PNAS RESEARCH ARTICLE | EVOLUTION BIOPHYSICS AND COMPUTATIONAL BIOLOGY OPEN ACCESS

Protein cost minimization promotes the emergence of coenzyme redundancy

Joshua E. Goldford[a,b,c,1], Ashish B. George[d,e,f], Avi I. Flamholz[g], and Daniel Segrè[c,d,h,i,1]

Edited by Paul Falkowski, Rutgers, The State University of New Jersey, New Brunswick, NJ; received June 17, 2021; accepted February 14, 2022

Coenzymes distribute a variety of chemical moieties throughout cellular metabolism, participating in group (e.g., phosphate and acyl) and electron transfer. 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. Although the physiological roles of various coenzyme systems are well established, it is unclear what selective pressures may have driven the emergence of coenzyme redundancy. Here, we use genome-wide metabolic modeling approaches to decompose the selective pressures driving enzymatic specificity for either NAD(H) or NADP(H) in the metabolic network of *Escherichia coli*. We found that few enzymes are thermodynamically constrained to using a single coenzyme, and in principle a metabolic network relying on only NAD(H) is feasible. However, structural and sequence analyses revealed widespread conservation of residues that retain selectivity for either NAD(H) or NADP(H), suggesting that additional forces may shape specificity. Using a model accounting for the cost of oxidoreductase enzyme expression, we found 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 protein minimization generically promotes coenzyme redundancy and that coenzymes typically thought to exist in a single pool (e.g., coenzyme A [CoA]) may exist in more than one form (e.g., dephospho-CoA).

metabolism | evolution | coenzymes | constraint-based modeling

Significance

Metabolism relies on a small class of molecules (coenzymes) that serve as universal donors and acceptors of key chemical groups and electrons. Although metabolic networks crucially depend on structurally redundant coenzymes [e.g., NAD(H) and NADP(H)] associated with different enzymes, the criteria that led to the emergence of this redundancy remain poorly understood. Our combination of modeling and structural and sequence analysis indicates that coenzyme redundancy may not be essential for metabolism but could rather constitute an evolved strategy promoting efficient usage of enzymes when biochemical reactions are near equilibrium. Our work suggests that early metabolism may have operated with fewer coenzymes and that adaptation for metabolic efficiency may have driven the rise of coenzyme diversity in living systems.

Group and electron transfer reactions form the foundation for more complex metabolic pathways in biological systems. For specific types of biochemical transformations, molecular components are shuttled within the cell through the action of small-molecule coenzymes. For the majority of these transformations, there often exist multiple chemically distinct coenzymes capable of similar chemistry (Fig. 1*A*). For example, electron transfer between common functional groups widely distributed throughout cellular metabolism (e.g., alcohols, aldehydes, and activated carboxylic acids) 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) (Fig. 1*B*). 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 thermodynamic 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 (3–7). For example, in *Escherichia coli* grown aerobically with glucose as the sole source of carbon, the NADH/NAD^+ ratio is ∼0.03, while the NADPH/$NADP^+$ ratio exceeds 57 (3, 8). This thermodynamic difference is commonly assumed to be a universal way for cells to simultaneously oxidize and reduce metabolites that might be impossible with just a single nicotinamide coenzyme (2, 9). Although the difference between the NADH/NAD^+ and NADPH/$NADP^+$ ratios in vivo enables tuning of thermodynamic potentials, the quantitative mechanism(s) that relate cellular fitness to these thermodynamic differences have not been addressed.

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

Author contributions: J.E.G. and D.S. designed research; J.E.G., A.B.G., and A.I.F. performed research; J.E.G. analyzed data; and J.E.G., A.I.F., and D.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: goldford@mit.edu or dsegre@bu.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2110787119/-/DCSupplemental.

Published March 28, 2022.

A

Chemical moiety	Coenzyme class	Examples
Electrons	Ferredoxins	Many
	Nicotinamides	NAD^+ , $NADP^+$
	Disulfides	Thioredoxins, Glutaredoxins
	Flavins	FAD, FMN
	Quinones	Ubiquinone, Plastoquinone
Phosphate	Nucleoside triphosphates	ATP, CTP, GTP
Methyl	Many	SAM, Cobalamin, Folate

B

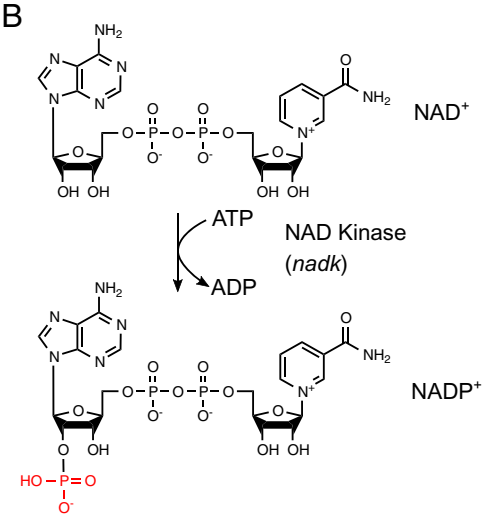

Fig. 1. Coenzyme redundancy is a widespread biological phenomenon. (*A*) A table of examples of redundant coenzymes for various classes of coenzymes. Note that for electron transfer reactions, different classes of coenzymes have varying characteristic ranges of standard redox potentials, enabling electron transfer between distinct classes of functional groups (27). (*B*) Structure of the nicotinamide adenine diphosphate (NAD^+) and the phosphorylated derivative ($NADP^+$). The phosphoryl transfer enzyme NAD kinase phosphorylates NAD^+ on the 2′ hydroxyl to make $NADP^+$.

sufficient, is it possible to quantitatively explain how evolution could gradually lead to the selective advantage of two coenzymes? Is the rise of multiple coenzymes an idiosyncratic step along a complex historical process, or a predictable universal feature of enzyme-driven biochemical networks?

While prior experimental work has shown that a range of cellular phenotypes can be sensitive to the coenzyme preferences for various oxidoreductases, they do not rule out the possibility of a simpler precursor to the NAD(H)/NADP(H) system. NAD(P)-dependent oxidoreductases can display selectivity for either NAD(H) or NADP(H), and perturbations to this specificity can strongly affect cellular phenotypes and fitness (10, 11). For example, mutagenesis experiments have shown that altering coenzyme specificity of isopropylmalate dehydrogenase and isocitrate dehydrogenase can severely reduce growth rate in *E. coli* (11–15). 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 by-products (10, 16–18). 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 (19). However, it is unclear if most enzymes are typically constrained to use only NAD(H) or NADP(H), and whether a single coenzyme could support the functioning of metabolism. It is also unclear how the thermodynamic potentials controlled by $NADH/NAD^+$ and $NADPH/NADP^+$ ratios in vivo translate into quantitative mechanism(s) that affect cellular fitness.

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 (20–25) 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) (26). Although recent quantitative work has explored the logic for why NAD(P)H emerged as a prominent coenzyme in biochemistry (27), a quantitative theory predicting the emergence of both NAD(H) and NADP(H), rather than just a single coenzyme, remains unexplored.

Here we use computational systems biology and various network-level models of metabolism to systematically probe, in a way that could not be easily addressed experimentally, the possible advantages of maintaining multiple pools of biochemically redundant redox coenzymes. By taking advantage of large-scale computational experiments using both core and genome-scale stoichiometric models, we show that the *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. The widespread sequence conservation of NAD(H)/NADP(H) selective residues, however, seems to contradict this low sensitivity, suggesting that factors beyond flux stoichiometry may drive coenzyme redundancy. By developing a model of enzyme cost as a function of enzyme–coenzyme specificity we found that this conundrum can be resolved by viewing coenzyme redundancy as a graded improvement that enables efficient usage of protein resources when reactions are near thermodynamic equilibrium. These results suggest that widespread coenzyme redundancy in metabolic networks is the outcome of evolution for increased metabolic efficiency and parsimonious use of protein at the whole-cell level. Our findings may cast new light on the early stages and emergence of coenzyme couplings and could serve as design principles for the construction of modified or artificial metabolic circuits for synthetic biology and metabolic engineering applications.

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 (15). To investigate the consequences of altering the NAD(P)-specificity of individual oxidoreductases, we used flux balance analysis (FBA) (28) to simulate

the growth rates of in silico mutants with altered coenzyme specificity in 117,180 different growth media. Specifically, we altered the iJO1366 (29) *E. coli* genome-scale metabolic model (GEMM) 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. Since the in vivo $NADH/NAD^+$ ratio is drastically different from the $NADPH/NADP^+$ ratio (3, 8), some reactions that are reversible with the original coenzyme coupling may become irreversible upon switching coenzymes (*SI Appendix*, Fig. S2 *A* and *B*), even when taking into consideration the possible range of concentrations of other metabolites in the cell (*Materials and Methods*). Thus, upon switching coenzymes, we reevaluated the thermodynamic feasibility of each reaction in either direction and translated this information into constraints included into our altered metabolic model (*SI Appendix*, Fig. S2 *A* and *B* and *Materials and Methods*). We compiled a list of 76 genes encoding NAD(P)-coupled oxidoreductases (*SI Appendix*, Table S1) and generated 76 single-mutant metabolic models by switching the coenzyme preference of each enzyme, one at a time. For each model mutant, we then 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 as electron acceptors, on combinations of 180 carbon sources and 93 nitrogen sources, using sulfate and orthophosphate as sulfur and phosphorus sources, respectively (*Materials and Methods* and *SI Appendix*, Table S2).

We first counted the frequency of lethal mutations across all in silico media conditions (Fig. 2*B* and *Materials and Methods*), expecting coenzyme switching to cause major fitness defects for each gene under at least some conditions. To our surprise, the vast majority (97.3%, 8,103,217 of 8,323,596) of environment and coenzyme mutation pairs were nonlethal, suggesting that metabolism is overall robust to individual swaps of coenzyme specificity. However, of the 219,092 lethal mutant–environment pairs, all came from forcing one of six NAD(H)-coupled oxidoreductases to use NADP(H). These genes are *leuB*, *pdxB*, *paaH*, *gatD*, *lgoD*, and *fucO* (Fig. 2*B*). By further constraining the range of allowed metabolite concentrations, we also found environments where sugar alcohol oxidoreductases encoded by *srlD* and *mltD* require coupling to NAD(H) rather than NADP(H) (*SI Appendix*, Fig. S1). Only two oxidoreductases, encoded by the two genes *leuB* and *pdxB*, were predicted to require NAD(H) as a coenzyme rather than NADP(H) in all conditions tested (Fig. 2*B*). This is in agreement with extensive work by Dean and coworkers, who demonstrated that the gene product of *leuB*, isopropylmalate dehydrogenase, is strongly specific for NAD(H) and that introducing mutations to alter specificity toward NADP(H) always reduced fitness (13). Notably, all the observed lethal gene–environment pairs are alleviated when we do not take into consideration the thermodynamic differences between NAD(H) and NADP(H) (*SI Appendix*, Fig. S2), demonstrating that the thermodynamic driving force is a dominant factor in constraining the coenzyme preference in our model.

Our flux predictions suggest that enzymes that are sensitive to switching coenzymes are limited to NAD(H)-dependent, and not NADP(H)-dependent, oxidoreductases, suggesting that a metabolism using only NAD(H) might be feasible. To test this scenario, we constructed an NAD(H)-only model of *E. coli* metabolism by swapping all NADP(H)-coupled reactions with NAD(H) and removing $NADP^+$ synthesis and degradation reactions (*Materials and Methods*). We systematically computed maximum growth rates across a broad gradient of $NADH/NAD^+$ ratios (Fig. 2*C*) and found that a wide range of NADH/NAD+ ratios enabled growths similar to that of the unmodified model (Fig. 2*D*). The feasibility of an NAD(H)-only *E. coli* metabolism is also corroborated by an alternative approach, i.e., thermodynamic flux balance analysis (TFBA), applied to a core and genome-scale *E. coli* model (*SI Appendix*, Figs. S3 and S4). Interestingly, TFBA modeling of a NAD-only genome-scale model suggests a decreased growth rate relative to the unmodified model, where flux through lower glycolysis was blocked (*SI Appendix*, Fig. S4 *A* and *B* and Table S3). We also computed relative growth rates of each of the 76 mutant models described previously using TFBA and found that, unlike results described in Fig. 2*A*, no single coenzyme swap reduced growth rate (*SI Appendix*, Fig. S4*C*). This discrepancy likely results from the fact that TFBA allows for variation in $NAD(P)^+/NAD(P)H$ ratios while our previous modeling approach (Fig. 2 *A* and *B*) fixed these ratios to measured values.

Altogether, these results challenge the commonly accepted notion that two redox coenzymes are essential for running metabolism at the whole-cell level and suggest that the reasons for coenzyme redundancy and NAD(H) or NADP(H) specificity might have to be sought elsewhere.

Amino Acid Residues Conferring Selectivity to NAD(H) or NADP(H) Are Well Conserved. Flux modeling predictions suggest that while the maximum organismal growth rate is insensitive to the coenzyme specificity for the majority of enzymes, a few genes are expected to be highly constrained to a single coenzyme specificity, 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.

Since protein structural features that shape coenzyme selectivity are well known, we reasoned that the analysis of conservation of residues that occlude or coordinate the 2′ phosphate on the ribosyl moiety of NADP(H) (10) would enable us to access the degree of coenzyme preference conservation independent of cosubstrate specificity. We compiled a list of 57 NAD(P)-dependent oxidoreductases from *E. coli* with NAD(P)-bound crystal structures of the enzyme (or a homolog with greater than 30% homology) in the Protein Data Bank (PDB) and annotated residues that confer selectivity to NAD(H) or NADP(H) (*SI Appendix*, Table S4). We next performed multiple sequence alignment for orthogroups for each oxidoreductase, and quantified the conservation of the NAD(P)-selective residues across orthologs (*Materials and Methods*). We found that residues that confer selectivity for NAD(H) or NADP(H) were significantly less variable than the remaining portion of the protein sequences but more variable than active-site residues (Fig. 3 *A* and *B*). Interestingly, residues that confer selectivity to NAD(H) were more conserved than NADP(H), which is consistent with our predictions that metabolic systems appear more sensitive to NAD(H) coupling compared to NADP(H).

Despite the overall high conservation of NAD(H)-binding residues relative to NADP(H)-binding ones, the flux balance prediction suggest that NAD(H)-dependent enzymes have a broad range of dependency on the coenzyme preference, with some being clearly much more sensitive than others. In exploring the relationship between NAD(H) selectivity and flux

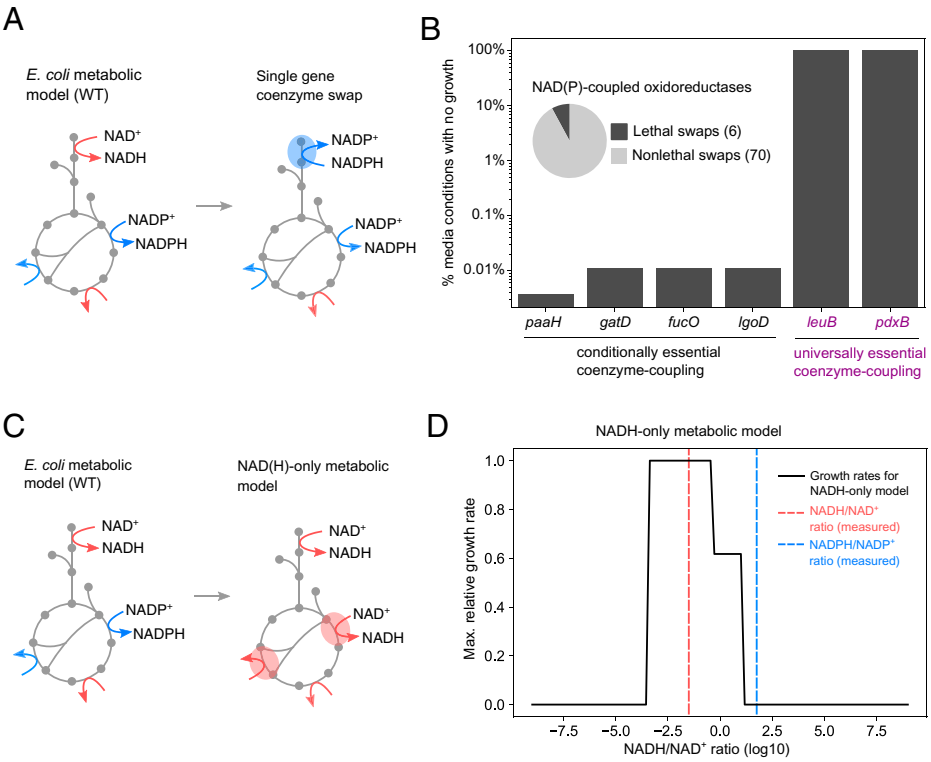

Fig. 2. Coenzyme "rewiring" in genome-scale metabolic models. (*A*) We determined the consequences of "rewiring" the *E. coli* metabolic network iJO1366 by changing the coenzyme specificity of 76 genes encoding NAD(P)-dependent oxidoreductases and stimulated the growth across 117,180 growth media spanning different combinations of 180 carbon sources, 93 nitrogen sources, and 6 electron acceptors (oxygen, nitrite, nitrate, TMAO, DMSO, and fumarate) as well as fermentative growth using FBA. (*B*) Only 6 of 76 genes were predicted to require a specific nicotinamide cofactor in one or more media conditions (*x* axis, *Inset*). The percentage of the media conditions that led to no growth is plotted as a bar graph (*y* axis), showing that only two oxidoreductases are constrained to the endogenous coupling in all conditions (purple text): the NAD(H)-coupled isopropylmalate dehydrogenase *leuB* and erythronate 4-phosphate dehydrogenase *pdxB*. (*C*) We next altered the *E. coli* metabolic model by replacing all NADP(H)-coupled reactions with NAD(H)-coupled reactions (*Materials and Methods*) and simulated growth at various concentration ratios of NADH/NAD^+. (*D*) We computed the maximum growth rate of the NAD(H)-only metabolic model (*y* axis) at various concentration ratios for NADH/NAD^+ (*x* axis, black line) and found that a wide range of ratios enabled growth at comparable levels to the unmodified model. The red and blue dashed line indicates the measured concentration ratios in *E. coli* for NADH/NAD^+ (0.032) and NADPH/$NADP^+$ (57.14), respectively. Note that a single coenzyme poised at the measured NADH/NAD^+ ratio could enable cellular growth.

balance predictions, we first investigated in detail *leuB* and *pdxB*, which both require NADH coupling to permit biomass production and growth in silico. While *leuB* has been previously shown to use only NAD(H) (13), *pdxB*, which encodes the enzyme erythronate-4-phosphate dehydrogenase involved in de novo pyridoxal 5′-phosphate biosynthesis, has received less attention. We computed the degree of conservation of the coenzyme-binding selective residue (Asp146; *SI Appendix*, Fig. S5) across all *pdxB* orthologs from the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (orthogroup K03473, $n = 1087$) and performed multiple sequence alignment on the coenzyme-binding Rossman fold (*Materials and Methods*). We found that Asp146 was conserved in all *pdxB* orthologs, consistent with predictions that *pdxB* is universally required to use NAD(H) like *leuB* (Fig. 3*B*).

The above analysis of the relationship between flux balance prediction and structural conservation can be extended to all NAD(H)-dependent oxidoreductases. We found 25 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) (*SI Appendix*, Table S4). We performed multiple sequence alignment for orthogroups for each oxidoreductase and quantified the conservation of the aspartic acid or glutamic acid residue across orthogroups. We next compared this degree of conservation to the minimum relative growth of coenzyme mutant models across all 117,180 media conditions. We found that all enzymes predicted to have a severe fitness defect upon cofactor switching (<10% of wild type) have a high conservation (>80%) of the coordinating Asp or Glu residues (Fig. 3*D*). Conversely, oxidoreductases with low conservation of the coordinating Asp or Glu residue were not predicted to have a fitness defect after forcing coupling to NADP(H) with flux modeling (e.g., *rfbD*). However, several oxidoreductases with high conservation of the coordinating Asp or Glu residue either were not predicted to have a fitness defect after forcing coupling to NADP(H) in any media condition (e.g., *hisD*). Thus, flux balance predictions only partially predict the observed cofactor preference, as measured by structural conservation.

Altogether, the results from genome-scale simulations (previous section) suggest that metabolism, in principle, can operate with a single nicotinamide coenzyme system, strikingly contradicting the textbook assumption that having two coenzymes poised at two distinct in vivo thermodynamic potentials is essential. Models also predict that NAD(H)-dependent reactions are more sensitive to altering the coenzyme coupling than NADP(H)-dependent reactions. While this hypothesis is supported by the overall pattern of conservation at the specificity-determining residues in NAD(P)-dependent oxidoreductases, a detailed analysis indicates that several enzymes with high conservation are predicted to have no detectable sensitivity based on flux balance models, suggesting that additional selective pressures not captured by

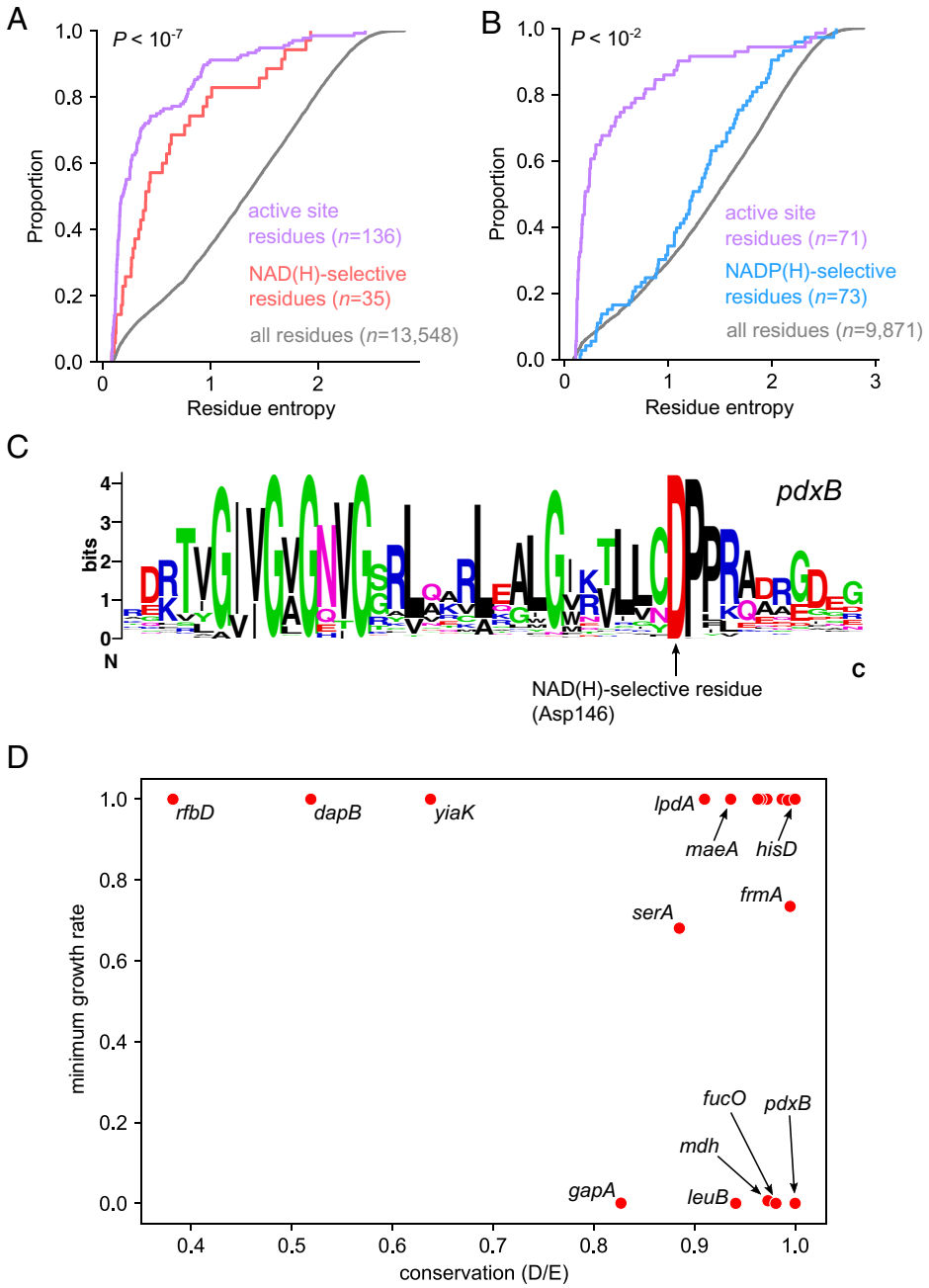

Fig. 3. Structural analysis of NAD(P)-dependent enzymes and connection to FBA predictions. (*A*) We computed the entropy for each residue that conferred selectivity for NAD(H) in 35 KEGG orthogroups encoding NAD-coupled oxidoreductases and plotted the empirical cumulative distribution (eCDF) for residues that are responsible for NAD(H) selectivity (red) vs. all other residues in the protein sequences (gray). (*B*) The same as *A*, but for 25 NADP(H)-dependent oxidoreductases, where the entropy distribution for NADP(H)-selective residues are plotted in blue. NAD(H) and NADP(H) selective residues displayed significantly lower entropy than nonselective residues [Mann–Whitney *U* test: $P < 10^{-7}$ for NAD(H)-dependent oxidoreductases and $P < 10^{-2}$ for NADP(H)-dependent oxidoreductases]. For both *A* and *B*, we plotted the eCDFs for residue entropy for active site residues (purple lines), obtained from M-CSA (44). (*C*) Orthologs of *pdxB* were retrieved from the KEGG database (K03473). The Rossman fold was identified using HMMR 3.1 (45) to generate a sequence logo for the NAD(H) binding domain of these enzymes (*Materials and Methods*). The arrow denotes the 100% conservation of the aspartic acid residue (Asp146) responsible for discriminating between NAD(H) and NADP(H). (*D*) For each enzyme with an Asp or Glu residue that specified selectivity for NAD(H) ($n = 25$), we constructed flux balance models with single oxidoreductases coupled to an alternative coenzyme and computed maximum growth rate across 117,180 different media. We computed then plotted the minimum growth rate (relative to wild-type growth) across all media conditions (*y* axis) versus the conservation of the ribosyl-coordinating Asp or Glu (*x* axis).

our flux modeling might drive coenzyme redundancy. In particular, while metabolism with an individual cofactor seems potentially viable, it may be highly inefficient, due to hidden costs that cells may need to pay in order to maintain metabolic fluxes. In the following section, we explore a model of metabolic fluxes that explores the potential relationship between these crucial hidden costs (i.e., enzyme abundance) and coenzyme thermodynamic potentials to test the hypothesis that coenzyme redundancy might lead to a quantitative fitness advantage in extant metabolism.

Enzyme Cost Minimization Drives Separation of Coenzyme Preference Globally. Thus far, the fitness consequences of modifying oxidoreductase specificity for different coenzymes has been assessed by computing maximum growth rates using stoichiometric models of metabolism with altered coenzyme

couplings for individual enzymes. 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 models also suggest that a metabolism with a single coenzyme (poised close to the in vivo NADH/NAD^+ ratio) can satisfy these constraints, which is in stark contrast with the observed widespread specificity for NAD(H) or NADP(H) by oxidoreductases (Fig. 3).

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 (30–33). This effect comes directly from the following flux–force relation (34):

$$\frac{v_f}{v_r} = e^{-\frac{\Delta G'_r}{RT}},$$

where $\Delta G'_r$ is the free energy at physiological metabolite concentrations and v_f and v_r are the forward and reverse reaction flux, respectively. Note that as the free energy decreases (i.e., it becomes more negative), the proportion of flux that is in the forward direction increases, and thus decreases the enzyme abundance required for catalysis (30, 31). We hypothesized that coenzyme redundancy might allow the cell to lower the total protein required to maintain intracellular reaction fluxes. As in previous work, we used the flux–force relation to model the interrelationship between thermodynamics and enzyme cost (30–33). Specifically, 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. 4*A* and *Materials and Methods*).

Given the uncertainty about many model parameters, we generated predictions based on extensive random sampling of key parameters, including the reaction network topology, the effective equilibrium constant for each reaction r (K'_r), the maximal turnover rates (κ_r), and the flux through each reaction (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 (*SI Appendix, Supporting Information Text*). Motivated by the fact that specific concentration profiles in the flux modeling enable a viable single-coenzyme metabolism, we investigated how coupling reactions to multiple coenzymes influences enzyme cost at these fixed concentrations. Our model also relies on quantities related to the thermodynamic drive of each coenzyme as well as the flux partitioning between different coenzyme pools. These features are captured by the ratio of charged to uncharged species for various coenzyme pairs ($\Gamma_\alpha = \frac{[\mathrm{NADH}]}{[\mathrm{NAD}]}$, $\Gamma_\beta = \frac{[\mathrm{NADPH}]}{[\mathrm{NADP}]}$) 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 used nonlinear optimization techniques to find the coenzyme concentration ratios ($\Gamma_\alpha, \Gamma_\beta$) and the reaction fluxes through each

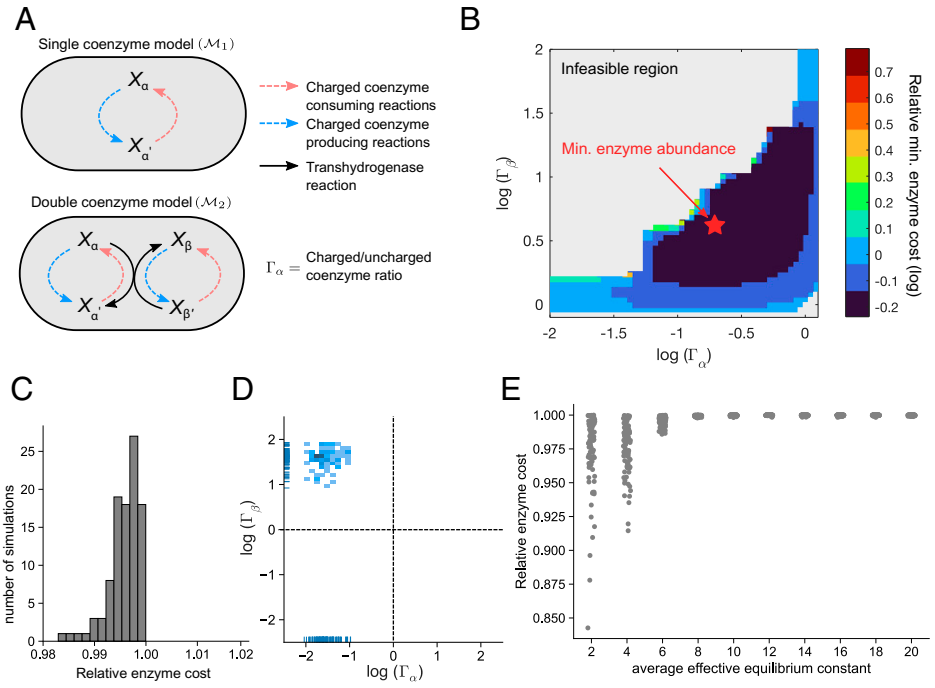

Fig. 4. A minimal model of coenzyme-utilizing enzyme cost predicts the emergence of coenzyme redundancy. (*A*) Model of a reaction network where multiple reactions (dashed arrows) produce and consume a charged coenzyme, where reaction flux is feasible with a single coenzyme (*Top*), compared to a model where reaction flux is partitioned between two coenzyme pools connected by a transhydrogenase reaction (*Bottom*). (*B*) A heat map of the minimum enzyme cost (scaled by the enzyme cost for a single-coenzyme model) as a function of various coenzyme ratios (*x* and *y* axis) for randomly sampled reaction fluxes and thermodynamic and kinetic parameters (*Materials and Methods*). The color scale denotes the minimum relative enzyme cost at each coenzyme ratio (log-scale), and the gray region denotes the region of space that is thermodynamically infeasible. Note that the dark blue regions denote the space of coenzyme ratios that enables enzyme cost to be less than the single coenzyme, and the red star denotes the point where the minimum enzyme abundance is achieved. (*C*) We simulated 100 random realizations and plotted the distribution of the minimal enzyme cost relative to the cost using a single coenzyme system (*x* axis) and found that all simulations were lower than unity. (*D*) The coenzyme ratio for coenzyme α (*x* axis) vs. coenzyme β (*y* axis) that minimizes total protein abundance were plotted for each realization as a two-dimensional 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 *SI Appendix, Supporting Information Text* for definitions and derivations).

coenzyme pool ($v_{r\alpha}, v_{r\beta}$) that minimized the total protein abundance in the network (Fig. 4*B* and *SI Appendix*, *Supporting Information Text*). We found that partitioning flux through two coenzyme pools decreased the minimum protein abundance required relative to the single coenzyme scenario (Fig. 4*B*). Notably, at the optimum, 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 (*SI Appendix*, Fig. S6*A*). The emergence of enzyme specificity for one of the two coenzymes was driven by the combination of both kinetic and thermodynamic factors (*SI Appendix*, Fig. S6 *B* and *C*), where reactions that required less protein to maintain the given flux demand were used to rebalance coenzyme pools, operating against the coenzyme concentration gradient (*SI Appendix*, Fig. S6 *B* and *C*).

To see if these results were generic for different realizations of parameters, we computed optimal flux distributions and coenzyme ratios for 100 random instances. The two-coenzyme model always achieved lower total protein cost than the one-coenzyme model (Fig. 4*C*). 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. 4*D*). 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 (3, 8). Interestingly, our model also predicts that the largest reduction in enzyme cost comes when the mean effective equilibrium constants are closer to unity (Fig. 4*E*), suggesting that two-coenzyme systems may be particularly beneficial if the coupled reactions are held closer to equilibrium.

Motivated by the expectation that enzymes catalyzing reactions near equilibrium should display stronger coenzyme preference, we revisited results presented in Fig. 3*D* to see if protein-cost minimization could explain any of the enzymes that were predicted to have low sensitivity to coenzyme preference using FBA but displayed a high degree of conservation of key residues. We found that the NAD(H)-dependent histidinol dehydrogenase (encoded by *hisD*) catalyzes a reaction operating far from equilibrium (−26.6 kJ/mol) in physiological conditions. After swapping for NAD(P), the free energy rises to slightly below equilibrium (−1.4 kJ/mol), which translates to an approximately threefold increase in estimated protein cost to run this reaction (*SI Appendix*, Fig. S7). Even after accounting for variable metabolite concentrations, the minimum protein cost increase would be ∼6%, indicating the protein cost considerations for coenzyme-catalyzed reactions could be a factor shaping coenzyme selectivity.

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 (Fig. 5*A*). To test this prediction, we computed free energies for all NAD(P)-coupled oxidoreductases in the KEGG database using eQuilibrator (35, 36) and computed binding preferences of protein sequences for NAD(H) and NADP(H) using an artificial neural network (37), 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 was bimodal, where oxidoreductases using oxygen as a cosubstrate were driven far from equilibrium ($\Delta G_r'^{\circ} = 421.8^{+/-}155$ kJ/mol) compared to oxygen-independent oxidoreductases ($\Delta G_r'^{\circ} = 12.4^{+/-}48.5$ kJ/mol; *SI Appendix*, Fig. S8), providing a simple way to categorize oxidoreductases by thermodynamic potential. We computed binding preferences for NAD(H) and NADP(H) (Fig. 5*B*) and computed the proportion of Rossman folds that were single-coenzyme binders [e.g., NAD(H) or NADP(H)] or multicoenzyme binders [e.g., NAD(H) and NADP(H)]. Consistent with predictions from the theory, we found that 61% (2,139/3,497) of folds from oxygen-coupled NAD(P)-dependent oxidoreductases are expected to be multicoenzyme binders, compared to just 23% (28992/126718) from nonoxygen-coupled NAD(P)-dependent oxidoreductases (Fisher's exact test: $P < 10^{-5}$) and that the estimated binding preferences are more promiscuous for either NAD(H) or NADP(H) in nonoxygen-coupled NAD(P)-dependent oxidoreductases (Mann–Whitney *U* test: $P < 10^{-5}$; *SI Appendix*, Fig. S5*C*). Since many confounding factors may shape coenzyme dependencies for oxygen-dependent oxidoreductases, we corroborated these results by analyzing binding specificities for protein folds from oxygen-independent reactions only by comparing folds that catalyze reactions near equilibrium versus reactions far from equilibrium (*Materials and Methods* and *SI Appendix*, Fig S5*D*). We found that folds are more promiscuous for either NAD(H) or NADP(H) if the folds are used to catalyze far-from-equilibrium reactions versus protein folds used to catalyze near-equilibrium reactions (Mann–Whitney *U* test: $P < 10^{-5}$; *SI Appendix*, Fig. S5*C*).

Dephospho-Coenzyme A Might Be a Redundant Coenzyme for Acyl Transfer Reactions. The simple phenomenological model presented above (Fig. 4) 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. Indeed, according to the model presented in Fig. 4, additional coenzyme redundancy decreases protein cost monotonically (*SI Appendix*, Fig. S9*A*), which potentially explains the vast degree of redundancy for protein-coded coenzymes like ferredoxins, glutaredoxins, and thioredoxins in microbial genomes (*SI Appendix*, Fig. S9*B*). This begs the question of why several coenzymes appear to exist as only one variant in the cell.

One interesting example is coenzyme A (CoA), a molecule involved in the redistribution of intracellular acyl groups. We computed free energies for various group transfer reactions (Fig. 6*A*) 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 CoA-coupled acetyl-transfer reactions. Interestingly, the last step of CoA biosynthesis resembles the last step of NADP biosynthesis, where a kinase acts on the ribosyl moiety of the coenzyme structure (Fig. 6*B*). For the case of $NADP^+$ biosynthesis, NAD kinase phosphorylates NAD^+ at the 2′ hydroxyl of the ribosyl moiety to create $NADP^+$. For the case of CoA, dephospho-CoA kinase phosphorylates the 3′ hydroxyl of the ribosyl moiety of dephospho-CoA to produce CoA. The

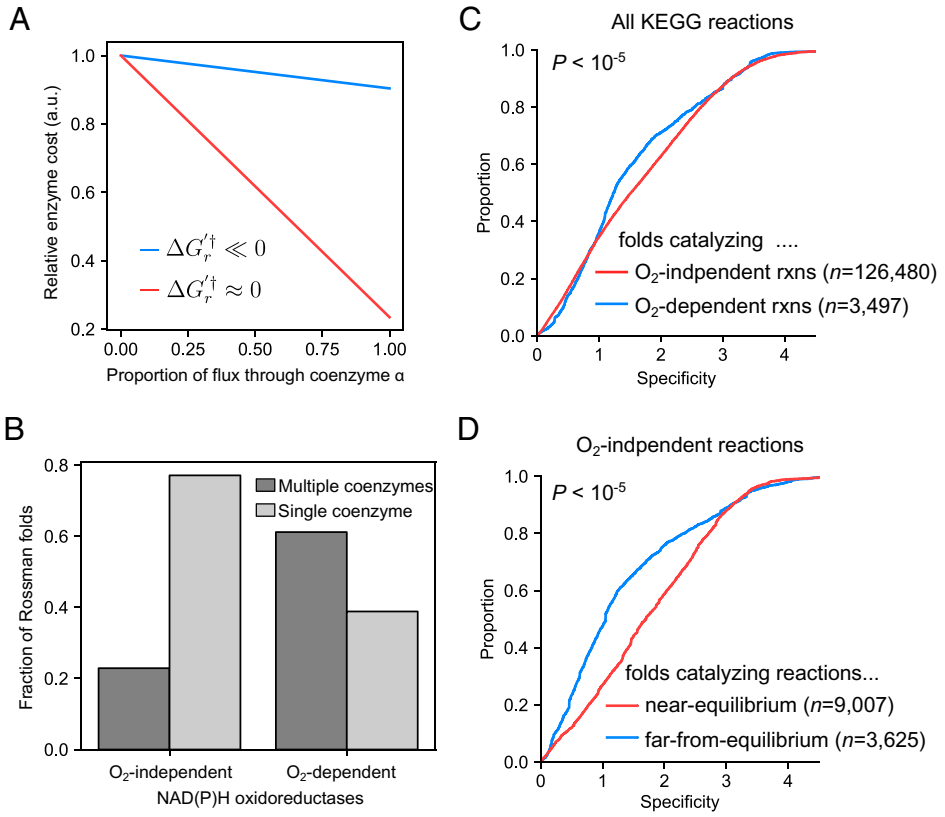

Fig. 5. NAD(P)-oxidoreductases catalyzing reactions far from equilibrium are predicted to be more coenzyme-promiscuous. (*A*) The relationship between the degree of flux partitioning between two coenzyme pools (*x* axis) and enzyme cost (*y* axis, rescaled by the maximum enzyme cost per model) at both low (red) and high (blue) coenzyme-independent driving forces ($\Delta G_r'^{\dagger}$). Note that at high driving forces, the sensitivity of enzyme cost to flux partitioning is significantly less than at low driving forces. (*B*) The proportion of Rossman folds that were predicted using an artificial neural network model (37) to bind a single coenzyme [light gray bars, e.g., NAD(H) or NADP(H)] or multiple coenzymes [dark gray 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 (2,139/3,497) were multicoenzyme binders, compared to just 23% among oxygen-dependent NAD(H)/NADP(H) oxidoreductases (28,992/126,718) (Fisher's exact test: $P < 10^{-5}$). (*C*) The empirical cumulative distribution (eCDF) of binding specificities for NAD(H) or NADP(H) for protein folds catalyzing NAD(P)-coupled redox reactions (*Materials and Methods*). We partitioned folds based on whether the folds catalyzed reactions that either used (blue line) or did not use (red line) oxygen as a substrate or product. Specificity is lower for folds involved in O_2-dependent reactions than in folds involved in O_2-independent reactions (Mann–Whitney *U* test: $P < 10^{-5}$). (*D*) Same as *C*, but just limited to folds involved in O_2-independent reactions. We partitioned folds as far from equilibrium (blue line) if they catalyzed reactions with absolute standard molar free energies above 62.7 kJ/mol (top 10% of all free energies for all O_2-independent reactions) or near equilibrium (red) if they catalyzed reactions with absolute standard molar free energies below 6.6 kJ/mol (bottom 10% off free energies for all O_2-independent reactions). Specificity is lower for folds involved in far-from-equilibrium than in folds involved in near-equilibrium reactions (Mann–Whitney *U* test: $P < 10^{-5}$).

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, mass spectrometry experiments have shown that dephospho-CoA can be acylated in rat liver (38). We thus propose that dephospho-CoA may be a redundant coenzyme for acyl-transfer reactions rather than simply serving as an intermediate during CoA biosynthesis.

Discussion

The evolution of coenzyme coupling in redox biochemistry constitutes a fascinating puzzle central to the emergence and complexification of life itself (24, 39). In order to understand the constraints and the selection pressure that led to the enzyme–coenzyme couplings observed today, one would ideally shuffle all of these couplings (e.g., by reassigning coenzymes 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 both core and 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 universally constrained to use NAD(H) or NADP(H) and that most 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 (13).

Our flux modeling results suggest that thermodynamic constraints play a role in shaping the preferences of just a small number of oxidoreductases for specific coenzymes. Structural and bioinformatics analyses of *pdxB* suggest that these thermodynamic constraints can create strong selective pressures shaping protein sequence evolution, as 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) (Fig. 3*C*). However, our results also suggest that residues conferring selectivity for NAD(H) or NADP(H) in oxidoreductases are well conserved (Fig. 3 *A* and *B*), indicating that constraint-based modeling may not adequately capture the principal factors driving enzyme specificity for individual coenzymes.

The model presented in Fig. 4 suggests that the emergence of widespread enzyme specificity toward one of many coenzymes

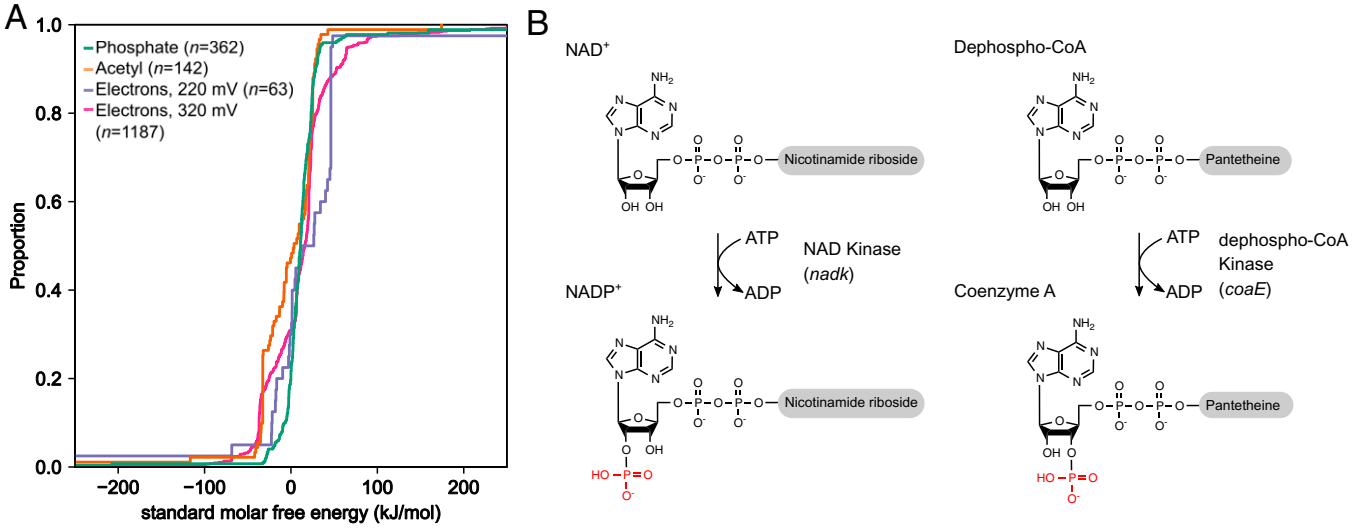

Fig. 6. Comparison between NAD(P) and CoA biosynthesis reveals potential redundancy for acyl-CoAs. (*A*) Free energies for group transfer reactions of phosphate groups, acetyl groups, and electron transfer reactions in KEGG were computed using eQuilibrator (35, 36), and the empirical cumulative distributions across reactions were plotted. Free energies for all group and electron transfer reactions are similarly distributed near equilibrium, suggesting that CoA-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: CoA/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 2′ hydroxyl group on the ribose moiety of NAD^+, compared to the biosynthesis of CoA, where the 3′ hydroxyl of the ribose moiety is phosphorylated by dephospho-CoA kinase (E.C. 2.7.1.24).

could be induced by a selective pressure to minimize the total abundance of enzymes in the cell. Importantly, our model does not rely on chemical details specific to nicotinamide coenzymes or electron transfer and could potentially explain the ubiquity of multicoenzyme systems (e.g., *SI Appendix*, Fig. S9 *A* and *B*) in cellular metabolism. Notably, this model suggests that enzyme specificity toward 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. The quantitative mechanistic link that our model establishes between cofactor redundancy and cellular fitness could be directly tested through physiological measurements of kinetic constants, metabolite abundances, and thermodynamic potentials of all NAD(P)-dependent oxidoreductases in an organism. Additionally, feasibility of an NAD(H)-only metabolism supports the hypothesis that ancient metabolic networks may have originally relied on NAD(H) only and gradually evolved NADP(H) dependencies. Future computational work could further assess the feasibility of an NAD(H)-only metabolic network by analyzing metabolite concentration bounds and comparing them to physiological constraints or by analyzing the feasibility of NAD(H)-only metabolic pathways beyond those found in *E. coli*, and thought to be important for ancient metabolism. While it has been proposed that an NADP(H)-dependent isocitrate dehydrogenase evolved from an ancestral NAD(H)-dependent enzyme in prokaryotes (14), future efforts to reconstruct ancestral sequences from various NAD(P)-dependent oxidoreductases can explore whether this is a general feature of NADP-coupled enzymes (40).

The flux analysis presented in this study only considers stoichiometric and thermodynamic factors shaping cell fitness. However, other constraints or metabolic demands exist that may constrain oxidoreductase coenzyme preference. For example, our flux models may not accurately capture metabolic demands induced by oxidative stress, as NADPH is the primary electron donor for enzymatic detoxification of reactive oxygen species and other damaging oxidants (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 cosubstrate specificity, kinetics (41), or thermostability (42). Nonetheless, our results suggest that thermodynamics and enzyme cost minimization are important factors that can constrain the oxidoreductase specificity for NAD(H) or NADP(H).

In addition to offering an explanation for the occurrence of multicoenzyme systems, our model can be used for the targeted search for unexplored redundant coenzymes. Our results point to the intriguing possibility that dephospho-CoA might serve as an acyl-transfer coenzyme rather than simply an intermediate during CoA and could in principle explain the occurrence of acyl-dephospho-CoAs (38). More broadly, the principle that enzyme cost minimization universally promotes coenzyme redundancy may help guide the interpretation of molecular diversity in the biosphere (43), as well as serve as a design principle for metabolic engineering and synthetic biology applications.

Materials and Methods

Software Availability. Code is available in 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 GEMM and by simulating the growth rate under various media conditions. We used the GEMM iJO1366 (29) downloaded from the BiGG database (46), with Gibbs free energies derived using the component contribution method (47), using the eQuilibraitor application programming interface (API) at pH 7 and ionic strength 0.1 M (35, 36). 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 specificities, 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 linear program:

$$\underset{\ln(x)}{\text{minimize/maximize}} \quad z_r = \Delta G_r'^{\circ} + RTs_r{}^T \ln(x)$$
$$\text{subject to :} \quad \ln(10^{-6}) \leq \ln(x) \leq \ln(10^{-2}),$$

where s_r is the vector of stoichiometric coefficients for reaction r, $\Delta G_r'^{\circ}$ is the change in free energy at standard molar conditions for reaction r, R is the gas constant, T is temperature in Kelvin, and $\ln(x)$ denotes the vector of natural log-transformed metabolite concentrations. Note that we fixed the concentration of coenzymes based on previously measurements, where $[\text{NAD}] = 2.6 \times 10^{-3}$ M, $[\text{NADH}] = 8.3 \times 10^{-5}$ M, $[\text{NADP}] = 2.1 \times 10^{-6}$ M, and $[\text{NADPH}] = 1.2 \times 10^{-4}$ M. 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 recomputed the minimum and maximum driving force and used these values to set the upper and lower bounds of reactions, such that

$$l_r' = \begin{cases} 0 & \text{if } \max(z_r) < 0 \\ -1,000 & \text{otherwise} \end{cases}$$

and

$$u_r' = \begin{cases} 0 & \text{if } \min(z_r) > 0 \\ 1,000 & \text{otherwise} \end{cases},$$

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 computing the maximum growth rate using flux balance analysis (28) under 117,180 conditions, consisting of various carbon sources, nitrogen sources, and electron acceptors (oxygen, nitrite, nitrate, TMAO, DMSO, and fumarate) as well as fermentative growth, using the following linear program:

$$\underset{v}{\text{maximize}} \quad v_{\text{biomass}}$$
$$\text{subject to :} \quad S'v = 0,$$
$$l' < v < u'$$

where S', l', u' is the altered stoichiometric matrix, flux lower bound, and flux upper bound after updating stoichiometric and thermodynamic constraints. Note that for the NAD(H)-only metabolism presented in Fig. 2 *C* and *D*, we performed the same procedure described above but for all NADPH-coupled reactions, and under a range of different coenzyme ratios (Fig. 3 *C* and *D*). For these simulations, we removed NAD kinase (NADK), NAD phosphatase (NADPPPs), NAD diphosphatase (NADDP), and NAD transhydrogenase (NADTRHD) from the model and performed FBA in aerobic conditions with glucose as the sole carbon source.

For results presented in Fig. 2*B*, we restricted the analysis to the media sets where growth of the unperturbed network was feasible (109,521 conditions). We modeled mutations for 76 genes encoding oxidoreductases in *E. coli*. This set of genes was determined using the following criteria: 1) the gene encoded an enzyme that could be classified using the Enzyme Commission code E.C. 1.X.1.Z, where X was any number except 6 and Z = any number and 2) 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 (48).

Thermodynamic flux balance analysis was performed using pyTFA (49), with the reduced core *E. coli* metabolic model irJO1366 (*SI Appendix*, Fig. S3) (50) and with the genome-scale model iJO1366 (*SI Appendix*, Fig. S4). We modified both the core model, irJO1366, and the genome-scale model, iJO1366, 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 (49). To run TFBA on the genome-scale model iJO1366, we used the relax_dgo subroutine from pyTFA to find minimal free energy relaxations to achieve feasible growth, which are provided in *SI Appendix*, Table S5.

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(H) vs. NADP(H). To this end, we identified NAD(P) binding orthogroups (KO) using the KEGG REST API and developed a custom Python script to identify sequence of NAD(P)-bound structures from the PDB 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 interactions with the 2′ 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 orthologs in the KEGG database for *pdxB* (orthogroup K03473, $n = 1087$), and performed multiple sequence alignment on the coenzyme-binding Rossman fold. We 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 such that gap opening and extension were penalized much more than nonmatching 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 25 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) (*SI Appendix*, Table S3). For orthogroups with >1,000 orthologs, we downsampled the sequences to 1,000 before identifying running multiple sequence alignment.

We annotated residues that confer coenzyme selectivity for 57 NAD(P)-coupled oxidoreductases in *E. coli* based on structures available in the PDB (*SI Appendix*, Table S3). 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), such that

$$H_j = \sum_i x_{ij} \log(x_{ij})$$

and x_{ij} is the proportion of sequences with amino acid residue i in position j. Prior to computing distributions of entropy across residues that confer selectivity to either NADH or NADPH, we removed all residues with >50% gaps, as these represented poorly aligned regions of the protein sequence.

In Fig. 5 *B–D*, we computed coenzyme binding likelihoods using an artificial neural network model called Cofactory (37). 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. We identified 779 orthogroups that either use oxygen as a cosubstrate and downloaded all sequences within each orthogroup from the KEGG REST API (51). Like in the previous case, we down sampled the sequences to 1,000 if the orthogroup contained more than 1,000 sequences before running Cofactory. After obtaining predictions of binding likelihoods to NAD(H), NADP(H), and FAD(H2), we filtered out Rossman folds that either 1) unambiguously bound FAD(H2) or 2) were not predicted to bind any of the coenzymes. We then categorized the remaining Rossman folds as "multicoenzyme 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. We estimated coenzyme specificities (S_r) for each NAD(P)-oxidoreductase with a Rossman fold by computing binding likelihoods for NAD(H) (L_r^{NAD}) and NADP(H) (L_r^{NADP}) using the following formula:

$$S_r = \left| \log_2 \left(\frac{L_r^{\text{NAD}}}{L_r^{\text{NADP}}} \right) \right|.$$

Phenomenological Modeling of Proteome Cost for Multicoenzyme Systems. A detailed derivation of the phenomenological model used to analyze protein cost for multicoenzyme systems in Fig. 4 is provided in *SI Appendix, Supporting Information Text*. Here we present the final model used in the main text. The following nonlinear optimization problem was constructed to find the minimal protein cost by varying coenzyme concentration ratios and flux partitioning, such that

$$\underset{v_{r\alpha},\, v_e,\, \Gamma_\alpha}{\text{maximize}} \quad \sum_{r\alpha} \frac{v_{r\alpha}}{\kappa_r} \left(\frac{K^{\dagger}_r + \Gamma_\alpha^{S_{r\alpha}}}{K^{\dagger}_r - \Gamma_\alpha^{S_{r\alpha}}} \right) + \sum_e \frac{v_e}{\kappa_e} \left(\frac{\Gamma_\beta + \Gamma_\alpha}{\Gamma_\beta - \Gamma_\alpha} \right)$$
$$\text{subject to :} \quad \sum_\alpha v_{r\alpha} = v_r \quad \forall\, r,$$
$$\sum_r s_{r\alpha} v_{r\alpha} + \sum_e s_{e\alpha} v_e = 0 \quad \forall\, \alpha$$
$$\Gamma_\alpha \leq \Gamma_\beta \quad \forall\, \alpha, \beta$$

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 is the ratio of charged to uncharged species for coenzyme pair α, and $v_{r\alpha}$ 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 subnetwork plus the cost of all generalized exchange reactions (see *SI Appendix, Supporting Information Text* for derivation). All enzyme cost simulations were performed in MATLAB R2021a, using the COBRA toolbox (48) and GLPK solver for linear programs.

Data Availability. Code and data have been deposited in GitHub (https://github.com/jgoldford/coenzymes). All other study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We thank Adrian Jinich, Igor Libourel, Pankaj Mehta, and Gina Mawla for helpful discussions, and wish to acknowledge inspiring conversations with Arren Bar-Even, who sadly passed away prematurely. We acknowledge the support provided by the Directorates for Biological Sciences (BIO) and Geosciences (GEO) at the NSF and NASA under agreements 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 a Physics of Living Systems Fellow through grant GBMF4513. D.S. also acknowledges funding from the US Department of Energy, Office of Science, Office of Biological & Environmental Research through the Microbial Community Analysis and Functional Evaluation in Soils Science Focus Area Program (m-CAFEs) under contract number DE-AC02-05CH11231 to Lawrence Berkeley National Laboratory, from the NSF Center for Chemical Currencies of a Microbial Planet (C-CoMP, publication #005), and from the NIH National Institute on Aging, award number UH2AG064704.

Author affiliations: [a]Physics of Living Systems, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]Blue Marble Space Institute of Science, Seattle, WA, 98154; [c]Bioinformatics Program, Biological Design Center, Boston University, Boston, MA, 02215; [d]Department of Physics, Boston University, Boston, MA, 02215; [e]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [f]Department of Plant Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [g]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA; [h]Department of Biomedical Engineering, Boston University, Boston, MA, 02215; and [i]Department of Biology, Boston University, Boston, MA, 02215

1. D. L. Nelson, M. M. Cox, *Lehninger Principles of Biochemistry* (W. H. Freeman, ed. 4, 2005).
2. L. Agledal, M. Niere, M. Ziegler, The phosphate makes a difference: Cellular functions of NADP. *Redox Rep.* **15**, 2–10 (2010).
3. B. D. Bennett *et al.*, Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat. Chem. Biol.* **5**, 593–599 (2009).
4. C. S. Henry, L. J. Broadbelt, V. Hatzimanikatis, Thermodynamics-based metabolic flux analysis. *Biophys. J.* **92**, 1792–1805 (2007).
5. 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).
6. 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).
7. A. 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).
8. J. O. Park *et al.*, Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat. Chem. Biol.* **12**, 482–489 (2016).
9. 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).
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. R. Chen, A. Greer, A. M. Dean, Redesigning secondary structure to invert coenzyme specificity in isopropylmalate dehydrogenase. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12171–12176 (1996).
13. M. Lunzer, S. P. Miller, R. Felsheim, A. M. Dean, The biochemical architecture of an ancient adaptive landscape. *Science* **310**, 499–501 (2005).
14. G. Zhu, G. B. Golding, A. M. Dean, The selective cause of an ancient adaptation. *Science* **307**, 1279–1282 (2005).
15. E. Armingol, E. Tobar, R. Cabrera, Understanding the impact of the cofactor swapping of isocitrate dehydrogenase over the growth phenotype of Escherichia coli on acetate by using constraint-based modeling. *PLoS One* **13**, e0196182 (2018).
16. N. S. Scrutton, A. Berry, R. N. Perham, Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature* **343**, 38–43 (1990).
17. 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).
18. L. Calzadiaz-Ramirez *et al.*, *In vivo* selection for formate dehydrogenases with high efficiency and specificity toward NADP. *ACS Catal.* **10**, 7512–7525 (2020).
19. M. Bouzon *et al.*, Change in cofactor specificity of oxidoreductases by adaptive evolution of an *Escherichia coli* NADPH-auxotrophic strain. *MBio* **12**, e0032921 (2021).
20. 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).
21. D. Heckmann *et al.*, Predicting C4 photosynthesis evolution: Modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell* **153**, 1579–1588 (2013).
22. S.-R. Hosseini, O. C. Martin, A. Wagner, Phenotypic innovation through recombination in genome-scale metabolic networks. *Proc. Biol. Sci.* **283**, 20161536 (2016).
23. J. E. Goldford, H. Hartman, R. Marsland III, D. Segrè, Environmental boundary conditions for the origin of life converge to an organo-sulfur metabolism. *Nat. Ecol. Evol.* **3**, 1715–1724 (2019).
24. J. E. Goldford, D. Segrè, Modern views of ancient metabolic networks. *Curr. Opin. Syst. Biol.* **8**, 117–124 (2018).
25. A. Jinich *et al.*, A thermodynamic atlas of carbon redox chemical space. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 32910–32918 (2020).
26. R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, U. Sauer, Multidimensional optimality of microbial metabolism. *Science* **336**, 601–604 (2012).
27. A. Jinich *et al.*, Quantum chemistry reveals thermodynamic principles of redox biochemistry. *PLOS Comput. Biol.* **14**, e1006471 (2018).
28. J. D. Orth, I. Thiele, B. Ø. Palsson, What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
29. J. D. Orth *et al.*, A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. *Mol. Syst. Biol.* **7**, 535 (2011).
30. 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).
31. 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).
32. E. Noor *et al.*, Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLOS Comput. Biol.* **10**, e1003483 (2014).
33. E. Noor *et al.*, The protein cost of metabolic fluxes: Prediction from enzymatic rate laws and cost minimization. *PLOS Comput. Biol.* **12**, e1005167 (2016).
34. D. A. Beard, H. Qian, Relationship between thermodynamic driving force and one-way fluxes in reversible processes. *PLoS One* **2**, e144 (2007).
35. A. Flamholz, E. Noor, A. Bar-Even, R. Milo, eQuilibrator–The biochemical thermodynamics calculator. *Nucleic Acids Res.* **40**, D770–D775 (2012).
36. M. E. Beber *et al.*, eQuilibrator 3.0: A database solution for thermodynamic constant estimation. *Nucleic Acids Res.* **50**, D603–D609 (2022).
37. 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* **82**, 1819–1828 (2014).
38. 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).
39. A. D. Goldman, B. Kacar, Cofactors are remnants of life's origin and early evolution. *J. Mol. Evol.* **89**, 127–133 (2021).
40. J. W. Thornton, Resurrecting ancient genes: Experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366–375 (2004).
41. D. J. 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. R. R. da Silva, P. C. Dorrestein, R. A. Quinn, Illuminating the dark matter in metabolomics. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12549–12550 (2015).
44. A. J. M. Ribeiro *et al.*, Mechanism and catalytic site atlas (M-CSA): A database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res.* **46** (D1), D618–D623 (2018).
45. S. R. Eddy, Accelerated profile HMM searches. *PLOS Comput. Biol.* **7**, e1002195 (2011).
46. Z. A. King *et al.*, BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **44**, D515–22(2015).
47. 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).
48. J. Schellenberger *et al.*, Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nat. Protoc.* **6**, 1290–1307 (2011).
49. P. Salvy *et al.*, pyTFA and matTFA: A Python package and a Matlab toolbox for thermodynamics-based flux analysis. *Bioinformatics* **35**, 167–169 (2019).
50. 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).
51. M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).