Published in final edited form as:

ACS Synth Biol. 2017 February 17; 6(2): 326–333. doi:10.1021/acssynbio.6b00188.

A general tool for engineering the NAD/NADP cofactor preference of oxidoreductases

Jackson KB Cahn[†,§], **Caroline A Werlang**[†,⊥], **Armin Baumschlager**[†,‖], **Sabine Brinkmann-Chen**[†], **Stephen L Mayo**[‡], **Frances H Arnold**[†]

[†]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, 91125.

[‡]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, 91125.

[§]Current Address: Institute for Microbiology, ETH Zürich, CH-8093 Zürich. Switzerland.

[⊥]Current Address: Institute of Bioengineering, EPFL, CH-1015 Lausanne, Switzerland.

[‖]Current Address: Department of Biosystems Science and Engineering, ETH Zürich, CH-4058 Basel, Switzerland.

Abstract

The ability to control enzymatic nicotinamide cofactor utilization is critical for engineering efficient metabolic pathways. However, the complex interactions that determine cofactor-binding preference render this engineering particularly challenging. Physics-based models have been insufficiently accurate and blind directed evolution methods too inefficient to be widely adopted. Building on a comprehensive survey of previous studies and our own prior engineering successes, we present a structure-guided, semi-rational strategy for reversing enzymatic nicotinamide cofactor specificity. This heuristic-based approach leverages the diversity and sensitivity of catalytically productive cofactor binding geometries to limit the problem to an experimentally tractable scale. We demonstrate the efficacy of this strategy by inverting the cofactor specificity of four structurally diverse NADP-dependent enzymes: glyoxylate reductase, cinnamyl alcohol dehydrogenase, xylose reductase, and iron-containing alcohol dehydrogenase. The analytical components of this approach have been fully automated and are available in the form of an easy-to-use web tool: Cofactor Specificity Reversal – Structural Analysis and LibrAry Design (CSR-SALAD).

AUTHOR CONTRIBUTIONS

This project was conceived by J.K.B.C., S.B-C., S.L.M., and F.H.A. J.K.B.C. designed and programmed CSR-SALAD, which was implemented online by C.A.W. J.K.B.C. performed the validation experiments with assistance from A.B., and produced the figures and tables. J.K.B.C., S.B-C., and F.H.A. wrote the manuscript with input from all authors.

SUPPORTING INFORMATION

The Supporting Information is available free of charge on the ACS Publications website at DOI [TBD].

• Previous cofactor specificity reversals in the literature; residue distributions of natural NAD-and NADP-bound structures; examples of structural classifications; CSR-SALAD interface examples; detailed kinetic parameters; wild-type enzyme structures and modeled mutants; detailed results of literature mutant recapitulation

• CSR-SALAD Users' Manual Version 1.1; additional references; designed libraries for proteins described in this work.

Keywords

cofactor specificity; oxidoreductases; protein engineering; library design; semi-rational engineering

The vast majority of oxidoreductases, which make up the largest group of enzymes in the Enzyme Commission nomenclature, use the functionally-equivalent cofactors nicotinamide adenine dinucleotide (NAD) or nicotinamide adenine dinucleotide phosphate (NADP) for the transport and storage of electrons in the form of hydride groups (Figure 1). The phosphate group that distinguishes these cofactors is sufficiently distant spatially and covalently from the chemically-active nicotinamide moiety that it plays no role in the chemistry. Nonetheless, nearly all enzymes that use these cofactors exhibit a strong preference for one or the other of them. This specificity enables cells to regulate different classes of enzymes and pathways separately, prevent futile reaction cycles, and maintain chemical driving forces by controlling the availability of the oxidized and reduced forms of NAD(P).

Interest in nicotinamide cofactor specificity has been driven not only by scientific curiosity but also by its importance to the engineering of cellular metabolism[1]. Several studies have shown how balancing cofactor availability can increase pathway yields by removing carbon inefficiencies and side products, eliminating oxygen requirements, or improving steady-state metabolite levels[2–8]. For this reason, switching enzyme cofactor preference has been a frequent target of protein engineering efforts ever since the first report of engineered specificity reversal in 1990 by Scrutton and co-workers[9]. Supporting Tables 1 and 2 list more than 25 examples of cofactor switching in each direction (NADP-to-NAD and *vice versa*), updated from that of Khoury *et al*[10]. However, a closer examination of these studies shows that cofactor specificity reversal remains an unsolved problem: many of these efforts were only marginally successful in achieving reversed specificity without compromising catalytic activity.

A number of factors combine to make reversing enzyme cofactor specificity a challenging task. Although the protein features and interactions which make up (phospho-)adenosine binding pockets are distal from the catalytic sites of the enzymes, they have been shown to have an outsize influence on enzyme activity. Subtle chemical changes to the cofactor can have a dramatic effect on activity[11], and mutations to the adenosine-interacting part of the cofactor binding pocket can affect reaction kinetics[12] and even substrate specificity[13]. Combined with the dynamic nature of cofactor binding[14–17], this sensitivity to structural perturbation has proven a major obstacle to rational and computational design approaches[10].

These same factors have also been a hurdle to homology-guided approaches[16,18], as has the structural diversity of cofactor binding and specificity motifs[19]. Despite the sensitivity of catalytic activity to cofactor binding pose, evolution has produced a diverse array of structural motifs for binding NAD(P) and for conferring specificity for NAD or NADP[15,19–23]. NAD(P) utilization has been documented in proteins with the canonical Rossmann fold as well as many others, including TIM-barrel, dihydroquinoate synthase-like, and FAD/NAD-binding folds. Considerable structural diversity may also exist within a given

fold and even within an enzyme family[19]. Furthermore, many enzyme families contain representatives with both specificities[24–27], suggesting that specificity reversal is relatively facile in natural evolution. In a previous study, we identified at least seven independent cases where cofactor specificity had switched to NAD-utilization in the evolutionary history of the otherwise exclusively NADP-preferring ketol-acid reductoisomerase enzyme (KARI) family. Notably, each of these switches was achieved through unique combinations of amino acid substitutions, insertions, and deletions[24]. Within the structural diversity of cofactor binding pockets, specificity for NAD or NADP appears to be dictated largely by the charge and polarity of the binding pocket. The negatively-charged phosphate of NADP is often coordinated by positively-charged residues, particularly arginine, and hydrogen-bond donating residues (Supporting Figures 1 and 2). In NAD-specific proteins, by contrast, negatively-charged residues often serve to repel the NADP phosphate and accept hydrogen bonds from the 2’ and 3’ ribose hydroxyls. More discrete recognition elements have been noted, such as an arginine in NADP-binding Rossmann folds that forms a cation-pi interaction with the adenine ring system, but none are either universal or deterministic[20].

Finally, random mutagenesis and screening has also proven of limited utility: several amino acids are nearly always involved in controlling specificity, and reversing specificity has required multiple simultaneous mutations[5], leading to an intractably large combinatorial space of mutations to explore. This problem is compounded by strong non-additivity in the effects of mutations[5,28,29], which renders uphill-walk modes of optimization ineffective. As a result, no single engineering approach has proven universally applicable, and the full promise of manipulating the NAD(P) cofactor utilization of enzymes and metabolic pathways has not been attained.

RESULTS AND DISCUSSION

Approach

We set out to develop a comprehensive and user-friendly approach to reversing the nicotinamide cofactor specificity of any NAD(P)-utilizing enzyme. Though we had previously developed a simple recipe for such a reversal for the enzymes of the KARI enzyme family[18], early tests showed that this recipe was not sufficient for the more general problem. The more general strategy, which was developed with a focus on ease-of-use for non-experts, is laid out in detail in Supporting Material 1. As with the recipe we developed for KARIs, it comprises three steps (Figure 2): enzyme structural analysis, design and screening of focused mutant libraries for reversing cofactor preference, and, finally, recovery of catalytic efficiency. The development of each step proceeded from a careful study of available literature on the structural biology, biophysics, and engineering of nicotinamide cofactor specificity followed by development of rational engineering heuristics. We then formalized these heuristics and implemented them in a computational framework: Cofactor Specificity Reversal: Structural Analysis and LibrAry Design (CSR-SALAD), which is freely available online at http://www.che.caltech.edu/groups/fha/CSRSALAD/index.html. Supporting Figure 4 shows the web interface of CSR-SALAD and an example of its output.

As might be intuitively expected, nearly all of the mutations previously required for cofactor specificity reversal are in the immediate vicinity of the 2’ moiety of the NAD/NADP

cofactor (Supporting Tables 1 and 2, column 5). While distal residues are occasionally mutated in cofactor-switched proteins, only rarely do they actually contribute to the reversal of specificity. Based on these past studies, we hypothesized that targeting a limited set of residues would be sufficient for cofactor switching. In CSR-SALAD, the specificity-determining residues are therefore defined as those that contact the 2' moiety directly, those in position to contact it through water-mediated interactions, or – specifically for NAD-to-NADP switching – those that can be mutated to contact the expanded 2' moiety of the NADP cofactor.

To assist with engineering, it is useful to obtain more information about a residue than simply whether it contacts the cofactor. Therefore, CSR-SALAD utilizes a system of classifications (described in Supporting Material 1 and illustrated in Supporting Figure 3) that we developed to describe a residue's role in forming the cofactor-binding pocket. This classification system is informed by that introduced in 1997 by Carugo and Argos[20] and is used in the library design process to discriminate among different sets of potential mutations. Examples of these classes include residues interacting with the face of the adenine ring system (what Carugo and Argos called S10), the edge of the rings (S8) or interacting with both the 2'-moiety and the 3'-hydroxyl (S9).

The next step is to design a library of mutations directed at the identified specificity-determining residues. To keep library sizes small and thus keep the screening for higher activity with the new cofactor experimentally tractable, CSR-SALAD predicts sub-saturation degenerate codon libraries[30–33], wherein specified mixtures of nucleotides are used to generate combinations of amino acids at each targeted position. So that library sizes can be tailored to the experimental capabilities of the user, CSR-SALAD possesses, for each residue in each structural class, a range of degenerate codons coding for different numbers of amino acids. The selection of these degenerate codons is guided primarily by the inclusion of mutations to structurally similar residues that have already been shown to be useful for cofactor specificity reversal in the desired direction. This selection is subjective and based on the accumulated lessons of prior studies by ourselves and others. We hope that application of the approach and of the CSR-SALAD software will result in feedback that will improve the algorithm as more data become available.

The final step is recovery of enzymatic activity. Cofactor-switched enzymes – in fact any enzyme having multiple mutations – often suffer a significant loss of activity[18,34,35], and compensatory mutations must be identified to recover it. These mutations re-stabilize or re-activate the protein for catalysis with the new cofactor, reversing whatever detrimental side effects the specificity-reversing mutations produced. Because these mutations are often remote from the cofactor-switching mutations, in our previous work on KARIs we recommended using random mutagenesis and screening to discover them[18]. While highly effective, the large library sizes for random mutagenesis render this approach time-consuming and equipment- and labor-intensive. Here, we have taken the unprecedented step of using structural information to predict positions in the amino acid sequence with dramatically increased probabilities of harboring compensatory mutations. This allows the production of highly active enzymes from the screening of just a handful of single-site saturation libraries and combination of the most beneficial mutations. We

have identified several types of activity recovery positions based on common features of previous engineering efforts, as outlined in Supporting Material 1. The most effective have consistently been mutations around the adenine ring, and all of the mutations used in our experimental validation fall into this category. In a previous study we demonstrated the power of these mutations to boost the *in vitro* and *in vivo* activities of a number of NAD(P)-dependent enzymes[12], and these effects are even more dramatic in cofactor-switched enzymes.

Experimental Validation

To demonstrate the general utility of CSR-SALAD for finding cofactor-switching mutations, we selected a test set of four oxidoreductases drawn from those previously studied in our lab and enzymes of industrial interest: *Arabidopsis thaliana* glyoxylate reductase (GR), *Saccharomyces cerevisae* cinnamyl alcohol dehydrogenase (CinADH), *Talaromyces emersonii* xylose reductase (XR), and *Thermotoga maritima* iron-containing alcohol dehydrogenase (FeADH) (Table 1). Although CSR-SALAD is capable of processing NAD- and NADP-bound structures, we focused only on NADP-preferring enzymes because switching in this direction is more relevant for applications[36–38] and has also proved to be more difficult to achieve in the past (Supporting Tables 1 and 2). The selected enzymes include the most common NAD(P)-binding fold, the Rossmann fold[39], as well as other, less common cofactor binding folds (TIM barrel and dihydroquinoate synthase-like folds). To demonstrate the robustness of CSR-SALAD analysis, we also included one protein (XR) without a crystal structure, but for which the structure of a homologous enzyme with 55% sequence identity was known, and one enzyme (GR) crystallized in the absence of cofactor, for which the cofactor-binding pose could be modeled. We also included proteins with strict (XR) and loose (FeADH) natural cofactor specificity. For each enzyme, we generated the mutant library recommended using CSR-SALAD default parameters and screened it for cofactor-switched variants. The best variants – those which had the greatest activity toward NADH provided that activity was greater than that toward NADPH – were then subjected to between zero and two rounds of single-site saturation mutagenesis for activity recovery, and the best final variants were purified and characterized for Michaelis-Menten kinetics (Table 1 and Supporting Table 3).This procedure led to successful cofactor specificity reversal for all four enzymes, and for three of the four test enzymes the catalytic efficiency was recovered back to at least 10% of wild type. Indeed, by this metric these represent three of the most successful NADP-to-NAD cofactor switches ever reported (Supporting Table 1). For XR, which we were also able to switch, activity recovery was less successful, but this enzyme showed the most dramatic reversal of cofactor specificity, nearly 5,000-fold, due to the extremely low initial activity on NADH. We suspect that, should greater activity be required, further compensatory mutations could be found by directed evolution[18].

Cofactor specificity was reversed in all four enzymes by a concerted set of three binding-pocket mutations, although for several enzymes more than three positions were probed in the libraries designed by CSR-SALAD (Supporting Material 3). Further catalytic activity was recovered in three of the four enzymes by single mutations in the adenine-binding pocket; FeADH had sufficient activity after specificity reversal that no additional mutations were warranted. Notably, the final sets of cofactor-switching mutations identified for each

enzyme show no significant overlap. Indeed, the only mutation that appears twice, serine-to-glycine in XR and FeADH, occurs at residues that are in strikingly different regions of their respective binding pockets. While some simple patterns emerge, broader engineering principles are not readily apparent. For instance, while each set of mutations contains one or more mutations to a negatively-charged residue, these mutations come from residues which play different structural roles in the wild-type binding of the cofactors (Supporting Figure 5). The rest of the mutations introduce amino acids whose role in cofactor specificity is less immediately evident. Even though the precise determinants of cofactor specificity are subtle and often difficult to predict, the successful reversal of cofactor specificity by CSR-SALAD demonstrates that active proteins can be selected from small libraries containing residues that are primed to alter specificity.

Comparison to Previous Studies

Because the components of the approach described here were developed based on a synthesis of the previous literature, we sought to determine whether modestly-sized CSR-SALAD libraries could also recapitulate cofactor-switching mutations identified in previous experiments. Indeed, for the majority of the previously published enzymes with a reversal of specificity and final activity at least 10% of initial activity, the CSR-SALAD libraries contained the precise combination of mutations used (not including distal compensatory mutations) (Table 2). Given that there may be multiple sets of mutations that reverse cofactor specificity for a given enzyme[5,40–42], it seems reasonable to hypothesize that CSR-SALAD libraries would have been able to successfully switch enzymes where the precise combination of mutations was not recapitulated, particularly because the same sites are often targeted for mutagenesis (full data on mutant recapitulation can be found in Supporting Tables 4 and 5). Even without these, the ability to recapitulate known mutation combinations means that this approach has been validated not merely on the four structurally diverse proteins described above, but has been shown capable of producing switched and active enzymes in 22 different proteins.

As mentioned above, we did not experimentally target NAD-preferring proteins in this study because these enzymes tend to be easier to switch. This tendency is likely a result of the significant potential binding energy that the phosphate provides. While the recapitulation of the mutations used in nearly a dozen of the best prior engineered switches in this direction provides strong evidence that the approach described here is capable of providing guidance for both classes of switching efforts, there may exist structures of NAD-binding pockets with unforeseen challenges that require modifications to the library design algorithm. Another potential challenge comes from enzymes that use NAD(P) in the context of multi-step electron transfer pathways, such as mono- and dioxygenases. Previous attempts to switch these enzymes' cofactor preference have resulted in nicotinamide cofactor oxidation uncoupled from substrate turnover[43,44], and preliminary CSR-SALAD-based engineering of a Baeyer-Villiger monooxygenase encountered this same hurdle (data not shown). We propose that the more complex electron transfers of these enzymes are more susceptible to lethal perturbation, and that reversing specificity in these enzymes may require simultaneous monitoring of uncoupled activity and likely directed evolution to recover coupling of cofactor utilization to product formation.

It is also worth noting that the libraries designed by CSR-SALAD are inspired by previous laboratory results more than by natural patterns of specificity. Hence, whereas it recapitulates protein engineering studies well, CSR-SALAD does not necessarily replicate the natural patterns of cofactor specificity. For example, the two known naturally NAD-preferring xylose reductases[45,46] possess 0 and 1 of the amino acids used here to alter the specificity of *T. emersonii* XR. Although protein engineers typically restrict themselves to amino acid substitutions, natural evolution often involves insertions and deletions[47], as observed in the KARI enzyme family[19]. No simple methods exist for making random in-frame insertion/deletion mutants, and the results of insertions and deletions can be challenging to predict, as loop geometries may change significantly. However, CSR-SALAD's ability to find novel engineering solutions based solely on substitutions supports the basic hypothesis that multiple cofactor-switching solutions exist for a given protein.

Conclusion and Outlook

Protein engineers generally approach engineering tasks in a detail-oriented fashion and on case-by-case basis. This is an 'artisanal' approach that requires considerable expertise. Here, we have taken on an important and challenging engineering problem and developed a general approach that reliably produces the desired enzyme properties with a minimum of expertise and labor. As protein engineering continues to mature as a discipline, we foresee the development of a number of similar tools to make enzyme optimization more accessible to users in the broader fields of metabolic and biological engineering.

In developing this method, the diversity of cofactor binding motifs in nature inspired us to hypothesize that multiple mutational solutions exist for each specificity reversal problem. Combined with the ability to recover activity without affecting specificity, we believed that this would allow for at least one satisfactory enzyme to be found in a focused pool of mutants. Careful consideration of the previous literature on nicotinamide cofactor specificity allowed us to develop heuristics that limit the mutational space in this semi-rational approach. Formalization into the CSR-SALAD computational framework makes this approach broadly accessible. Nicotinamide cofactor specificity reversal was a particularly inviting problem for this approach, as we were able to take advantage of the extensive literature and a facile high-throughput screen based on the unique spectral signature of reduced nicotinamide. It remains to be seen whether similar structure-based and knowledge-guided semi-rational approaches can be developed for other protein engineering problems.

We foresee that CSR-SALAD will be useful in the fields of synthetic biology, metabolic engineering, and biocatalysis and will make the reversal of cofactor specificity a routine task rather than a formidable engineering endeavor.

METHODS

CSR-SALAD development

CSR-SALAD was built in Python, relying heavily on the PDB module of the Biopython package[48]. The analysis component was tested on a representative set of

499 NAD-bound structures and 463 NADP-bound structures selected based on sequence identity and resolution. Degenerate codon selection was assisted by the LibDesign[31] and AA-Calculator[49] tools and optimized using the Ambiguous Nucleotide Tool (ANT) framework[32]. CSR-SALAD has been implemented as a PHP-based web-server that is available at http://www.che.caltech.edu/groups/fha/CSRSALAD/index.html. Further documentation of CSR-SALAD's internal workings and guidelines for use can be found on the website or in the Supporting Material 1.

Structure analysis

Homology models were created using the SWISS-MODEL server[50]. For proteins crystallized without bound cofactor, cofactor was manually placed by alignment with a homologous protein, followed by manual side-chain rotamer adjustments in PyMol[51].

Cloning and library creation

Genes were either obtained from Integrated DNA Technologies (IDT) as gBlock linear fragments or cloned from pre-existing vectors and were cloned into pET22b(+) in frame with the C-terminal His_6-tag for expression in *Escherichia coli* BL21(DE3) 'E. Cloni' (Lucigen) using Gibson cloning[52] with overlap at the T7 promoter and terminator sequences. Mutagenic primers for site-saturation mutagenesis were obtained from IDT and treated as suggested by IDT protocols. Libraries were generated using a modified version of the QuikChange method (Stratagene) as described previously[53]. Some of the libraries screened were based on early versions of the CSR-SALAD algorithm, but all reported mutants are found in the CSR-SALAD libraries generated using default parameters.

Following transformation with library DNA, single colonies were picked with sterile toothpicks and inoculated into 300 μL of Luria broth supplemented with 100 μg/mL ampicillin (LB-Amp) in shallow-well 96-well plates. Following overnight growth at 37 °C with shaking at 225 rpm and 80% humidity, 50 μL of the pre-cultures were added to 600 μL of fresh LB-Amp media in deep-well 96-well plates and grown for 3 h at 37 °C. Then, 50 μL of additional LB-Amp containing 0.25 mM isopropyl thiogalactopyranoside (IPTG) were added and expression was continued at a reduced temperature. For expression temperatures and times, see Table 3. The expression cultures were harvested through centrifugation, and the plates containing cell pellets were stored at −20 °C until screening.

Library screening

Assay procedures varied depending on the protein. The following is the general protocol, but deviations for specific proteins are listed in Table 3. *E. coli* cells were resuspended in the appropriate lysis buffer (Table 3) containing 750 mg/L lysozyme, 10 mg/L DNase I, and 2 mM $MgCl_2$. Lysis was accomplished at 37 °C for 1 h. Enzyme activities were then assayed by monitoring NAD(P)H consumption in the presence of the substrate molecule at 340 nm in a plate reader.

Enzyme expression, purification, and kinetic measurements

For larger-scale expression, 5–50 mL pre-cultures were grown overnight and diluted 1:250 into fresh medium for expression. After expression cultures reached an OD_{600} of 0.6–0.9,

they were induced with IPTG to a final concentration of 0.5 mM. Expression then proceeded for the same times and temperatures used for library expression. Expression cultures were harvested by centrifugation, the supernatant was discarded, and the pellets were frozen at −20 °C.

For purification, cell pellets were resuspended in 10–20 mL buffer A (25 mM Tris, 100 mM NaCl, 20 mM imidazole, pH 7.4) and lysed by sonication or using BugBuster protein extraction reagent (EMD Millipore). The lysate was clarified by centrifugation, and the enzymes were purified via their C-terminal His_6-tag using High Performance (HP) Ni-NTA Sepharose columns (GE Healthcare, Waukesha, WI, USA) on an Äkta Xpress FPLC (GE Healthcare). The concentration of purified protein was determined using the Bradford assay (Bio-Rad, Hercules, CA, USA).

For rate measurements, k_{cat} values were determined using the same assay conditions as above with saturating cofactor and substrate. Michaelis–Menten constants were determined by varying cofactor concentration, and activity was monitored using fluorescence (excitation 340 nm, emission 440 nm) for improved sensitivity. At least six cofactor concentrations were used for these determinations, and all measurements were performed at least three times. MATLAB (MathWorks, Natwick, MA, USA) was used for parameter fitting.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by the Gordon and Betty Moore Foundation through grant number GBMF2809 to the Caltech Programmable Molecular Technology Initiative and by American Recovery and Reinvestment Act (ARRA) funds through the National Institutes of Health Shared Instrumentation Grant Program, grant number S10RR027203, to F.H.A. J.K.B.C. acknowledges the support of the Resnick Sustainability Institute (Caltech). The authors thank Ruchi Jahagirdar and Lisa Mears for experimental assistance and Tilman Flock for providing the list of nonredundant PDBs used for CSR-SALAD testing and validation. They also thank numerous former and current members of the Arnold and Mayo labs for invaluable suggestions and discussions.

REFERENCES

1. Wang Y, San K-Y and Bennett GN (2013) Cofactor engineering for advancing chemical biotechnology, Curr. Opin. Biotechnol. 24, 994–999. [PubMed: 23611567]
2. Matsushika A, Watanabe S, Kodaki T, Makino K, Inoue H, Murakami K, Takimura O, and Sawayama S (2008) Expression of protein engineered $NADP^+$-dependent xylitol dehydrogenase increases ethanol production from xylose in recombinant *Saccharomyces cerevisiae*, Appl. Microbiol. Biotechnol. 81, 243–255. [PubMed: 18751695]
3. Bengtsson O, Hahn-Hagerdal B and Gorwa-Grauslund MF (2009) Xylose reductase from *Pichia stipitis* with altered coenzyme preference improves ethanolic xylose fermentation by recombinant *Saccharomyces cerevisiae*, Biotechnol. Biofuels 2.
4. Hasegawa S, Uematsu K, Natsuma Y, Suda M, Hiraga K, Jojima T, Inui M, and Yukawa H (2012) Improvement of the redox balance increases L-valine production by *Corynebacterium glutamicum* under oxygen deprivation conditions, Appl. Environ. Microbiol. 78, 865–875. [PubMed: 22138982]
5. Bastian S, Liu X, Meyerowitz JT, Snow CD, Chen MMY, and Arnold FH (2011) Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in *Escherichia coli*, Metab. Eng. 13, 345–352. [PubMed: 21515217]

6. Tamakawa H, Ikushima S and Yoshida S (2011) Ethanol production from xylose by a recombinant *Candida utilis* strain expressing protein-engineered xylose reductase and xylitol dehydrogenase Biosci., Biotechnol., Biochem 75, 1994–2000. [PubMed: 21979076]
7. Shen CR, Lan EI, Dekishima Y, Baez A, Cho KM, and Liao JC (2011) Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*, Appl. Environ. Microbiol. 77, 2905–2915. [PubMed: 21398484]
8. Xiao Z Lv C, Gao C, Qin J, Ma C, Liu Z, Liu P, Li L, and Xu P (2010) A novel whole-cell biocatalyst with NAD^+ regeneration for production of chiral chemicals, PLoS ONE 5, e8860. [PubMed: 20126645]
9. Scrutton NS, Berry A and Perham RN (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering Nature 343, 38–43. [PubMed: 2296288]
10. Khoury GA Fazelinia H, Chin JW, Pantazes RJ, Cirino PC, and Maranas CD (2009) Computational design of *Candida boidinii* xylose reductase for altered cofactor specificity, Prot. Sci. 18, 2125–2138.
11. Mesecar AD, Stoddard BL and Koshland DE Jr. (1997) Orbital steering in the catalytic power of enzymes: Small structural changes with large catalytic consequences, Science 277, 202–206. [PubMed: 9211842]
12. Cahn JKB, Baumschlager A, Brinkmann-Chen S and Arnold FH (2016) Mutations in adenine-binding pockets enhance catalytic properties of NAD(P)H-dependent enzymes, Prot. Eng., Des. Sel 29, 31–38.
13. Maddock DJ, Patrick WM and Gerth ML (2015) Substitutions at the cofactor phosphate-binding site of a clostridial alcohol dehydrogenase lead to unexpected changes in substrate specificity, Prot. Eng., Des. Sel 28, 251–258.
14. Naylor CE, Gover S, Basak AK, Cosgrove MS, Levy HR, and Adams MJ (2001) NADP(+) and NAD(+) binding to the dual coenzyme specific enzyme *Leuconostoc mesenteroides* glucose 6-phosphate dehydrogenase: Different interdomain hinge angles are seen in different binary and ternary complexes, Acta Cryst. D 57, 635–648. [PubMed: 11320304]
15. Kavanagh KL, Klimacek M, Nidetzky B and Wilson DK (2003) Structure of xylose reductase bound to NAD(+) and the basis for single and dual co-substrate specificity in family 2 aldo-keto reductases, Biochem. J. 373, 319–326. [PubMed: 12733986]
16. Campbell E, Chuang S & Banta S (2013) Modular exchange of substrate-binding loops alters both substrate and cofactor specificity in a member of the aldo-keto reductase superfamily, Prot. Eng., Des. Sel. 26, 181–186.
17. Plapp BV (2010) Conformational changes and catalysis by alcohol dehydrogenase, Arch. Biochem. Biophys. 493, 3–12. [PubMed: 19583966]
18. Brinkmann-Chen S Flock T, Cahn JKB, Snow CD, Brustad EM, McIntosh JA, Meinhold P, Zhang L, and Arnold FH (2013) General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH, Proc. Nat. Acad. Sci. USA 110, 10946–10951. [PubMed: 23776225]
19. Cahn JKB, Brinkmann-Chen S, Spatzal T, Wiig JA, Buller AR, Einsle O, Hu Y, Ribbe MW, and Arnold FH (2015) Cofactor specificity motifs and the induced fit mechanism in class I ketol-acid reductoisomerases, Biochem. J. 468, 475–484. [PubMed: 25849365]
20. Carugo O and Argos P (1997) NADP-dependent enzymes. I: Conserved stereochemistry of cofactor binding, Proteins: Struct., Funct., Bioinf 28, 10–28.
21. Moodie SL, Mitchell JB and Thornton JM (1996) Protein recognition of adenylate: An example of a fuzzy recognition template, J. Mol. Biol. 263, 486–500. [PubMed: 8918603]
22. Denessiouk KA, Rantanen VV and Johnson MS (2001) Adenine recognition: a motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins, Proteins 44, 282–291. [PubMed: 11455601]
23. Wallen JR, Paige C, Mallett TC, Karplus PA and Claiborne A (2008) Pyridine nucleotide complexes with *Bacillus anthracis* coenzyme A-disulfide reductase: A structural analysis of dual NAD(P)H specificity, Biochemistry 47, 5182–5193. [PubMed: 18399646]
24. Brinkmann-Chen S, Cahn JKB and Arnold FH (2014) Uncovering rare NADH-preferring ketol-acid reductoisomerases, Metab. Eng. 26, 17–22. [PubMed: 25172159]

25. Di Luccio E, Elling RA and Wilson DK (2006) Identification of a novel NADH-specific aldo-keto reductase using sequence and structural homologies, Biochem. J. 400, 105–114. [PubMed: 16813561]
26. Hatrongjit R and Packdibamrung K (2010) A novel NADP(+)-dependent formate dehydrogenase from *Burkholderia stabilis* 15516: Screening, purification and characterization, Enzyme Microb. Technol. 46, 557–561.
27. Jensen CN, Ali ST, Allen MJ and Grogan G (2014) Exploring nicotinamide cofactor promiscuity in NAD(P)H-dependent flavin containing monooxygenases (FMOs) using natural variation within the phosphate binding loop. Structure and activity of FMOs from *Cellvibrio* sp. BR and *Pseudomonas stutzeri* NF13 J. Mol. Cat. B 109, 191–198.
28. Rodriguez-Arnedo A, Camacho M, Llorca F and Bonete MJ (2005) Complete reversal of coenzyme specificity of isocitrate dehydrogenase from *Haloferax volcanii*, Protein J. 24, 259–266. [PubMed: 16284723]
29. Rosell A, Valencia E, Ochoa WF, Fita I, Pares X, and Farres J (2003) Complete reversal of coenzyme specificity by concerted mutation of three consecutive residues in alcohol dehydrogenase, J. Biol. Chem. 278, 40573–40580. [PubMed: 12902331]
30. Reetz MT and Wu S (2008) Greatly reduced amino acid alphabets in directed evolution: Making the right choice for saturation mutagenesis at homologous enzyme positions, Chem. Comm 5499–5501. [PubMed: 18997931]
31. Mena MA and Daugherty PS (2005) Automated design of degenerate codon libraries, Prot. Eng., Des. Sel. 18, 559–561.
32. Engqvist MKM and Nielsen J (2015) ANT: Software for generating and evaluating degenerate codons for natural and expanded genetic codes, ACS Syn. Biol. 4, 935–938.
33. Cornish-Bowden A (1985) Nomenclature for incompletely specified bases in nucleic acid sequences: Reccomendations 1984, Nucleic Acids Res. 13, 3021–3030. [PubMed: 2582368]
34. Fasan R, Meharenna YT, Snow CD, Poulos TL and Arnold FH (2008) Evolutionary history of a specialized P450 propane monooxygenase, J. Mol. Biol. 383, 1069–1080. [PubMed: 18619466]
35. Bloom JD, Arnold FH and Wilke CO (2007) Breaking proteins with mutations: Threads and thresholds in evolution, Mol. Syst. Biol. 3, 76. [PubMed: 17262035]
36. Woodyer R, van der Donk WA & Zhao HM (2003) Relaxing the nicotinamide cofactor specificity of phosphite dehydrogenase by rational design, Biochemistry 42, 11604–11614. [PubMed: 14529270]
37. Verho R, Londesborough J, Penttila M and Richard P (2003) Engineering redox cofactor regeneration for improved pentose fermentation in *Saccharomyces cerevisiae*, Appl. Environ. Microbiol. 69, 5892–5897. [PubMed: 14532041]
38. Huffer S, Roche CM, Blanch HW and Clark DS (2012) *Escherichia coli* for biofuel production: Bridging the gap from promise to practice, Trends Biotechnol. 30, 538–545. [PubMed: 22921756]
39. Rossmann MG, Moras D and Olsen KW (1974) Chemical and biological evolution of nucleotide-binding protein Nature 250, 194–199. [PubMed: 4368490]
40. Rane MJ and Calvo KC (1997) Reversal of the nucleotide specificity of ketol acid reductoisomerase by site-directed mutagenesis identifies the NADPH binding site, Arch. Biochem. Biophys. 338, 83–89. [PubMed: 9015391]
41. Liang L, Zhang JQ and Lin ZL (2007) Altering coenzyme specificity of *Pichia stipitis* xylose reductase by the semi-rational approach CASTing, Microb. Cell Fact. 6, 36. [PubMed: 18028553]
42. Zeng Q-K, Du H-L, Wang J-F, Wei D-Q, Wang X-N, Li Y-X, and Lin Y (2009) Reversal of coenzyme specificity and improvement of catalytic efficiency of Pichia stipitis xylose reductase by rational site-directed mutagenesis, Biotechnol. Lett. 31, 1025–1029. [PubMed: 19330484]
43. Maurer SC, Kühnel K, Kaysser LA, Eiben S, Schmid RD, and Urlacher VB (2005) Catalytic hydroxylation in biphasic systems using CYP102A1 mutants, Adv. Synth. Catal. 347, 1090–1098.
44. Fasan R, Crook NC, Peters MW, Meinhold P, Buelter T, Landwehr M, Cirino PC, and Arnold FH (2011) Improved product-per-glucose yields in P450-dependent propane biotransformations using engineered *Escherichia coli*, Biotechnol. Bioeng. 108, 500–510. [PubMed: 21246504]

45. Lee JK, Koo BS and Kim SY (2003) Cloning and characterization of the xyl1 gene, encoding an NADH-preferring xylose reductase from *Candida parapsilosis*, and its functional expression in *Candida tropicalis*, Appl. Environ. Microbiol. 69, 6179–6188. [PubMed: 14532079]

46. Cahn JKB (2016) Discovery of an NAD-preferring xylose reductase. In Engineering, Predicting, and Understanding Nicotinamide Cofactor Specificity pp 45–55, California Institute of Technology - Ph.D. Thesis.

47. Tóth-Petróczy Á and Tawfik DS (2013) Protein insertions and deletions enabled by neutral roaming in sequence space, Mol. Biol. Evol. 30, 761–771. [PubMed: 23315956]

48. Hamelryck T and Manderick B (2003) PDB file parser and structure class implemented in Python, Bioinformatics 19, 2308–2310. [PubMed: 14630660]

49. Firth AE and Patrick WM (2008) GLUE-IT and PEDEL-AA: New programmes for analyzing protein diversity in randomized libraries, Nucleic Acids Res. 36, W281–285. [PubMed: 18442989]

50. Arnold K, Bordoli L, Kopp J & Schwede T (2006) The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling, Bioinformatics 22, 195–201. [PubMed: 16301204]

51. Schrodinger LLC. (2010) The PyMOL Molecular Graphics System, Version 1.3r1.

52. Gibson DG (2011) Enzymatic assembly of overlapping DNA fragments, In Methods in Enzymology (Christopher V, Ed.), pp 349–361, Academic Press.

53. Engqvist MKM, McIsaac RS, Dollinger P, Flytzanis NC, Abrams M, Schor S, and Arnold FH (2015) Directed evolution of *Gloeobacter violaceus* rhodopsin spectral properties, J. Mol. Biol. 427, 205–220. [PubMed: 24979679]

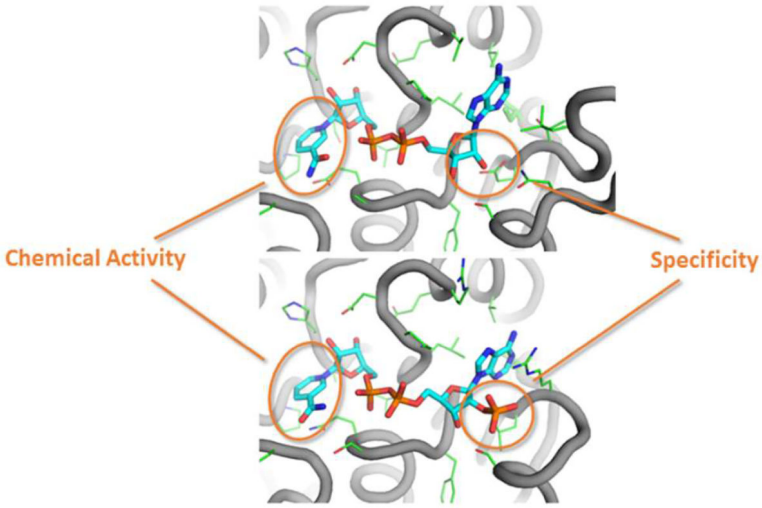

Figure 1.

Cofactors nicotinamide adenine dinucleotide (NAD, top) and nicotinamide adenine dinucleotide phosphate (NADP, bottom) in representative Rossmann fold binding pockets (PDBs 4XDY and 4TSK, respectively). The highlighted 2' recognition element (the phosphate of NADP or hydroxyl of NAD) and the chemically relevant hydride-bearing nicotinamide are separated in space and by multiple covalent bonds. In this paper, we use NAD and NADP, collectively NAD(P), when speaking generally about these cofactors, independent of their oxidation state, and only indicate the state, i.e. NADH / NAD+, when referring to those compounds specifically, such as for experimental details.

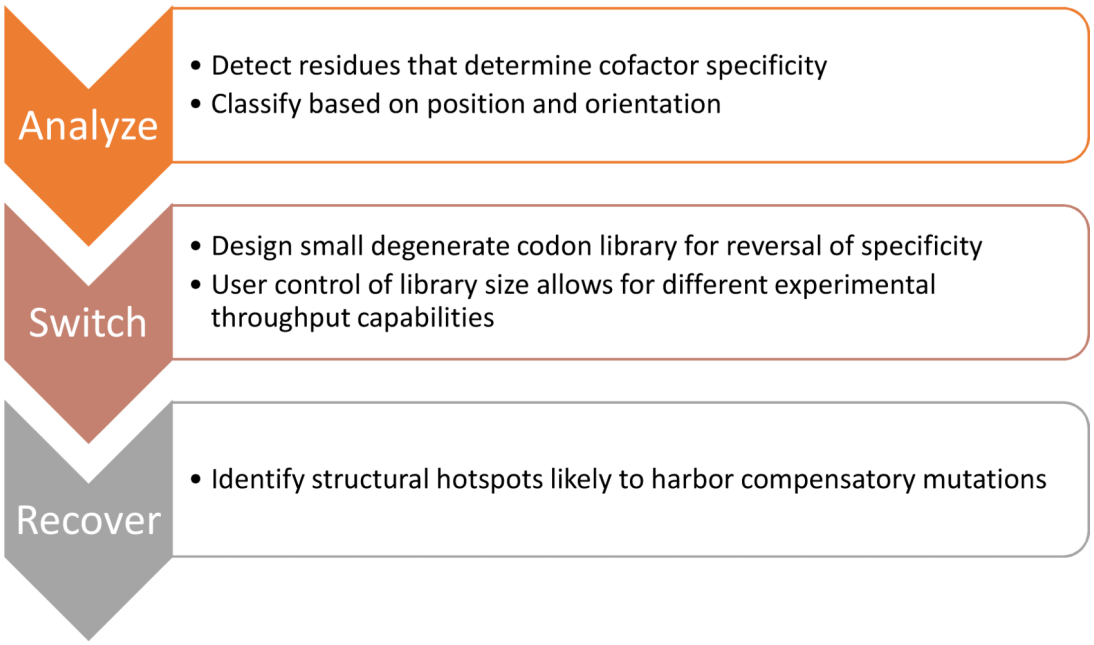

Figure 2.
CSR-SALAD performs three tasks: structure analysis, design of cofactor-switching libraries, and identification of positions for activity recovery.

Table 1.

NADP-preferring enzymes used for experimental validation of the CSR-SALAD method for switching cofactor preference to NAD while retaining catalytic activity. UniProt IDs, PDB accession codes, and Structural Classification of Proteins (SCOP) classification of their cofactor binding folds are listed for each. Mutations in bold are those used for switching cofactor specificity, those from recovery libraries are italicized. Fold changes from wild-type kinetics are provided, as is the log of the mutant's NADH k_{cat}/K_M divided by the wild-type NADPH k_{cat}/K_M. Specific values for wild-type and mutant enzymes are provided in **Supporting Table 3**.

					Fold change from wild type								
					NADH			NADPH					
Enzyme	**UniProt**	**RCSB PDB**	**Fold (SCOP)**	**Mutations**	k_{cat}	K_M	k_{cat}/K_M	k_{cat}	K_M	k_{cat}/K_M		**Specificity (NADH/ NADPH)[c]**	**log (NADH^{mut}/ NADPH^{wt})**
Arabidopsis thaliana glyoxylate reductase	Q9LSV0	3DOJ[a]	Rossmann (c.2.1.6)	**R31L, T32K, K35D,** *C68R*	2.4	0.9	2.7	0.4	4.6	0.1	33	−0.7	
Saccharomyces cerevisae cinnamyl alcohol dehydrogenase	Q04894	1PIW	Rossmann (c.2.1.1)	**S210D, R211P, K215E,** *S253P*	110	0.5	190	1.5	0.5	2.9	65	0.8	
Talaromyces emersonii xylose reductase	C5J3R6	1K8C[b]	TIM Barrel (c.1.7.1)	**S272G, N273D, R277Y,** *Q280E*	14	0.4	31	0.02	2.5	0.01	4900	−1.5	
Thermotoga maritima iron-containing alcohol dehydrogenase	Q9X022	1VHD	DHQS-like (e.22.1.2)	**G36E, S38N, S39G**	0.9	0.2	5.9	0.2	2.3	0.1	84	0.5	

[a] Structure 3DOJ has no cofactor, so the cofactor was introduced from structure 3PEF (*G. metallireducens* γ-hydroxybutyrate dehydrogenase) on the basis of a backbone alignment.

[b] *T. emersonii* xylose reductase (XR) has not been crystallized, so a homology model was generated from 1K8C, the structure of an XR from *C. tenuis*.

[c] Specificity is the ratio of NADH k_{cat}/K_M divided by NADPH k_{cat}/K_M for each enzyme. Presented here is fold change.

Table 2.

Recapitulation of mutations from 13 successful NADP-to-NAD and 17 successful NAD-to-NADP cofactor specificity reversals reported in the literature. Includes only mutants that had k_{cat}/K_M for the switched cofactor greater than the wild type's preferred cofactor and where mutant k_{cat}/K_M was at least 10% of wild type. This count excludes residues distal to the 2' binding pocket. For a full breakdown of mutant recapitulation, see **Supporting Tables 4** and **5**.

Percent mutations recapitulated	0%	1–49%	50–99%	100%
NADP-to-NAD	0	1	5	7
NAD-to-NADP	1	4	1	11

Table 3.

Details for expression, lysis, and assay conditions for enzymes used for validation of CSR-SALAD.

Enzyme	Expression Conditions	Lysis Conditions	Assay Conditions
***Arabidopsis thaliana* glyoxylate reductase**	Luria broth 24 °C, 18 h	100 mM hydroxyethyl-piperazineethanesulfonic acid (HEPES) pH 7.8	100 mM HEPES pH 7.8 1 mM sodium glyoxylate
***Saccharomyces cerevisae* cinnamyl alcohol dehydrogenase**	Luria broth 20 °C, 18 h	33 mM sodium phosphate pH 7 10% 10x BugBuster	33 mM sodium phosphate pH 7 0.5 mM *trans*-cinnamaldehyde
***Talaromyces emersonii* xylose reductase**	Luria broth 25 °C, 18 h	50 mM potassium phosphate pH 7	50 mM potassium phosphate pH 7 400 mM xylose
***Thermotoga maritima* iron-containing alcohol dehydrogenase**	Luria broth 20 °C, 18 h	33 mM sodium phosphate pH 7 10% 10x BugBuster	33 mM sodium phosphate pH 7 0.1% (v/v) *n*-butanal[a] (~10 mM)

[a]Enzyme had no prior experimental characterization. Initial tests showed it was more active on *n*-butanal than on propanal or furfural.