynamic constraints.

For results presented in Fig. 1c-d, we restricted the analysis to the media sets (109,521 conditions) where growth of the unperturbed network was feasible. We modeled mutations for 76 genes encoding oxidoreductases in *E. coli.* This set of genes were determined using the following criteria: (a) the gene encoded an enzyme that could be classified using the following

enzyme commission code - E.C. 1.X.1.Z, where X was any number except 6, and Z= any number, and (b) there were no other genes coding for enzymes that catalyze the same reaction. All simulations were performed using CobraPy and Gurobi (version 9.0.0) optimizer (47). Thermodynamic flux balance analysis was performed using pyTMFA (31), with the reduced core *E. coli* metabolic model irJO1366 (30). We modified irJO1366 by substituting all NADP(H)-coupled reactions with NAD(H)-coupled reactions, and removed the NAD(P)H transhydrogenase reaction, and performed TMFA as previously described(31).

Oxidoreductase specificity from compiled structural data and sequences

We compiled a list of NAD(P)-dependent oxidoreductases in the *E. coli* genome, and manually searched for structural features that confer selectivity of NAD vs NADP. To this end, we identified orthogroups (KO) using the KEGG REST API, and developed a custom python script to identify NAD(P)-bound protein data bank (PDB) codes from KEGG genes within each orthogroup. For each orthogroup that contained at-least one NAD(P)-bound structure for the *E. coli* ortholog, or an ortholog with at-least 30% homology, we manually identified residues that clearly showed hydrogen bonding with the 2′ and 3′ hydroxyl group of the ribosyl moiety in NAD(H), or electrostatic interactions with the 3′ phosphate in NADP(H).

We first analyzed the conservation of the residues conferring specificity in orthologs of *pdxB* (K03473), the gene that codes for erythronate-4-phosphate dehydrogenase. We downloaded all orthogroups in the KEGG database (orthogroup K03473, n=1087), and performed multiple sequence alignment on the coenzyme-binding Rossman fold. We first identified a 42 amino acid subsequence in the *E. coli* genome using HMMR 3.1 as the Rossman fold in the *pdxB* gene (b2320) from *E. coli* MG1655, and used the biopython “pairwise2” module to perform local alignment between the reference Rossman fold and the ortholog sequence. Parameters were chosen to penalize gap opening much more than allowing extensions or for allowing non-matching characters (matched character: 1; unmatched character: 0; gap opening penalty: -10; gap extension penalty: -1). For each ortholog, we identified the subsequence with the

largest score and labeled this subsequence as the Rossman fold for each ortholog. The sequences were then aligned using MUSCLE v3.8.31 using default parameters.

To compare other genes to *pdxB*, we identified 11 genes encoding enzymes in the iJO1366 model with structural evidence supporting the coordination of a single aspartic acid or glutamic acid residue and the 2′ and 3′ hydroxyl group of the ribosyl moiety in NAD(H), which were *leuB* (K00052), *pdxB* (K03473), *gapA* (K00134) , *mdh* (K00024), *frmA* (K00121), *tdh* (K00060), *rfbD* (K00067), *wecC* (K02472), *astD* (K06447), *ugd* (K00012), *dapB* (K00215), *maeA* (K00027), *hisD* (K00013). For orthogroups with >1000 orthologs, we downsampled the sequences to 1000 before identifying running multiple sequence alignment. In Supplementary Fig. 3, we annotated residues that confer coenzyme selectivity for 57 NAD(P)-coupled oxidoreductases in iJO1366 based on structures available in the PDB.

In Fig. 4c-d, we computed coenzyme binding likelihoods using a hidden markov model (HMM) dependent algorithm called Cofactory(39). Briefly, Cofactory first identifies the Rossman folds in a protein sequence using HMMR 3.1, and estimates the likelihood each fold binds FAD(H2), NADH and NADPH. In Fig. 4c-d, we identified 779 orthogroups that either use oxygen as a co-substrate, and downloaded all sequences within each orthogroup from the KEGG REST API(48). Like in the previous case, we downsampled the sequences to 1000 if the orthogroup contained more than 1000 sequences before running Cofactory. After obtaining predictions of binding likelihoods to NAD(H), NADP(H) and FAD(H2), we filtered out Rossman folds that were either (a) unambiguously bound FAD(H2) or (b) was not predicted to bind any of the coenzymes. We then categorized the remaining Rossman folds as “multi-coenzyme binders" if the binding likelihood exceeded 0.5 for more than one coenzyme, and “single coenzyme binders“ if the binding likelihood exceeded 0.5 for either NAD(H) or NADP(H), but not both.

Phenomenological modeling of proteome cost for multi-coenzyme systems

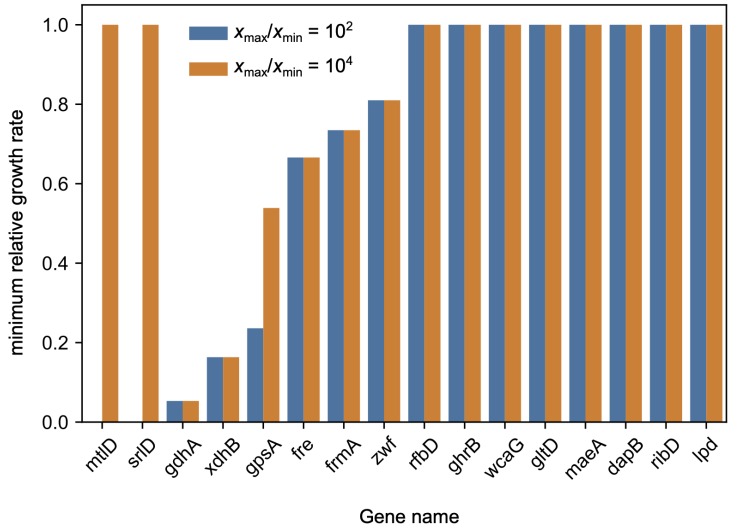
A detailed derivation of the phenomenological model used to analyze protein cost for multi-coenzyme systems in Fig. 3 is provided as a Supplementary Text. Here we present the

final model used in the main text. The following non-linear optimization computes the minimum protein cost of coenzyme-coupled enzymes, such that:

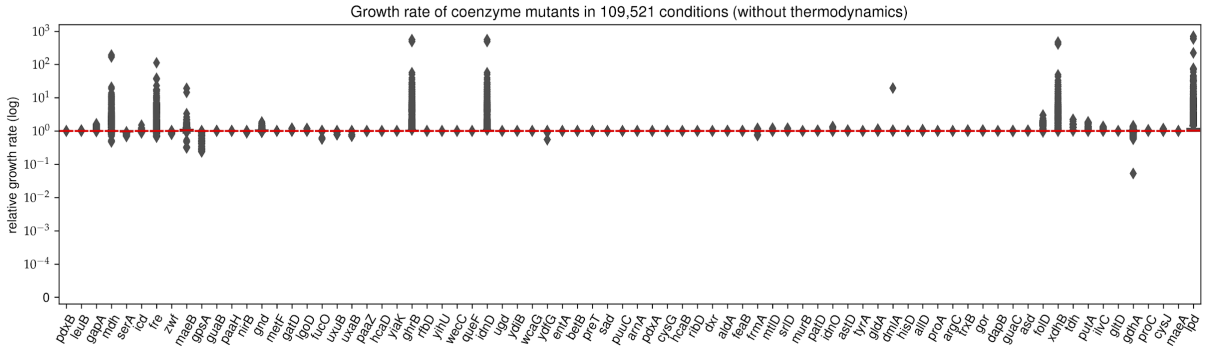
$$\begin{aligned}
\underset{v_{r\alpha}, v_e, \Gamma_\alpha, \Gamma_\beta, \ldots}{\text{minimize}} \quad & \sum_{r\alpha} \frac{v_{r\alpha}}{\kappa_r} \left(\frac{K_r^\dagger + \Gamma_\alpha^{s_{r\alpha}}}{K_r^\dagger - \Gamma_\alpha^{s_{r\alpha}}} \right) + \sum_e \frac{v_e}{\kappa_e} \left(\frac{\Gamma_\beta + \Gamma_\alpha}{\Gamma_\beta - \Gamma_\alpha} \right) \\
\text{subject to} \quad & \sum_\alpha v_{r\alpha} = v_r \quad \forall r \\
& \sum_r s_{r\alpha} v_{r\alpha} + \sum_e s_{e\alpha} v_e = 0 \quad \forall \alpha \\
& \Gamma_\alpha \leq \Gamma_\beta \quad \forall \alpha, \beta
\end{aligned}$$

where sampled parameters are the following: $K^\dagger_r$ is the effective equilibrium constant for reaction r, κ_r is the maximum turnover rate for reaction r, v_r is the net flux for reaction r, and $s_{r\alpha}$ is the stoichiometric coefficient for the charged coenzyme (negative if the charged-coenzyme is being consumed and is positive if being produced). Variables in our optimization approach are the following: Γ_α the ratio of charged to uncharged species for coenzyme pair α, $v_{r\alpha}$ is the fraction of flux from reaction r through the coenzyme pool α. In this model, we also include an additional set of reactions e, which are generalized exchange reactions that shuttle groups between coenzymes, much like NAD(P)H transhydrogenase enzymes. The objective function is the total enzyme cost of the coenzyme-coupled sub-network plus the cost of all generalized exchange reactions (see Supplementary Text for derivation). All enzyme cost simulations were performed in MATLAB R2021a, using the Cobra toolbox and GLPK solver for linear programs.

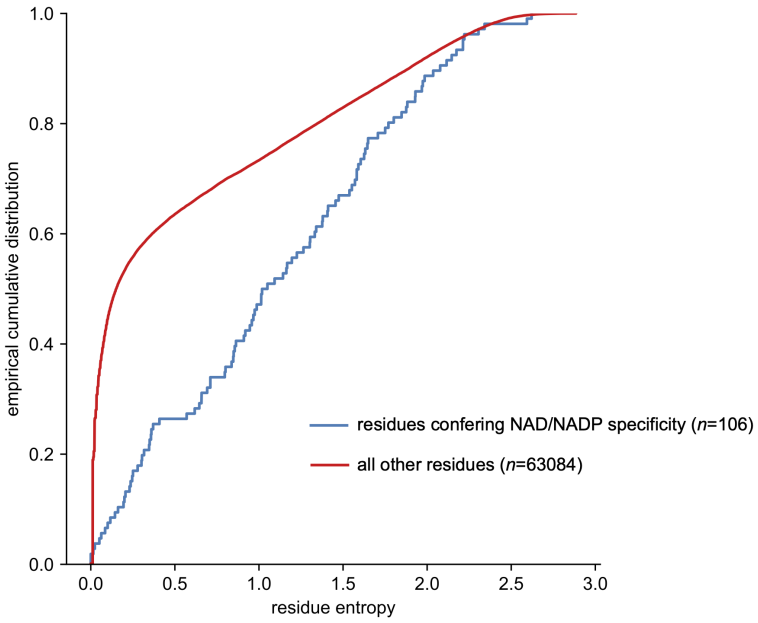
Supplementary Figures



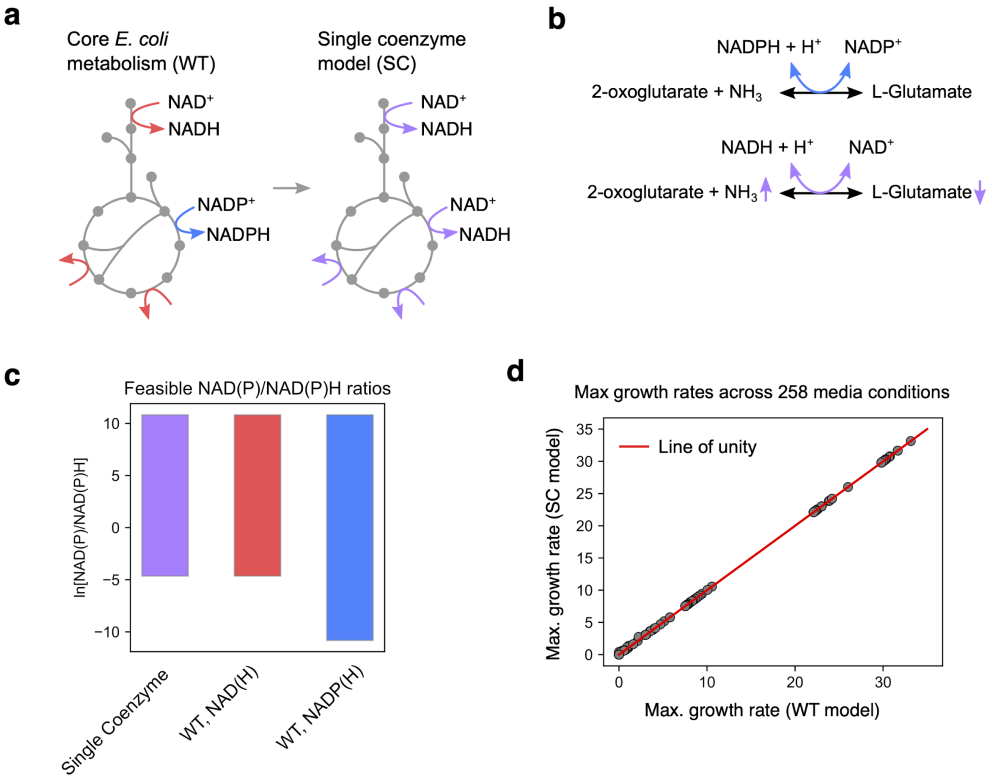
Supplementary Figure 1: Metabolite concentration ranges influence growth rates of coenzyme-swapping mutations in *E. coli*. We altered the modeling procedure presented in the Figure 1 of the main text to reduce the feasible range of metabolite concentrations, such that that maximum metabolite to minimum metabolite concentration ratio (x_{max}/x_{min}), was reduced from four orders of magnitude to two orders of magnitude. For each coenzyme mutation (x-axis), we computed maximum growth rates across all 109,521 media conditions, and plotted the minimum relative growth rate (μ_{mut} / μ_{wt}) across all conditions (y-axis).



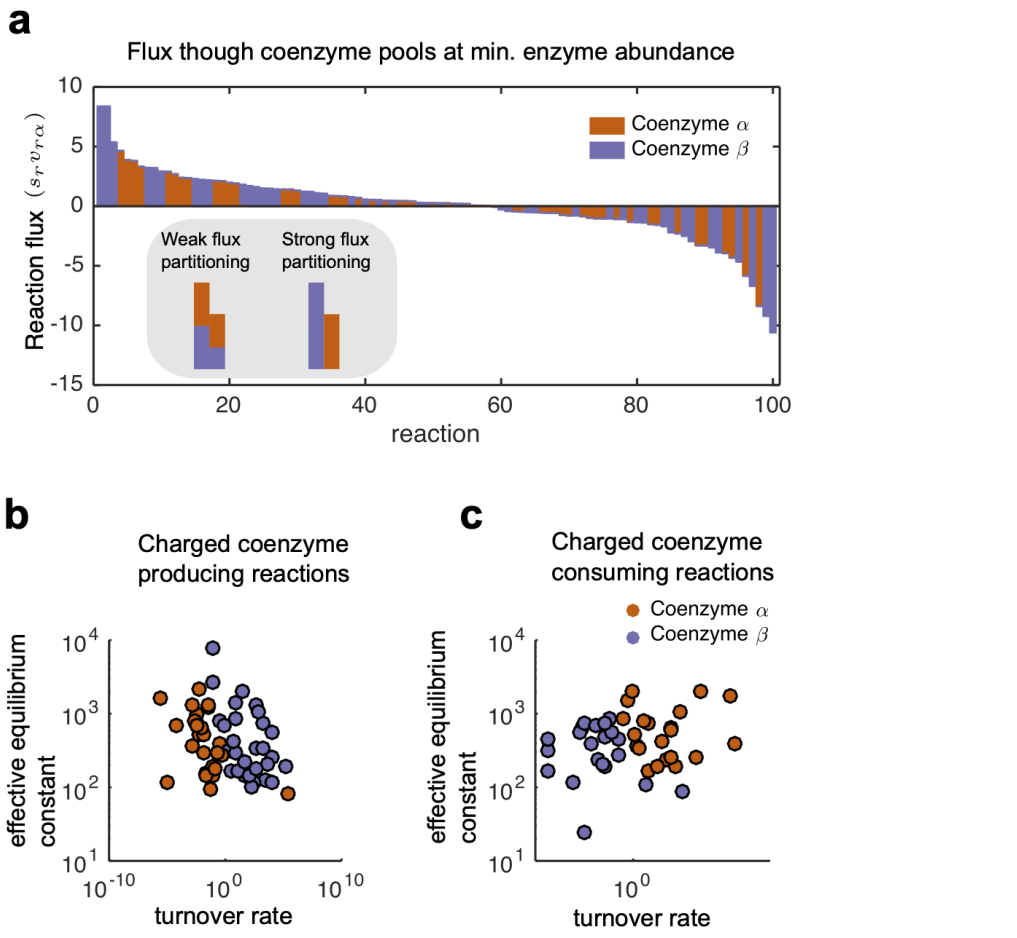
Supplementary Figure 2: Relative growth rates of various coenzyme mutations without thermodynamic constraints. We repeated the procedure described in the main text (Fig. 1c), but without updating thermodynamic constraints using knowledge of NAD(P)/NAD(P)H ratios. In this model, the only constraints that change are stoichiometric constraints, rather than reaction directionality constraints. Each mutant is plotted on the *x*-axis and the relative growth rate (μ_{mut} / μ_{wt}), is plotted on the *y*-axis. Notably, all essential coenzyme dependencies observed in Fig. 1c are no longer observed.



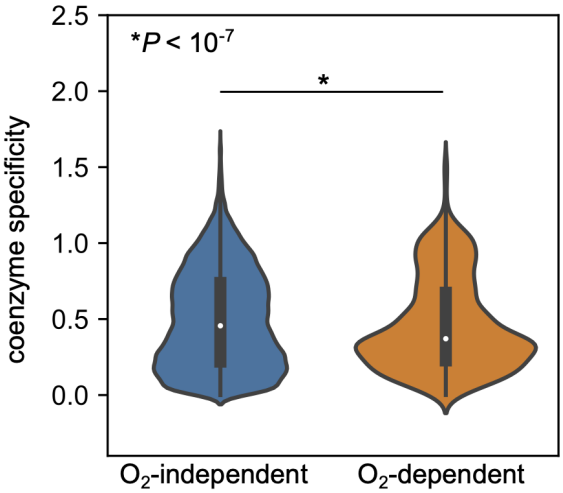
Supplementary Figure 3: Residues controlling selectivity for NAD(H) or NADP(H) are more conserved than other residues in oxidoreductase sequences. We identified all structural residues that form contacts with either the 2′ and 3′ hydroxyl group in the ribosyl portion of NAD(H), or the 2′ phosphate in NADP(H) in 57 KEGG Orthogroups (KOs) that mapped uniquely to reactions in iJO1366 (Supplementary Table 4). We performed multiple sequence alignment using MUSCLE, and computed the shannon entropy, H_j, at each residue position j, and classified a residue as conferring selectivity to NAD(H) vs. NADP(H), where $H_j = \sum_i x_{ij} log(x_{ij})$, and x_{ij} is the proportion of sequences with amino acids residue i in position j. We plotted the empirical cumulative distribution across all residues sites, where the blue line denotes residue sites that confer NAD(H)/NADP(H) selectivity, and the red line denotes all other residue sites in the protein sequence.



Supplementary Figure 4: Modeling the consequences of replacing all NADP(H) coupled reactions with NAD(H) in *E. coli* core metabolism. **(a)** We modified the reduced genome-scale metabolic model irJO1366 by replacing all NADP(H)-coupled reactions with NAD(H), and removed the transhydrogenase reaction and simulated growth in aerobic conditions with glucose as the sole carbon source. (**b**) Example of a reaction, glutamate dehydrogenase, which is natively originally coupled to NADP(H), that maintained thermodynamic feasibility by changing co-substrate concentrations (e.g., by decreasing L-glutamate and increasing 2-oxoglutarate and ammonia). **(c)** Feasible ranges of $NADP)^+$/NAD(P)H ratios (*y*-axis) for the single coenzyme model, and the wildtype model at maximum growth rate. (**d**) A scatterplot that shows the maximum growth rates of the wild-type (WT, *x*-axis) vs. single coenzyme model (SC, *y*-axis) growth rate on a variety of carbon and nitrogen sources, in both anaerobic and aerobic environments.



Supplementary Figure 5. Enzyme specificity emerges from proteome expression minimization. a. At the point of minimum total enzyme expression, we plotted the flux (*y*-axis) for each reaction (*x*-axis) as a stacked bar plot, where reaction flux through coenzyme α and β are shown in orange and purple, respectively. **b-c** For each reaction, we plotted the sampled turnover rate (*x*-axis) vs. the sampled effective equilibrium constant (*y*-axis), and colored each reaction by the preferred coenzyme shown in Fig. 3b in the main text. For the charged coenzyme producing reactions (**b**), enzymes with stronger driving forces and faster kinetics were preferentially used coenzyme β, where these reactions were responsible for operating against a concentration gradient.



Supplementary Figure 6. Coenzyme specificity is higher in enzymes that catalyze oxygen-independent reactions vs. oxygen-dependent reactions. We compute binding likelihoods for NAD(H) (L_{NAD}) and NADP(H) (L_{NADP}) for n = 130,215 Rossman folds that were predicted to bind at-least one coenzyme and were not predicted to only bind FAD(H2). Binding likelihoods were converted to specificities using the following formula: coenzyme specificity = $|log(L_{NAD}/L_{NADP})|$. We then plotted distributions of specificities for oxygen-independent (n = 126718) and oxygen-independent (n = 3497) Rossman folds, and found that folds from oxygen-independent reactions had higher specificity than folds from oxygen-dependent reactions ($P < 10^{-7}$, Mann-Whitney U-Test).