

Computer-Aided Engineering of a Non-Phosphorylating Glyceraldehyde-3-Phosphate Dehydrogenase to Enable Cell-Free Biocatalysis

Sam J. B. Mallinson,[§] Delphine Dessaux,[§] Sophie Barbe,* and Yannick J. Bomble*Cite This: *ACS Catal.* 2023, 13, 11781–11797

Read Online

ACCESS |

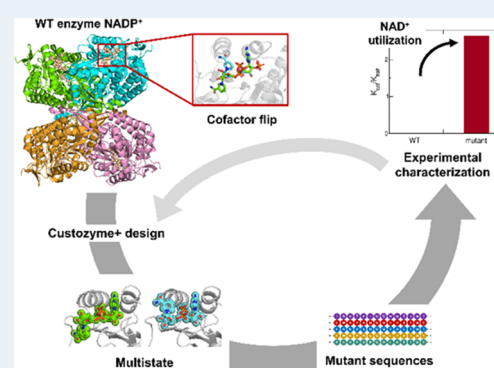
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Redox cofactor utilization is one of the major barriers to the realization of efficient and cost-competitive cell-free biocatalysis, especially where multiple redox steps are concerned. The design of versatile, cofactor balanced modules for canonical metabolic pathways, such as glycolysis, is one route to overcoming such barriers. Here, we set up a computer-aided design framework to engineer the non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GapN) from *Streptococcus mutans* for enabling an NADH linked efficient cell-free glycolytic pathway with a net zero ATP usage. This rational design approach combines molecular dynamics simulations with a multistate computational design method that allowed us to consider different conformational states encountered along the GapN enzyme catalytic cycle. In particular, the cofactor flip, characteristic of this enzyme family and occurring before product hydrolysis, was taken into account to redesign the cofactor binding pocket for NAD⁺ utilization. While GapN exhibits only trace activity with NAD⁺, a ~10,000-fold enhancement of this activity was achieved, corresponding to a recovery of ~72% of the catalytic efficiency of the wild-type enzyme on NADP⁺, with a GapN enzyme harboring only 5 mutations.

KEYWORDS: computational protein design, molecular dynamics, multistate enzyme design, glyceraldehyde-3-phosphate dehydrogenase, cofactor affinity, protein engineering, cell-free biocatalysis



INTRODUCTION

Natural metabolic pathways have been touted as a route to green synthesis of biofuels and biochemicals,^{1–3} but these processes can be difficult to engineer into chassis microorganisms and, when successful, are often hampered by the limitations imposed by complex cellular metabolism such as toxicity of products and intermediates, slow growth rates, and maintaining cell viability. One potential solution to these problems is to employ the reactions *in vitro* with partially purified enzymes or cell lysates, where many of the aforementioned problems are intrinsically resolved. This solution also has additional benefits such as simpler separations and higher volumetric yields. This strategy is not without its own limitations which primarily concern the cost and stability of the biocatalysts and associated cofactors, prominent among which are high-energy cofactors such as NAD(P) and ATP. This is exemplified by one of the most fundamental metabolic pathways—glycolysis. Recently, Bowie and co-workers demonstrated a cell-free glycolytic module with balanced cofactor usage for the synthesis of isobutanol⁴ in which a non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase from *Streptococcus mutans* (GapN, EC:1.2.1.9) is included in place of phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GapDH, EC:1.2.1.12). Deploying GapN, which is not

structurally homologous to GapDH, and whose natural function is to generate NADPH for biosynthetic reactions in either chloroplasts^{5,6} or non-photosynthetic organisms which lack the NADPH production enzymes of the hexose monophosphate pathway,⁷ hence results in a glycolysis pathway that has no net ATP production (and net production of two NADPH molecules). Since isobutanol synthesis from pyruvate uses no ATP, GapN allows a balance of ATP to be maintained to drive the reaction forward (Figure 1), in addition to catalyzing an energetically more favorable reaction than GapDH. This system is therefore not compatible with NADH-utilizing enzymes downstream, such as those in the 2,3-butanediol pathway described by Kay and Jewett.⁸

To our knowledge, only two NAD⁺ active GapN enzymes have been identified in nature^{9,10} with very limited activity on this cofactor. These natural variants are found in thermophilic

Received: March 30, 2023

Revised: June 2, 2023

Published: August 22, 2023



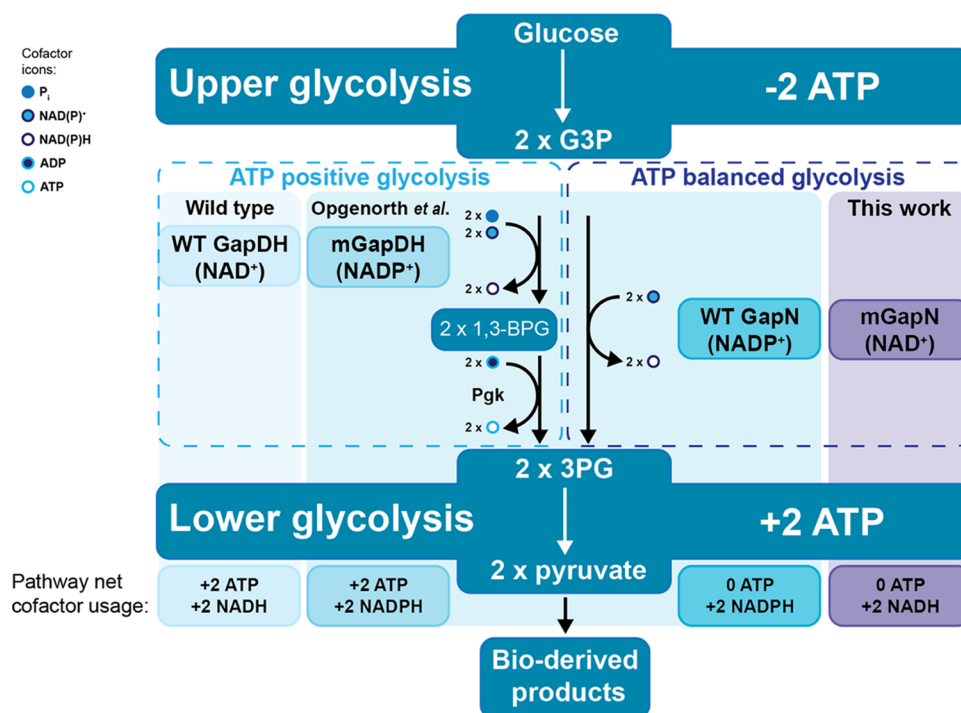


Figure 1. Glycolysis modules for cell-free applications. Standard glycolysis (WT GapDH) converts one molecule of glucose to two molecules of pyruvate with the net generation of two ATP and NADH molecules each. Opgenorth et al.⁴ developed modules that could generate two NADPH instead of NADH with a mutant phosphorylating glyceraldehyde dehydrogenase (mGapDH) and also showed that using the GapN from *S. mutans* (WT GapN) meant that glycolysis could be ATP-balanced while generating two NADPH molecules. In this study, we have engineered WT GapN to accept NAD⁺ (mGapN), allowing a module in which NADH is generated, and ATP usage is balanced. G3P: glyceraldehyde-3-phosphate; 1,3-BPG: 1,3-bisphosphoglycerate; 3PG: 3-phosphoglycerate; and P_{gk}: phosphoglycerate kinase.

microbes and therefore are even less active at standard temperatures. They also have a large range of activators and inhibitors in the glycolytic pathway that complicate *in vitro* reactions.^{10,11} Furthermore, the reported natural function of GapN as generating NADPH for biosynthetic reactions, as described above, indicates that a search for natural NAD⁺ active GapN enzymes would not be successful. We have engineered GapN from *S. mutans* (SmGapN) to drive the oxidation of G3P using NAD⁺, since this GapN has already been demonstrated as a biocatalytic tool,⁴ and is well characterized in the literature.^{12–16} Several strategies were attempted to introduce activity on NAD⁺, including mapping of amino acids from NAD⁺ utilizing homologous enzymes (of which there are few examples as mentioned previously), cofactor specificity reversal using the CSR-SALAD tool developed by the Arnold lab,¹⁷ and finally a fully rational computational protein design (CPD) strategy that proved highly successful. This computer-aided engineering approach combines AI-based computational design methods^{18–21} with molecular dynamics (MD) simulations. Based on multistate design, this approach can take into account the complex two-state mechanism of SmGapN, which involves a cofactor flip for progression through the catalytic cycle.^{13,14,16,22–24} Three rounds of engineering involving this computational approach and experimental characterization of the mutant enzymes (Figure 2) resulted in the dramatic increase of trace activity of SmGapN on NAD⁺, by over 10,000-fold, equating to 72% of the catalytic efficiency of the wild-type (WT) enzyme on NADP⁺.

Development of this engineered GapN variant with comparable kinetics to the WT enzyme on its native cofactor

represents an important step toward establishing an efficient, robust, and versatile cell-free glycolysis module for biochemical syntheses that could have broad applications in the field of green chemistry. The computer-aided engineering approach set up here, which combines multistate enzyme design and MD simulations, may also provide a template for other researchers wishing to conduct similar engineering campaigns.

RESULTS AND DISCUSSION

Initial Engineering Attempts with Residue Mapping and CSR-SALAD Protocols. Since crystallized structures of NAD⁺ utilizing thermophilic GapN and homologous ALDH enzymes are available, initial attempts were made to map residues onto SmGapN—several single and double mutations, and one whole domain swap—from *Thermoproteus tenax* GapN (PDB: 1UXT¹¹) or human ALDH1a2 (PDB: 4X2Q²⁵). These attempts were unsuccessful (data not shown), so a CSR-SALAD library was then generated using the X-ray structure of SmGapN (PDB: 2EUH¹⁵), which consisted of a total of 14 different single-site amino acid mutations across three positions (K177, T180, and G210; see Table S1). From these, two mutations were discovered that only slightly increase the enzyme activity on NAD⁺: G210V and G210I (Figure S1). G210V was the more active of these two, with 1.7% of the WT enzyme activity on NADP⁺ under the conditions in this preliminary screen, and was therefore taken forward into a second round of mutagenesis to be combined with mutations at sites K177 and T180, before being subjected to activity recovery engineering as per the CSR-SALAD workflow. Neither the combination of G210V with K177 and T180 mutations nor the activity recovery variants produced

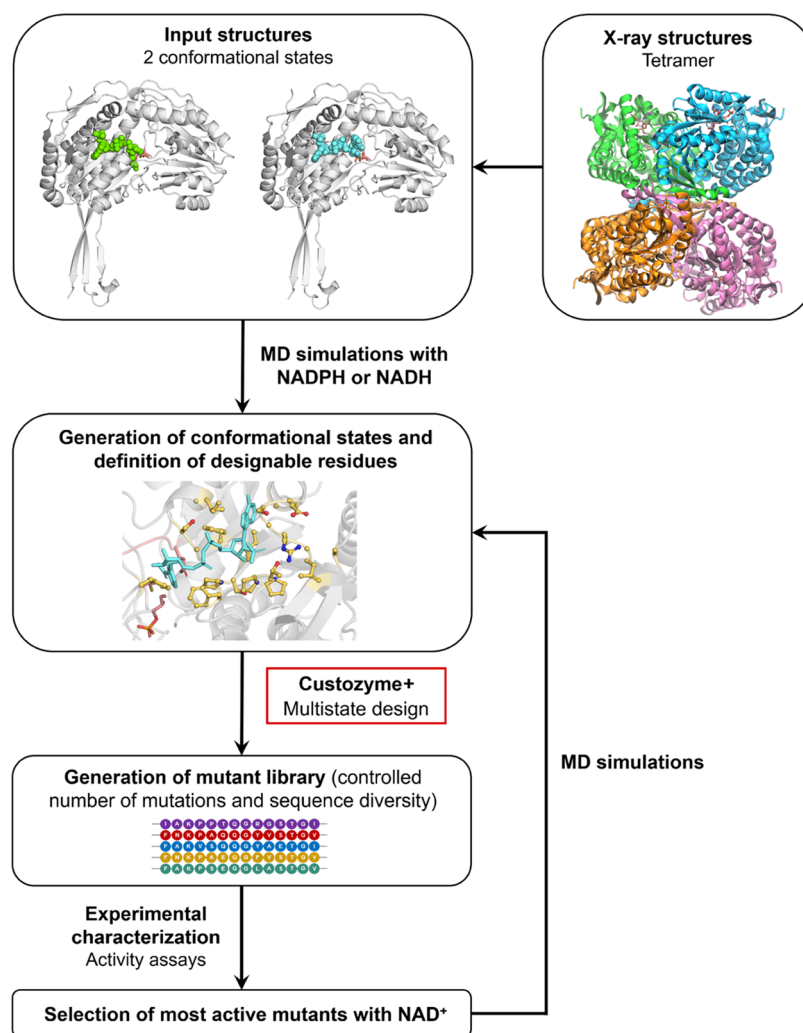


Figure 2. Computationally guided enzyme engineering approach to generate GapN variants with significant activity on NAD^+ . Initially, MD simulations on SmGapN covalently bound to G3P and in interaction with NADH or NADPH, before and after cofactor flip (state 1 and state 2), were carried out to analyze enzyme/cofactor binding and select designable residues. Conformations of SmGapN-G3P with NADH in both states were retrieved after the minimization procedure of MD preparations to be used in the first multistate designs, after their relaxation with the same energy function employed in the design process. The computer-aided engineering approach involved three iterative rounds, each consisting of the following steps: (i) multistate designs²¹ to generate a library of diverse, low-energy mutant sequences with a controlled number of mutations relative to the WT sequence; (ii) activity assays of GapN mutants with the best energy scores; and (iii) MD simulations on GapN mutants with the best NAD^+ activities to explore the impact of mutations and select conformations (based on geometric analyses of the MD trajectories) to be used for the next round.

any observable improvement in activity (data not shown). On the one hand, this can be explained by the crucial role of K177 in the enzymes of the ALDH family, which can stabilize both the structure of these enzymes and the arrangement of cofactors within the pocket, by interacting either with the phosphate group of NADP^+ or with the hydroxyl groups of NAD^+ .^{14,23,26–28} On the other hand, the loop carrying T180 also contains two successive prolines, which constrain it in a conformation that could influence the orientation of the mutated residues and limit their interaction with the cofactor. These characteristics are not taken into account by the CSR-SALAD tool.

Computational Investigation of the Dynamic Behavior of NAD(P) Cofactors in Interaction with SmGapN. After the limited success of these homology-based and semi-rational engineering attempts, a fully rational computational design approach, combining molecular dynamics with AI-based sequence optimization algorithms, was set up. Molecular

modeling and dynamics studies were thus first carried out to decipher the differences in the interaction of SmGapN with NAD^+ or NADP^+ along its catalytic process. In this regard, it should be noted that for several enzymes of the ALDH family, the nicotinamide mononucleotide (NMN) moiety of the NAD(P)^+ cofactor has been shown to undergo a conformational change between the acylation and the deacylation steps of the catalytic reaction^{13,14,23,24} (Figure 3A). This cofactor flip has been investigated in detail in the SmGapN enzyme,¹⁶ which has been crystallized with its NADP(H) cofactor in two distinct conformational states^{15,22} (Figure 3B). After the formation of the thioacylenzyme intermediate by hydride transfer between C284 and NADP^+ , a conformational change of the NMN moiety is required to allow activation of a water molecule by the catalytic E250 and the subsequent hydrolysis of the product.^{16,22} The ELGG conserved loop (residues 250–253) has been shown to play a critical role in this cofactor flip,¹⁶ which should also occur with the NAD^+ cofactor. Based

A



State 1
PDB: 2EUH

State 2
PDB: 2QE0

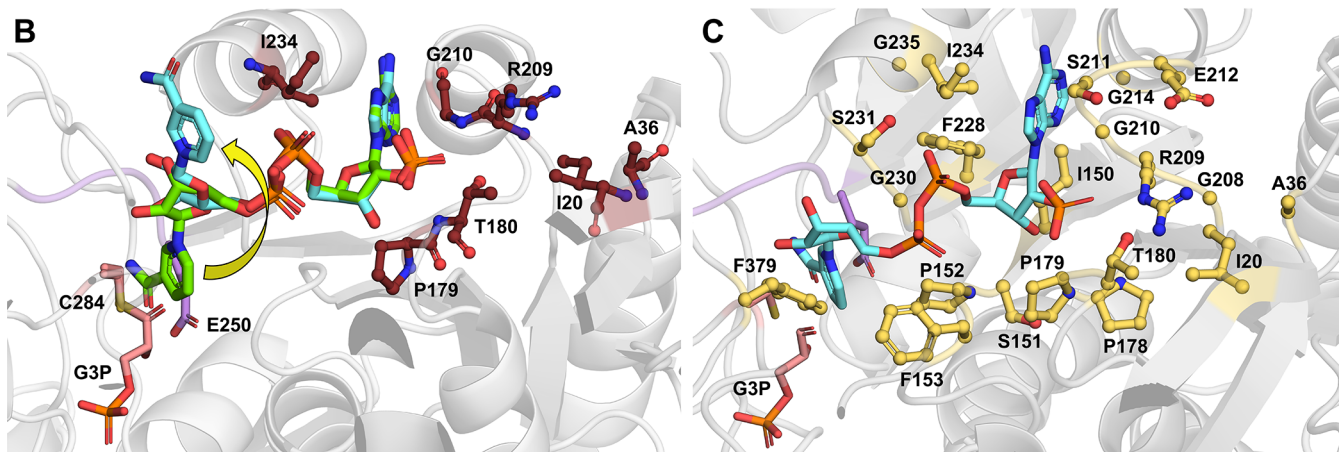


Figure 3. Mechanism of the reaction catalyzed by the SmGapN enzyme. (A) Scheme of the reaction catalyzed by SmGapN with the crystallographic structures indicated. G3P: glyceraldehyde-3-phosphate; 3PG: 3-phosphoglycerate; and CG3: thioacylzyme intermediate. (B) 3D structures of SmGapN enzyme interacting with NADP(H) before (PDB: 2EUH), in green, or after flip (PDB: 2QE0), in cyan. A yellow arrow highlights the flip of the NMN moiety. The ELGG conserved loop (residues 250–253) and C284 are colored, in purple and pink, with E250 and C284 shown in stick representation. The thioacyl intermediate formed by a covalent bond between G3P and C284 and recovered from the 2QE0 crystal structure is also visible in sticks. Favorable mutations identified in this paper are indicated and represented in brown balls and sticks, with backbone atoms also displayed. (C) Structure of SmGapN enzyme before flip of NADPH (state 1) with the side chains of residues targeted during computational design shown in ball and sticks representation and colored in yellow (with NADP⁺ in cyan).

on this knowledge, we compared the behavior of both cofactors inside the binding pocket, in *pre-* (state 1) or *post-* (state 2) cofactor flip state, by MD simulations. The cofactor flip occurring after the acylation step (see Figure 3A), reduced cofactors (NAD(P)H) were considered to build 3D structural models of the thioacylzyme intermediate from the X-ray structures of SmGapN interacting with NADP(H) (state 1, PDB: 2EUH¹⁵ and state 2, PDB: 2QE0²²). From these models, MD simulations of 100 ns were performed with both cofactors.

Analyses of these MD simulations showed that the absence of the phosphate group in NADH leads to a destabilization of the adenosine monophosphate (AMP) moiety and, thus, to an important conformational change, which also impacts the structure of the cofactor binding pocket at this location (Figure S2). The calculation of binding free energies using the MM/GBSA method shows a weaker interaction of SmGapN with NADH than with NADPH, especially in state 1 with a $\Delta G = -54.1 \pm 7.1$ kcal·mol⁻¹ as opposed to a $\Delta G = -124.5 \pm 12.0$ kcal·mol⁻¹ for NADPH (Table S2). This reinforces the crucial role of the phosphate group of NADPH in the binding of the cofactor into the enzyme pocket. Furthermore, depending on the state of the cofactor, two different effects can be noted with NADH compared to NADPH: In state 1, there is an increase in the flexibility of the enzyme and the cofactor, and in state 2, there is a loss of flexibility of both, as highlighted by the ΔRMSF between the NADH and NADPH simulations ($\Delta\text{RMSF} = \text{RMSF}_{\text{NADH}} - \text{RMSF}_{\text{NADPH}}$) of the enzyme backbone and of the cofactor (Figures 4 and S3). The ΔRMSF of the backbone of the Rossmann fold binding domain (Figure 4B), corresponding to residues 146–253,

emphasizes the difference in flexibility of the helices surrounding the adenosine group (residues 210–220 and 232–243 highlighted in Figure 4D). These observations are also supported by the RMSD calculations of the protein backbone and the NAD cofactor during the simulations (Figure S4).

The loss of the cofactor's anchoring by the phosphate group seems to induce its shift into the enzyme pocket (Figure S5), which could lead to a stabilization in an unfavorable conformation for catalysis. As a result, the NADH may not undergo the necessary flip, or it may be unable to exit the pocket after the release of the product. In state 1, the nicotinamide base of NADH is tilted compared to the NADPH (Figure S5). This change results in a different interaction with the thioacyl intermediate and the conserved ELGG loop, which has been shown to be involved in cofactor flip.¹⁶ Moreover, the distortion of the NADH inside the pocket induces a modification of the pocket conformation, as shown on Figure S2. Unlike NADPH, which is moving away from the thioacyl intermediate in state 2, NADH maintains the same distance from it (Figure S6), highlighting the stability of the NADH in state 2 inside the enzyme and therefore its lower propensity to leave the pocket.

Hydrogen bonds between SmGapN and the cofactor were computed during MD simulations (Figure S4D). In state 1, the change of conformation of the NMN moiety disturbs the interaction of the nicotinamide base with E377 and R437 (Figures S7 and S8). Moreover, the adenosine part of the NADH loses the hydrogen bonds formed between NADPH and T180, R209, G210, and S211 and also exhibits fewer

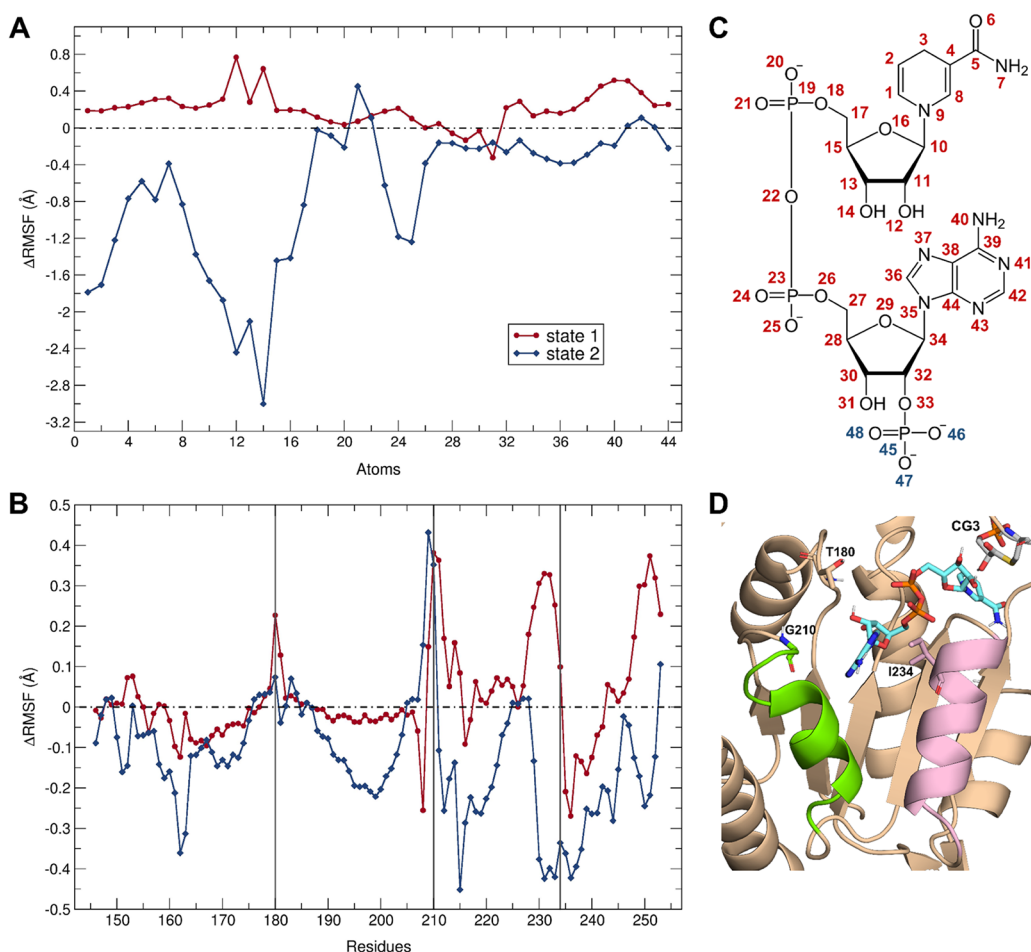


Figure 4. Differences between RMSF of the atoms of the NAD cofactors (A) or of the enzyme backbone residues of the cofactor binding domain (residues 146–253) (B) from NADH and NADPH MD simulations with SmGapN ($\Delta\text{RMSF} = \text{RMSF}_{\text{NADH}} - \text{RMSF}_{\text{NADPH}}$). Cofactor in state 1 is represented in red and cofactor in state 2 in blue. Atoms from the phosphate group cannot be compared across cofactors and are therefore omitted from graph A. Residues 180, 210, and 234, which are positions of interest for mutation, are highlighted by vertical lines on the graph B. (C) Numbering of the NAD(P)H atoms used in the RMSF calculation are shown. (D) Conformations of the helices 210 to 220 (in green) and 232 to 243 (in pink) after 100 ns of SmGapN simulation with NADH (in cyan) in state 1. The thioacyl intermediate (CG3), NADH, and residues T180, G210, and I234 are represented with sticks, and atoms are colored by type.

interactions with K177. This lack of interaction between the AMP moiety and the enzyme induces the conformation shift of NADH in the pocket for both states and is most likely the cause of only trace activity with this cofactor.

Computational Design of Mutant Libraries and Preliminary Experimental Screens. The differences in interactions between SmGapN and both cofactors described above were considered to define designable residues in the structure-based computational enzyme design approach carried out. Designable residues were selected from the regions of the binding pocket with the most variation between simulations with NADPH and NADH in terms of flexibility and interaction with the cofactor while favoring residues around the adenosine moiety. Conserved residues in the ALDH family enzymes, such as K177 and N154, that were shown to be critical for interaction with NAD(P)^+ cofactors or the reaction mechanism,^{13,15,23,26–31} were omitted from the designable region.

It has been previously shown for several ALDH enzymes that T180 plays an important role in cofactor specificity toward NADP^+ ; when a glutamine or glutamate residue is found in this position, NAD^+ tends to be the preferred cofactor.^{23,24,26,30,32,33} However, as previously mentioned, the loop

carrying T180 is strongly constrained by two successive proline residues, which could explain the lack of activity of our GapN T180 single mutants in our previous attempts with the CSR-SALAD protocol. In an effort to achieve a more favorable orientation of residue 180 for NAD^+ binding, combinations of mutations at positions 179 and 180 were thus explored.

All of this led to the definition of a designable region comprising 19 residues in the immediate vicinity to the NAD^+ cofactor, highlighted in yellow in Figure 3C. For each of the 19 designable positions, all 20 natural amino acid types were considered, thus defining a sequence search space of $20^{19} \approx 5.2 \times 10^{24}$. This space was automatically explored in the first round using our computational protein design method²¹ to identify the most favorable energy mutant sequences (with single or multiple mutations). For the second and third design rounds, two residues, I20 and A36, located further away from the cofactor but likely to interact with mutated residues in the first round, more specifically with the tyrosine or phenylalanine in position 209, were added in the designable region (see below for details on definition of the search space). These potential interactions could help stabilize helix 210–220, which is in contact with the adenosine moiety of the cofactor (Figure 4D).

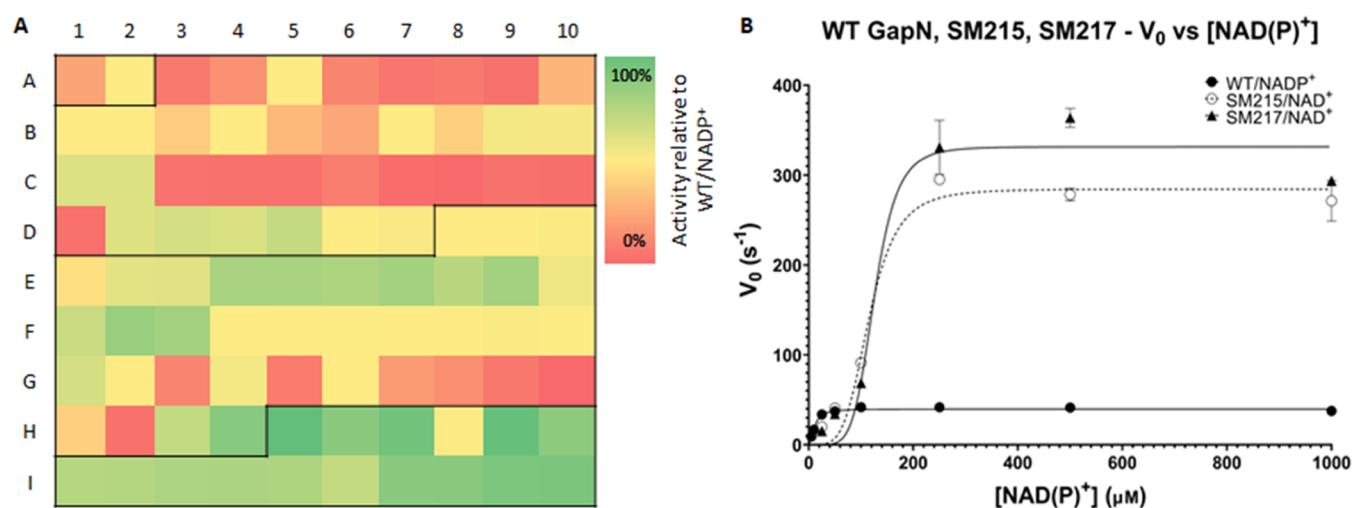


Figure 5. Experimental characterization of GapN mutants for activity on NAD⁺. (A) Heat map of GapN variant activities from the initial screen. Activities were normalized to WT SmGapN with NADP⁺ = 100%. Mutant identities can be found in Table S3, which references the cell ID shown here. Assays were conducted with 5 μ L of crude *Escherichia coli* lysate (see Methods Section for expression conditions) for the respective variants to a 96-well assay plate followed by 195 μ L of a reaction mastermix containing 50 mM HEPES pH 7.5, 5 mM KCl, 5 mM MgCl₂, 1 mM NAD(P)⁺, 5 mM FBP, 0.05 mg/mL FBA, and 0.05 mg/mL TIM. (B) Initial rate vs NAD(P)⁺ concentration for the two top mutants based on preliminary screens. Reactions were set up containing 0.02–20 μ M GapN (purified through HisTrap and then size-exclusion chromatography), 50 mM HEPES pH 8.2, 5 mM MgCl₂, 0.4 mM DL-G3P, and a range of NAD(P)⁺ concentrations. Error bars represent the standard deviation of 3 replicates.

For the optimization of cofactor utilization by ALDH family enzymes, which exhibit two distinct cofactor conformations depending on the reaction step,^{13,14,16,22–24} multistate design is crucial. While most computational design methods commonly consider only one molecular conformational state, we used a multistate design (MSD) method that aims to identify a sequence that simultaneously fits several conformational states required for catalysis. Sequence design thus relies on the energy contributions of conformational states of GapN enzyme with NADH before and after flip. This MSD method^{21,34} of our CPD-dedicated Custozyme+ software is based on state-of-the-art “automated reasoning” artificial intelligence algorithms, more specifically Cost Function Networks (CFNs),³⁵ to identify sequences that minimize the energy of an ensemble of conformational states described as the sum of the energies on each state. This deterministic approach provides optimality guarantees, *i.e.*, given an ensemble of conformational states and an energy function, the sequence returned is the global minimum of the energy function. It can also generate libraries of sequences that satisfy deterministic guarantees, both on sequence diversity and energy quality.¹⁸ Different designable residue spaces (*i.e.*, the list of residues allowed to mutate and the amino acid types allowed in each designable position) can be specified, various pairwise decomposable energy functions can be implemented, and the number of mutations introduced relative to a given sequence (*e.g.*, the WT sequence) can be controlled, which enables a degree of versatility for a broad range of applications. The combination of these various functionalities was used in our design approach.

Three rounds, each including the application of this computational approach to design mutants, followed by experimental testing of the designed enzymes, were carried out (Figure 2). As previously described, prior to the first design round, MD simulations of SmGapN covalently bound to G3P and in complex with NADH or NADPH in both states (states 1 and 2) were performed to identify the designable residues. Conformations of SmGapN-G3P/NADH, obtained after minimization during MD preparation, were used as a starting

point for the first round of design. Prior to the initiation of each multistate design run, all conformations were relaxed using the same energy function used for the design. Each of the three iterative rounds then involved: (i) generation of a mutant library through multistate design and applying inter-solution diversity constraints and distance constraints from the WT sequence; (ii) experimental testing of the designed enzymes with the best energy scores; and (iii) MD simulations of GapN variants with improved (or significant) activity on NAD⁺ to select conformations as a starting point for the next round. This iterative approach aimed to progressively refine the design process by leveraging the insights and results obtained from previous rounds. The three design rounds resulted in a total of 88 variant sequences that were experimentally tested (see Table S3 for the complete list of GapN mutants).

In the first round, 35 designed GapN mutants with the most favorable energy scores, harboring from one to four mutations (Figure 5A, cells A3 to D7, and highlighted in blue in Table S3), were experimentally tested. Among these mutants, 20 exhibited enhanced NAD⁺ activity. Notably, six variants (in bold in Table S3), with either three or four mutations, showed significant NAD⁺ activity, up to 39.4% relative to WT SmGapN on NADP⁺, based on preliminary screening data. These 6 mutants were selected as templates for the multistate design in the second round. In this round, a total of 37 designed mutants with the best energy scores, carrying four to eight mutations, were selected for experimental testing (Figure 5A, cells D8 to H4, and in yellow in Table S3). Of these mutants, 10 showed an improved activity with NAD⁺ compared to the best mutant obtained from the first round (in bold in Table S3). Relative to the WT SmGapN with NADP⁺, NAD⁺ activity up to 77.5% was achieved. Moving forward, these 10 mutants were used as templates in a final round of multistate design. In this round, the activity of the 16 sequences with the best energy scores, containing five to seven mutations, was measured (Figure 5A, cells H5 to I10, and in green in Table S3), resulting in 15 mutants with significant NAD⁺ activity, *i.e.*, higher than mutants from the first design

Table 1. Kinetics, with Respect to the Cofactor and Mutant Identities of Several of the Top-Performing Mutants

enzyme	cofactor	k_{cat} (s^{-1})	hill coefficient	K_{half} (μM)	$k_{\text{cat}}/K_{\text{half}}$ ($\mu\text{M}^{-1} \text{s}^{-1}$)	mutations from wild-type						
						I20	A36	P179	T180	R209	G210	I234
WT	NADP ⁺	39.8	1.90	10.9	3.66							
WT	NAD ⁺	0.224	1.63	918	0.000244							
SM215	NAD ⁺	284	4.33	116	2.45	F		A	Q		V	V
SM217	NAD ⁺	332	5.74	126	2.64			A	Q	Y	V	V
SM223	NAD ⁺	12.3	2.28	147	0.0843	F		A	Q	F	V	V
SM230	NAD ⁺	30.2	3.44	118	0.256	F	N	A	Q		V	V
SM232	NAD ⁺	6.97	3.17	287	0.0243		N	A	Q	L	V	V

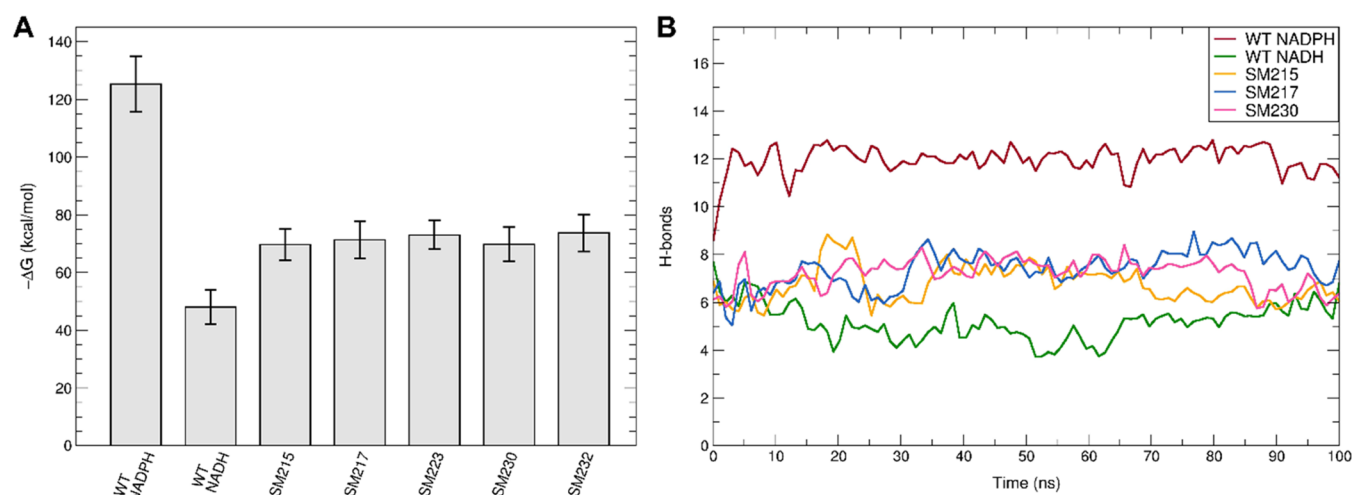


Figure 6. Interaction between NAD(P)H and GapN enzymes along MD simulations. (A) Binding free energy computed between 40 and 50 ns of simulations. (B) Number of hydrogen bonds between the cofactor and GapN enzymes (red: NADPH with WT, green: NADH with WT, orange: NADH with SM215, blue: NADH with SM217, pink: NADH with SM230).

round. Among these mutants, 7 achieved an improved NAD⁺ activity compared to the second-round mutants. The only mutant with no significant NAD⁺ activity (SM219) failed to express over several attempts. The best mutant from these preliminary screens reached an activity on NAD⁺ of 103.8% in comparison to the WT SmGapN activity on NADP⁺.

Experimental Characterization of Top-Performing Mutants. Based on the preliminary experimental screens of *in silico*-generated mutants (Figure 5A), 10 were selected for a final side-by-side comparison (Table S4) before determining kinetic parameters. From the results of this second screen, 3 top variants for activity on each of NAD⁺ and NADP⁺ were chosen (highlighted in bold in Table S4). Since SM215 appeared in both categories, a total of 5 variants were considered for full kinetic characterization with respect to both NAD⁺ and NADP⁺ (Figure S9). Kinetic characterization on both cofactors was performed to better understand the effects of these mutations on GapN cofactor recognition. Detailed results of activity with both cofactors for these 5 variants are reported in Tables S5 and S6. Table 1 and Figure 5B display the results for the 5 variants and top 2, respectively, with NAD⁺. These results show a discrepancy between the final side-by-side comparison (Table S4) and the kinetics (Tables 1 and S5), e.g., SM217 was found to be the most active variant on NAD⁺, despite having modest activity compared to the others in the previous screen. There are several possible explanations for this. First, the substrate was different—in the preliminary screen, in order to reduce cost, G3P was generated *in situ* from fructose 1,6-bisphosphate (FBP) via FBP aldolase (FBA) and triosephosphate isomerase (TIM) while DL-G3P

was directly used for the kinetics; second, because of the presence of FBA and TIM, experiments were conducted at different pH (7.5 for the preliminary screen vs 8.2 for the kinetics); third, the purification procedure was more thorough—the preliminary screen used His-tag purification followed by desalting, while for the kinetics, His-tag purification was followed by size-exclusion chromatography, which appeared to separate out different multimeric forms of the enzyme (data not shown).

The kinetics results revealed that the K_{half} values for the two best variants (SM215 and SM217) on both cofactors had converged to about an order of magnitude better than the WT value for NAD⁺ and an order of magnitude worse than the WT on NADP⁺. Meanwhile, the k_{cat} for NAD⁺ had significantly improved compared to the WT enzyme, with a k_{cat} for both variants about an order of magnitude better than the WT/NADP⁺ value (and 1,000-fold better than the WT enzyme on NAD⁺). Overall, this resulted in a 10⁴-fold better $k_{\text{cat}}/K_{\text{half}}$ than for the WT SmGapN enzyme on a NAD⁺, and 72% of the value of WT SmGapN on the preferred cofactor, NADP⁺, demonstrating successful engineering of this enzyme.

From the results of the second experimental screen (Table S4) and the kinetics parameters measurements (Figure S9 and Table S5), it can be noted that residues 20, 36, and 209 can interact with each other and with neighboring residues and can influence the enzyme behavior with both cofactors.

In the conditions of the second experimental screen, combining F20 with Y209 hinders the enzyme activity, as highlighted by comparison of the behavior of SM215, SM217, and SM221 mutants, whereas the addition of the mutation

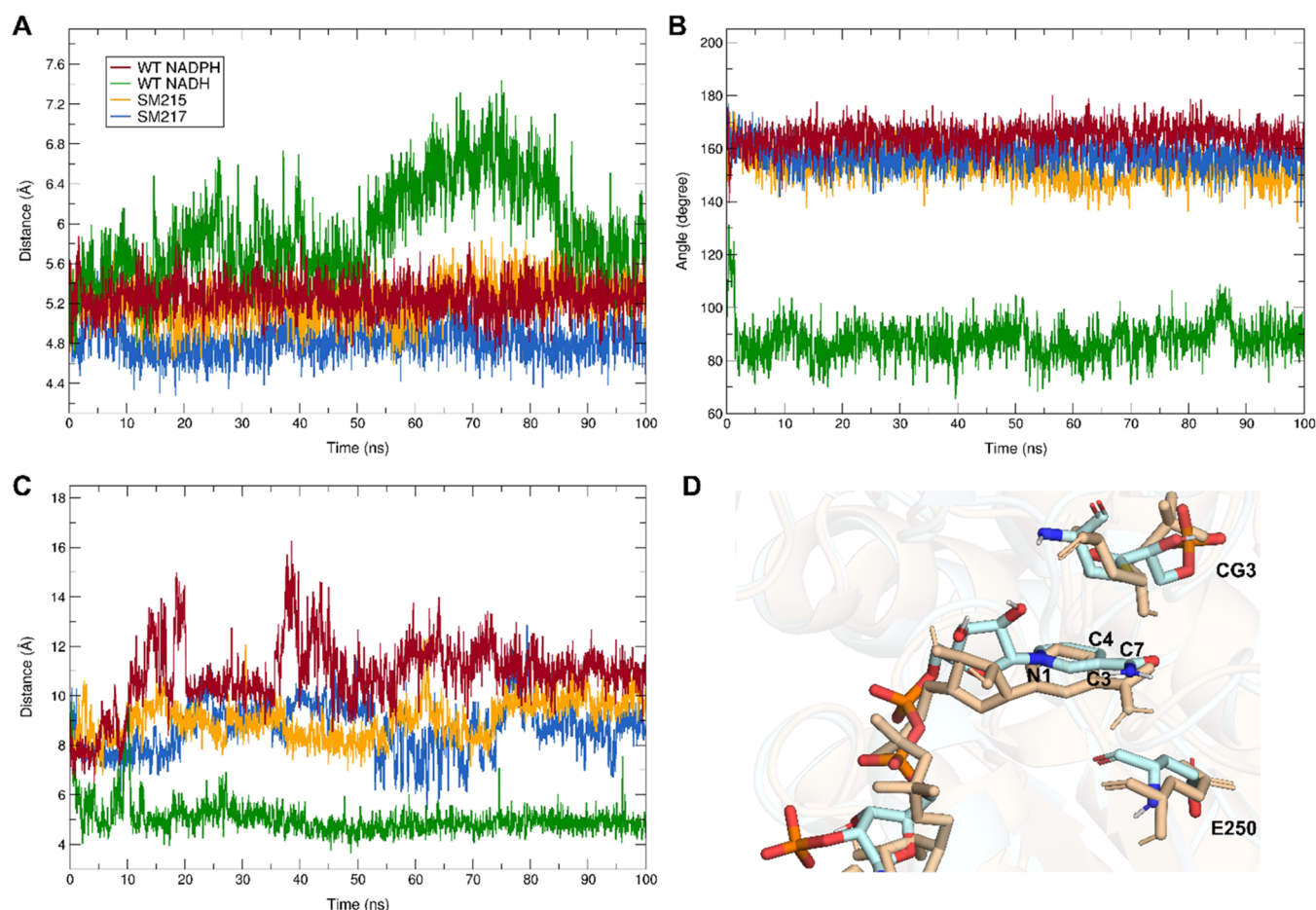


Figure 7. Geometric properties of the NADH cofactor in state 1 or 2 during the 100 ns of MD simulations. (A) Distance between atom $C\alpha$ from the thioacyl intermediate formed between G3P and C284 and atom C3 from the nicotinamide base of the NAD(P)H cofactor in state 1. (B) Dihedral angle between C4 and N1 from the NMN moiety and $C\alpha$ and N from the thioacyl intermediate, for the NAD(P)H in state 1. (C) Distance between atom C7 from the NMN moiety of the cofactor in state 2 and atom N from the backbone of G252 (red: NADPH with WT, green: NADH with WT, orange: NADH with SM215 and blue: NADH with SM217). (D) Representation of the nicotinamide moiety of NADH, colored in light brown, and NADPH, in pale blue, after 100 ns of MD simulation in state 1 with thioacyl intermediate and E250 also visible. Atoms from the cofactor used for the distance and dihedral calculations are labeled. Atoms are colored by type only on the conformation of SmGapN with NADPH. [Figure S5](#) shows the full conformations of the ligand after 100 ns of simulation with NADH or NADPH.

A36N can restore some of its NADP activity, or totally compensate for the effect of the combination of F20 and Y209 on its NAD activity, as evidenced by the differences in activities between SM221 and SM231 mutants. On the other hand, when N36 is combined with F20 without modification of residue R209, N36 seems to negatively impact enzyme activity with both cofactors, as revealed by activities of SM215 and SM230 mutants, while the mutation R209Y can restore some of its activity with both cofactors, as shown by the results for SM231 mutant. It appears from the results obtained for the SM215, SM221, and SM223 mutants with both cofactors that the combination of F20 with F209 has a lesser impact on the activity than that of F20 with Y209. Moreover, the activities of the SM233 and SM232 mutants suggest that a leucine is less favorable than an aromatic residue at position 209, such as tyrosine or phenylalanine, in enhancing NAD^+ activity.

By comparison of the kinetic results of SM215 and SM230 mutants ([Tables 1 and S5](#)), we can deduce that, in these conditions, A36N mainly affects the catalytic efficiency, with a strongly reduced k_{cat} with NAD^+ , and seems to hamper the allosteric effect provided by the mutations of SM215. However, the effect on $NADP^+$ activity is relatively weak,

with the SM230 mutant more active than SM215, and A36N has less impact than R209F, as highlighted by the results for the SM223 mutant.

Molecular Dynamics Simulations of GapN Mutants.

To further understand the mechanisms influencing mutant activities, MD simulations on GapN mutants with the NADH cofactor in both states were more deeply analyzed. The binding free energies of the cofactor, in state 1, with the enzymes (ΔG) were estimated by MM/GBSA calculations. The results for the five experimentally characterized mutants are reported in [Table S2](#) and shown in [Figure 6A](#). An improvement of the binding free energy of NADH with the mutant enzymes of about the same order of magnitude occurs compared to WT SmGapN, which is consistent with the improvement in K_{half} ([Table S5](#)). This enhancement in binding affinity is partially due to an increase in the number of hydrogen bonds between the cofactor and the enzyme, observed along the MD simulations for mutants SM215, SM217, and SM230 ([Figure 6B](#)). MD simulations of both SM215 and SM230 mutants reveal that N36 can interact with R209, N18, and E212, stabilizing this part of the structure ([Figure S10](#)), but also adding steric constraints that limit the

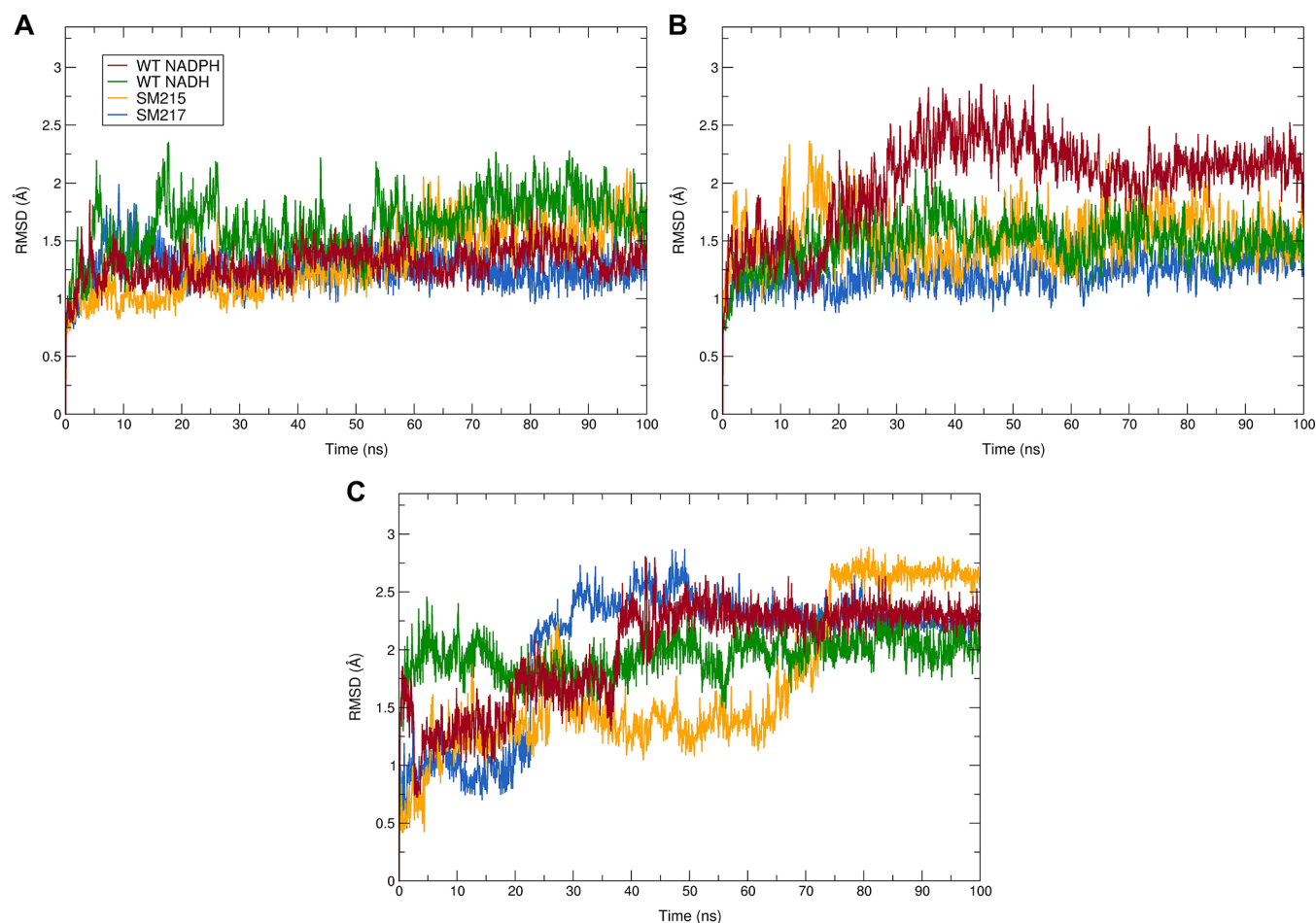


Figure 8. RMSD of the GapN enzymes or of the NAD(P)H cofactor during the 100 ns of MD simulations. (A) RMSD of the protein backbone with the cofactor in state 1 or (B) in state 2. (C) RMSD of the cofactor in state 2 (red: NADPH with WT, green: NADH with WT, orange: NADH with SM215, blue: NADH with SM217).

conformations of these residues and neighboring F20 and Q180. This could explain the loss by the SM230 mutant of the cooperative behavior exhibited by the SM215 mutant and, thereby, its lower catalytic efficiency.

For further analyses, we focus on the results for the two mutants showing the best $k_{\text{cat}}/K_{\text{half}}$ SM215, and SM217 (Figure 7). In state 1, the nicotinamide base of NADH in SM215 and SM217 mutants adopts a similar conformation to the NADPH in the WT SmGapN. The tilt of the base observed for NADH in the WT enzyme (Figure 7D) is not exhibited, as highlighted by the valence and dihedral angles calculations between the NMN moiety and the thioacyl intermediate (Figure 7A,B). In state 2, NAD(P)H should initiate its release from the binding pocket. However, as previously explained, for WT SmGapN, NADH is less flexible than NADPH (Figure S5), and the nicotinamide moiety seems to be caught in the pocket, preventing the cofactor from leaving. For both SM215 and SM217 mutants, the NMN base migrates apart from the thioacyl intermediate and G252 residue in a similar manner to the NADPH in the WT SmGapN (Figures 7C and S6), which is consistent with the RMSD of the cofactor in state 2 (Figure 8C). Indeed, for WT SmGapN, the NADH cofactor undergoes a conformational change early on and then appears to be stable. In contrast, NADPH is subject to more conformational changes. Similarly, NADH (especially its NMN moiety) is more flexible in both

mutants than in the WT SmGapN (Figures 8C, S11, and S12), which further supports an initiation of the exit from the enzyme pocket.

In addition, the structure of both SM215 and SM217 mutants seems stabilized for the cofactor in state 1 compared to the WT SmGapN with NADH (Figures 8A, S13, and S14). As shown on the RMSD of the enzyme backbone (Figure 8A), the structure of these mutants is overall less flexible than the WT enzyme with NADH and tends to resemble the behavior of the enzyme with NADPH. However, the rigidification of the backbone, previously noted for WT SmGapN with NADH in state 2, when compared to NADPH was not overcome by these mutations. In the case of SM217, the backbone flexibility is even further decreased (Figures 8B, S13, and S14), which could hinder the release of the cofactor.

Overall, the mutations of both SM215 and SM217 induced a stabilization of the backbone structure of both helices composed of residues 210 to 220 and 232 to 243, except for the SM215 mutant with NADH in state 2. In this conformational state, R209 interacts with E212, N18, Q180, and F20, and I213 is involved in hydrophobic interactions with neighboring residues. The G210V mutation cannot compensate for these interactions, which induce a large conformational change and thus a higher RMSF for residues 209 to 215 (Figures S13B and S14A,B).

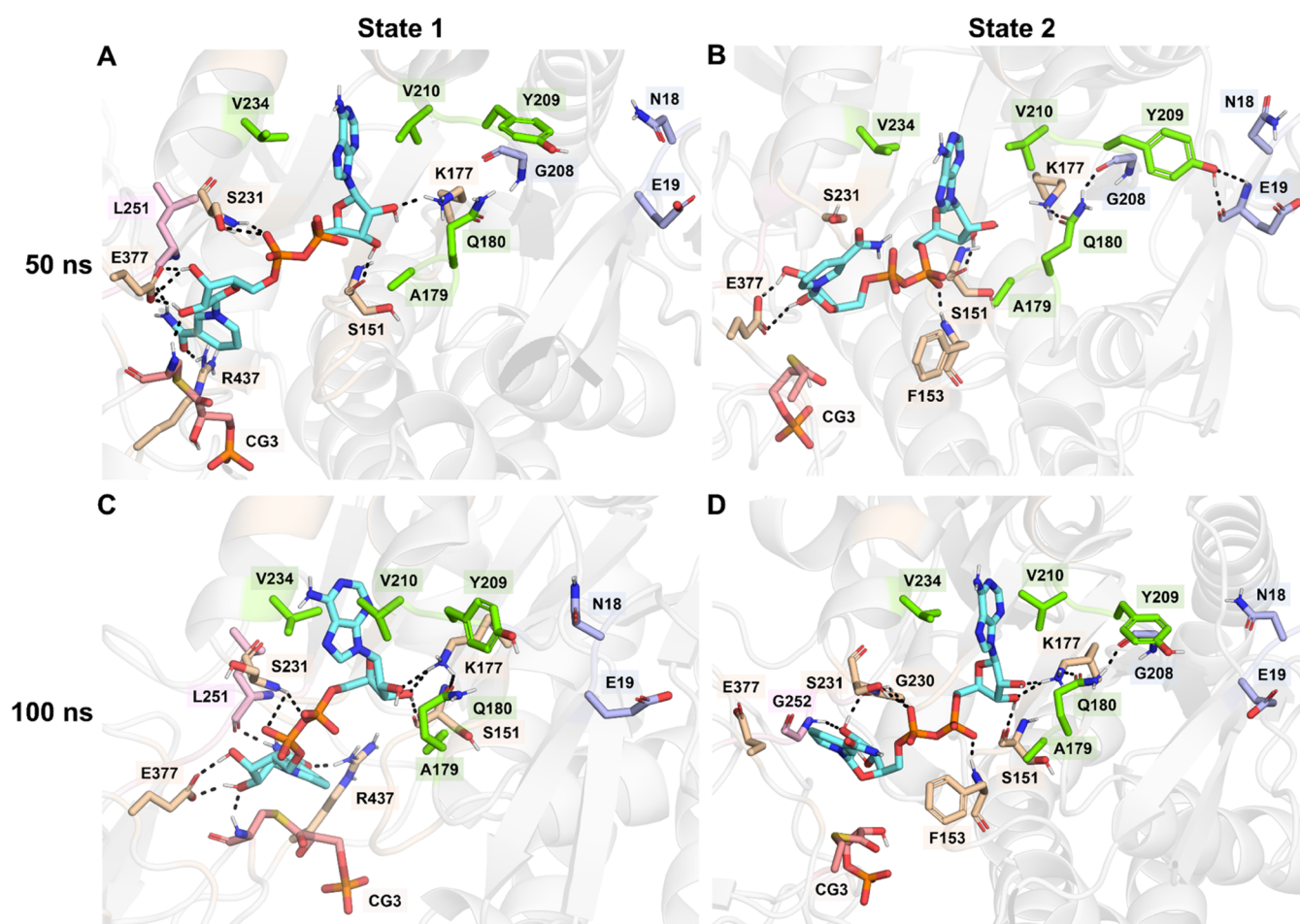


Figure 9. Conformations of SM217 mutant after 50 ns (A, B) or 100 ns (C, D) of MD simulation with NADH in state 1 (A–C) or 2 (B–D). Represented in sticks are: NADH in cyan, the mutations in green, thioacyl intermediate (CG3), L251, and G252 in pink, S151, F153, K177, G230, S231, E377, and R437 in light brown, and N18, E19, and G208 in blue. Atoms are colored by type, and backbone atoms are only represented when in interaction with the cofactor or with mutated residues. Hydrogen bonds between NADH and protein residues, or involving mutated residues, are also shown as black dashed lines.

As mentioned previously, the P179A mutation provides greater conformational flexibility of the loop and consequently of the side chain of residue 180, as illustrated by the conformational rearrangements observed after equilibration of the molecular systems (Figure S15). Overall, Q180 interacts mostly with K177 and the backbone of G208, as visible for SM217 and SM215 mutants in Figures 9, 10, and S16, and seldom with NADH. Therefore, Q180 (combined with A179) is more involved in the stabilization of the protein structure and the reduction of the pocket volume previously occupied by the phosphate group of the NADPH cofactor than in a direct interaction with NADH. However, Q180 also stabilizes K177 in a favorable orientation with respect to its interaction with NADH, thereby enhancing the ability of NADH to interact with the enzyme. Indeed, the AMP ribose of NADH is stabilized by interactions of its hydroxyl groups with the backbone of S151 and with the amine group of K177, as shown on both conformations of the SM217 and SM215 mutants (Figures 9 and 10) and with the calculation of the distances between these residues during the simulations of the SM217 mutant (Figure S16). Interactions between Q180 and K177 or G208 occur less frequently for NADH in state 2 (Figure S16B), which may lead to a pocket rearrangement more favorable to cofactor release.

It was observed for both states of the cofactor that, in addition to hydrophobic interaction with neighboring residues such as I20 or F20, Y209 can form hydrogen bonds with the backbone of E19 *via* its hydroxyl group, as visible for SM217 after 50 ns of simulation (Figure 9B). These interactions involving Y209 further stabilize the conformation of the binding pocket and, thereby, the position of the adenosine moiety of the NADH.

CONCLUSIONS

In this work, we presented the engineering of NAD⁺ activity of GapN to enable a NADH linked efficient cell-free glycolytic pathway similar to that developed by Oppenorth et al.⁴ on NADPH (Figure 1). We describe a rational, computationally guided, engineering workflow to dramatically increase dehydrogenase activity using a nonpreferred cofactor of SmGapN—NAD⁺—by roughly 10,000-fold, equating to 72% of the original catalytic efficiency on NADP⁺ of the wild-type enzyme. It is of note that while the $k_{\text{cat}}/K_{\text{half}}$ values for the wild-type vs SM217 GapN, the best engineered variant, are similar (3.66 vs 2.44 $\mu\text{M}^{-1}\cdot\text{s}^{-1}$), the k_{cat} and K_{half} values individually are about an order of magnitude better and worse, respectively, compared to the wild-type activity (Table 1). The SM217 mutant exhibits an 8.3-fold and 1482-fold higher k_{cat} with

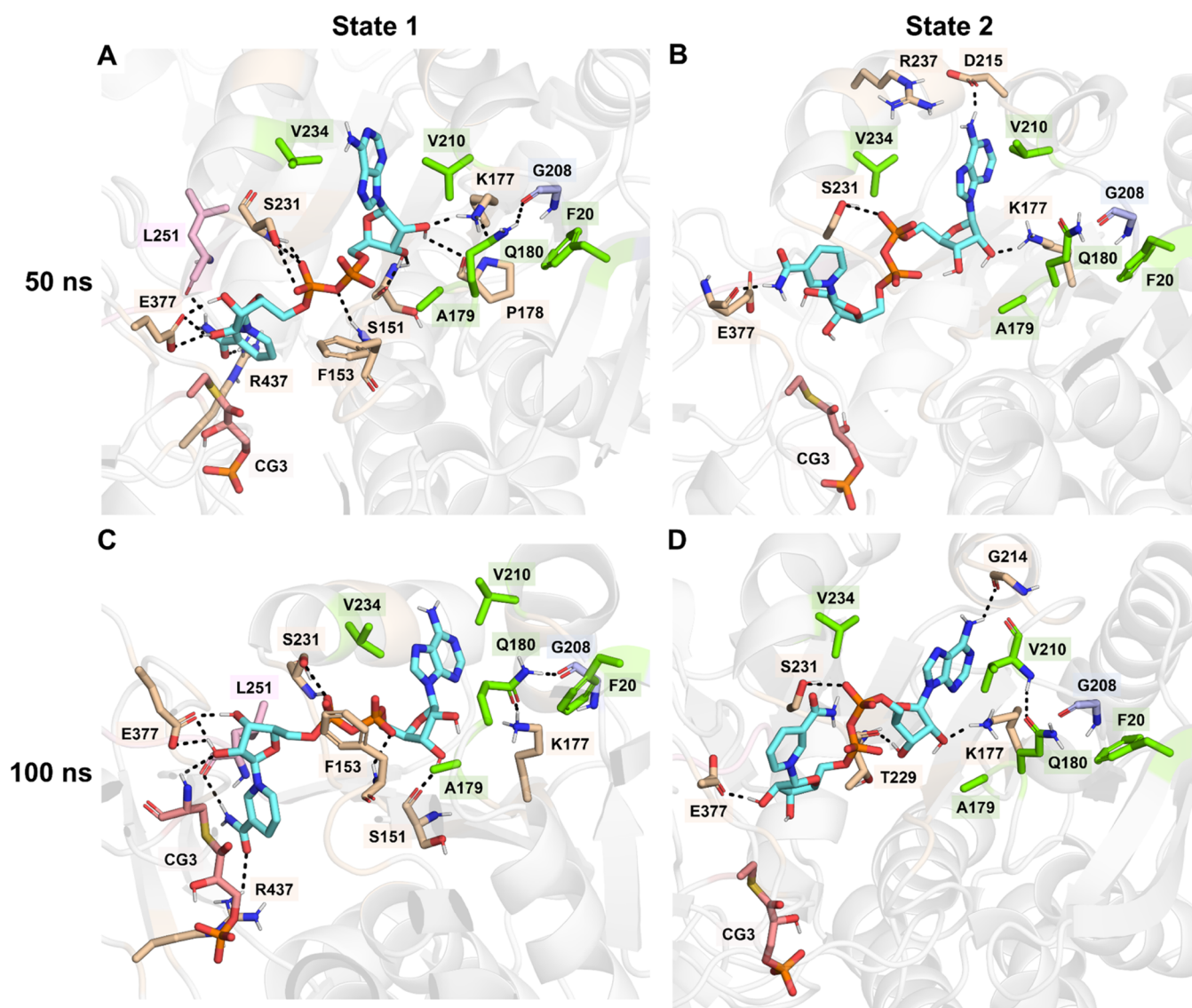


Figure 10. Conformations of SM215 mutant after 50 ns (A, B) or 100 ns (C, D) of MD simulation with NADH in state 1 (A–C) or 2 (B–D). Represented in sticks are: NADH in cyan, the mutations in green, thioacyl intermediate (CG3) and L251 in pink, S151, F153, K177, P178, G214, D215, T229, S231, R237, E377, and R437 in light brown, and G208 in blue. Atoms are colored by type, and backbone atoms are only represented when in interaction with the cofactor or with mutated residues. Hydrogen bonds between NADH and protein residues, or involving mutated residues, and protein residues are also shown as black dashed lines.

NAD⁺ compared with that of the wild-type enzyme with NADP⁺ and NAD⁺, respectively. However, the improvement in the affinity of the enzyme for NAD⁺ is less significant, up to 8-fold (for SM215), compared to the affinity of the wild-type enzyme for NAD⁺. Therefore, activities at lower concentrations of NAD⁺ may suffer when implementing reactions with our mutant enzymes. However, additional designs could further improve the binding affinity of NAD⁺, *i.e.*, achieve a lower K_{half} and thus the activity of GapN with this cofactor. Nevertheless, this engineering campaign has produced enzyme variants that display enhanced affinity for NAD⁺ over that of *T. tenax* GapN for NAD(P)⁺ even in the presence of its effectors,¹¹ and a greatly improved k_{cat} . Moreover, the mutations harbored by these variants are not similar to the residues found in the GapN binding site of *T. tenax*.

It should also be noted that our engineering effort appears to have increased the Hill coefficient of the best variants (Table S6 and Figures 5B and S9), indicating an increase in

cooperativity. In our computational design approach, a monomer of the enzyme was considered; future work will be aimed at analyzing the origin of this modification of the allosteric phenomenon in our GapN mutants.

Additionally, this study served to further establish our computational multistate protein design method that enables the identification of mutant sequences that simultaneously minimize the energy on several distinct molecular/conformational states, which is a tremendous advantage in enzyme designs. Indeed, catalytic processes typically involve several molecular states and conformational changes, which are crucial for enzyme activity and efficiency. Our computer-aided framework combining MD simulations and multistate enzyme design thus provides a valuable tool for the design of proteins that, for instance, exhibit a switch between several different conformational states or enzymes involved in complex multistep catalytic reactions.

As noted earlier in this study, the approach was aimed to augment residual activity on the nonpreferred cofactor, without regard for the effect on NADP⁺ activity, a viable strategy, given the intended application of this enzyme in cell-free systems. This may help explain why computational protein design was more successful than the established CSR-SALAD tool, whose target is cofactor specificity reversal in which native cofactor activity must be suppressed. Indeed, our initial screens often displayed engineered mutants as active or more so on NADP⁺ than on NAD⁺ (data not shown). Just as previous efforts to engineer NAD(P) utilization in enzymes have typically involved switching the specificity, *i.e.*, reducing or removing activity on the wild-type preferred cofactor in addition to introducing/augmenting activity on the non-preferred cofactor, the methods displayed in this paper can also be applied for specificity reversal. In particular, our CPD-dedicated Custozyme+ software can perform positive and negative multistate designs, *i.e.*, identify sequences that optimize the energy on desired states while disfavoring undesirable states. Conformational states of SmGapN in interaction with NADP(H) could be considered in a negative multistate design approach to favor its interaction with NAD⁺ while destabilizing its interaction with NADP⁺.

Other targets of interest in cofactor engineering include mimetics of NAD(P) for application in cell-free biocatalysis, as demonstrated by the Li group.³⁶ The interest in these (semi-)synthetic molecules derives from three weaknesses of natural cofactors—their instability, their cost, and the difficulty in balancing oxidation and reduction in a cell-free reaction. Synthetic NAD(P) cofactors offer the possibility of cheaper and more robust cofactors that could be reduced *in situ*, allowing long-lasting industrial biocatalytic reactions in which reducing equivalents could be supplied *via* a biomimetic cofactor from electricity through an anode. The major drawback is that enzyme activity on these (semi-)synthetic cofactors is difficult to predict and often almost nonexistent. The cost of screening natural mutants *in vitro* until a viable enzyme is found can be prohibitive, but the methods such as those described in this article open up the possibilities to reduce time and cost associated with these engineering campaigns.

METHODS

Construction of 3D Models. 3D models of the SmGapN covalently bound to its substrate, the glyceraldehyde-3-phosphate (G3P), and in interaction with NAD(P)H cofactor in two distinct conformational states (named state 1 and state 2) were constructed from one monomer (chain A) of the high-resolution X-ray structures of SmGapN/NADP⁺ complex (PDB: 2EUH for state 1)¹⁵ and E250A SmGapN covalently bound to G3P (thioacylenzyme intermediate CG3) in complex with NADPH (PDB: 2QE0 for state 2).²²

Both structures of SmGapN were aligned using Pymol,³⁷ and the G3P from the 2QE0 structure was inserted into the 2EUH structure and linked covalently to the C284 thiol to create the first conformational state. The second conformational state was constructed by restoring the catalytic residue E250 in the 2QE0 structure using the side-chain conformation of this residue from the 2EUH structure.

Two conformational states models of SmGapN monomer interacting with NADH were also built by removing the 2'-phosphate group (one phosphate and three oxygen atoms) of the NADPH adenosyl moiety from the previous models.

Protonation states of the charged amino acids and of the histidines were computed using the pdb2pqr server,^{38,39} resulting in protonated His69 and Glu119 and deprotonated Lys254.

Molecular Dynamics Simulations. The molecular all-atom ff14SB⁴⁰ and the general AMBER gaff⁴¹ force fields were used for the enzyme and the substrate, respectively. Parameters from the AMBER parameter database^{42–44} were used for the NADPH and NADH cofactors. Counterions Na⁺ or Cl[−] were added as required to neutralize the systems, which were then solvated with TIP3P water molecules, using a rectangular box with a minimum distance of 0.12 nm between the protein and the simulation box edge, and periodic boundary conditions were applied. Energy minimization cycles were performed on the molecular systems with the steepest descent, and conjugated gradient algorithms. The systems were then subjected to a heating to 298 K over 100 ps in the NVT ensemble, followed by a short 20 ps equilibration. Further equilibration was conducted during 120 ps in the NPT ensemble at a temperature of 298 K and a pressure of 1 bar. Pressure and temperature were maintained constant using the Berendsen algorithm⁴⁵ with a coupling constant of 2 ps for both parameters, and the particle-mesh Ewald (PME) method^{46,47} was used for long-range electrostatics calculations, with a cutoff radius of 12 Å. Protein backbone atoms were initially restrained to their initial position using a harmonic potential with a force constant that was progressively removed throughout the preparation procedure, with the exception of the backbone atoms of the subunit residues at the tetramer interface (regions 118–144 and 465–475) that were held restrained with constraints of 5 kcal/mol-Å². The production phase of simulations was carried out in the NPT ensemble over 100 ns. The integration time-step of each simulation was 2 fs, and the SHAKE algorithm was used to constrain the bond lengths involving hydrogen atoms to their equilibrium values.⁴⁸ Atomic coordinates of each simulation were saved every 10 ps.

MD simulations were performed for the SmGapN enzyme, associated with the G3P substrate, in interaction with NADH or NADPH cofactors in both possible conformations (state 1 = pdb 2EUH and state 2 = pdb 2QE0). They were also carried out with GapN mutants under the same conditions as the WT enzyme. All molecular dynamics simulations were performed using the AMBER 18 package,⁴⁹ and they were analyzed using *cpptraj* tools from the AMBER package.⁵⁰

GapN Mutant Design. Protein design was carried out using the POSitive Multistate Protein Design (POMP^d) method^{21,34} of our in-house CPD-dedicated Custozyme+ software that seeks for minimum energy mutant sequences from an ensemble of conformational molecular states. The approach relies on a pairwise decomposable energy function, a definition of enzyme amino acids as designable and/or flexible or rigid, a discretized description of the amino acid side-chain conformational space based on a library of rotamers, and an ensemble of rigid backbone conformations of the thioacylenzyme intermediate in complex with the NADH cofactor. Designable residues can be allowed to mutate to any of the 20 natural amino acid types or a subset, flexible residues can reorient their side chain without changing their amino acid type, and rigid residues are completely frozen.

Let us assume a rigid backbone and a pairwise decomposable energy function taking the form

$$E(s) = \sum_{i=1}^n E_i(s_i) + \sum_{i<j} E_{ij}(s_i, s_j)$$

with $E(s)$ the total energy of protein sequence s of length n , $E_i(s_i)$ the energy of the residue s_i , and $E_{ij}(s_i, s_j)$ the energy of interaction between the residues s_i and s_j .

Given the defined search space, the method looks for the sequences that minimize the energy of the ensemble of conformational states described as the sum of the energies on each state. It models this problem as a cost function network (CFN)³⁵ and solves it exactly by returning the global minimum of this energy function and proving its optimality. Weight attribution allows the designer to control the contribution of each state to the total energy. Based on an incremental CFN approach using sequence diversity constraints that lower-bound the Hamming distance between sequences, the method also enables the efficient identification of a set of sequence solutions that satisfy guarantees on both inter-solution sequence diversity and energy quality.¹⁸ Distance constraints can also be used to bound the minimum and maximum number of mutations that can be introduced relative to an input sequence (which can be different from the sequence of the input conformation).

Definition of the Search Space. Based on a comparative analysis of the interaction of the enzyme with each of the cofactors from molecular dynamics simulation results, designable amino acid residues were selected. Residues within the binding pocket that showed significant variation in flexibility and interaction with the cofactor in the simulations with NADPH and NADH were carefully chosen, with a focus on the residues surrounding the adenosine moiety. 19 designable residues (I150, S151, P152, F153, P178, P179, T180, G208, R209, G210, S211, E212, G214, F228, G230, S231, I234, G235, F379) were thus selected as designable for the first design round. Based on the analysis of the results of the first design round (types of mutations introduced and impact of the mutations on enzyme/NADH interaction and conformation from molecular dynamics simulations), two additional residues, I20 and A36, were added to the designable region for the second and third design rounds. The aim was to explore if introducing mutations in these distal residues could enable them to establish interactions with the neighboring residues surrounding the cofactor, resulting thus in the formation of a stabilizing interaction network. For all design runs, the 20 natural amino acids were considered at each designable position. Except for residues 250 to 253 in the conserved loop and the K254, which were defined as rigid, all other amino acid residues in the enzyme were considered as flexible. The Dunbrack⁵¹ rotamer library was used to define the amino acid side-chain conformational space. The extra chi3 sub-rotamer level was used for all residues.

Preparation of Input Enzyme/Cofactor Conformations. For the first design round, the conformations of SmGapN covalently bound to G3P (CG3) and in interaction with NADH in both states (state 1: before flip, and state 2: after flip) were retrieved after the minimization procedure of MD preparation. The conformations were selected immediately after minimization in order to retain the conformation of NADH in the binding pocket as in the initial model built from the X-ray structures of the NADH/enzyme complexes. For the second and third design rounds, the mutants resulting from the previous round that exhibited the best activities after

experimental testing were selected to generate input models. For this purpose, MD simulations of these mutant enzymes with NADH in both states were performed. One conformation for NADH in each state sampled during the early MD production phases was selected based on RMSD analyses and other geometric parameters along the simulations (detailed information on the MD analyses is provided below). Selection was made so that the enzyme backbone could rearrange to accommodate mutations, but the cofactor retained a favorable conformation in the binding site. Before initiating the design process, these WT and mutant conformations with NADH in both states underwent relaxation using Rosetta FastRelax.⁵² Two different procedures were used: (i) FastRelax1: the conformations were relaxed with harmonic constraints applied on the backbone and the side chains heavy atoms; (ii) FastRelax2: in addition to constraints on the backbone and side-chain heavy atoms, residues 251 to 254, CG3 and NADH were fixed during the relaxation process. The first and second design rounds were performed using *beta_nov16*⁵³ Rosetta energy function, while the *beta_genpot* Rosetta⁵⁴ energy function was used for the last design round. NADH and CG3 parameters were generated with *molfile_to_params.py*, *molfile_to_params_polymer.py*, and *mol2genparams.py* tools from the Rosetta package,^{55,56} and Rosetta FastRelax procedures were performed with the same energy function used for the design, either *beta_nov16* or *beta_genpot*.

Design Rounds of Mutant Libraries. For the first round, multistate designs were carried out considering either two or four conformations with NADH in both states, before (state 1) and after (state 2) flip. The two conformations (states 1 and 2) were derived from either the FastRelax1 or FastRelax2 procedure. The four conformations (2 for each NADH state) correspond to the ensemble of conformations resulting from the two Rosetta FastRelax procedures (2 conformations from FastRelax1 and 2 from FastRelax2). Multistate design was thus performed from three distinct sets of conformations. The same weight was assigned to each conformation. Distance constraints from the WT sequence were applied to produce sequences with an increasing number of mutations. Therefore, 12 design runs were performed: 4 runs (with distance constraints from 1 to 4) for each of the 3 sets of conformations. Six sequence solutions with an inter-solution diversity threshold equal to 1 were generated for each of these design runs. The 35 unique sequences with the best Rosetta energy scores were selected for experimental testing.

For the second round, multistate designs were performed using two distinct sets of two conformations (states 1 and 2) of the 6 GapN mutants (containing 3 or 4 mutations) with the best experimental activities in the first round. Each set was derived from either the FastRelax1 or FastRelax2 procedure. An equal weight was given to each conformation. Design processes involved two strategies. In the first strategy, all 21 positions were considered designable, as described in the previous section on search space definition. In the second strategy, the residues that were mutated in each of the 6 GapN variants used as input were excluded from the 21 positions, leaving 17 or 18 designable residues. As before, distance constraints from the WT sequence were applied to generate sequences with a controlled number of mutations (from 4 to 8) relative to the WT GapN. Six sequence solutions were produced for each design run with a diversity threshold of 1. The 37 unique sequences with the best Rosetta energy scores were selected for activity assays.

For the third design round, the 10 mutants (carrying 4 or 5 mutations) showing a better activity than the previous best mutant were used as input templates. As in the previous round, two strategies of design were employed; either all 21 positions were designable, or the residues that were mutated in each of the 10 GapN variants used as input were excluded from the designable positions, leaving 16 or 17 designable residues. Other parameters of the design procedure were also similar to the second round of design, except for the use of the Rosetta *beta_genpot*⁵⁴ energy function and the distance constraints from the WT sequence applied. Solution sequences containing between 5 and 7 mutations relative to the WT GapN were produced. The 16 unique sequences with the best Rosetta energy scores were selected for experimental testing.

In Silico Analyses of Molecular Systems (WT and Mutants). *MM/GBSA Calculations.* Binding free energies, ΔG , between NAD(P)H and GapN were estimated using the molecular mechanics generalized Born surface area (MM/GBSA) method with the *mmpbsa.py*⁵⁷ tool provided with the AMBER 18 package. The MM/GBSA method consists of calculating the binding free energy (ΔG_{bind}) of a ligand from molecular dynamics simulations. This binding free energy is estimated by the difference in free energy of the protein substrate complex and of the sum of the free energies of the unbound protein and ligand: $\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}})$, and it encompasses both enthalpy and entropy variations, as described by the Gibbs equation

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S \approx \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S$$

ΔE_{MM} includes the variation of the energy comprising the internal energy ΔE_{int} , the electrostatic energy ΔE_{elec} , and the van der Waals energy ΔE_{vdW} ($\Delta E_{\text{MM}} = \Delta E_{\text{int}} + \Delta E_{\text{elec}} + \Delta E_{\text{vdW}}$). The variations of the solvation free energy ΔG_{solv} combine the changes of the electrostatic ΔG_{GB} and non-electrostatic solvation free energies ΔG_{SA} ($\Delta G_{\text{solv}} = \Delta G_{\text{GB}} + \Delta G_{\text{SA}}$). The electrostatic and non-electrostatic solvation free energies are estimated by solving the Generalized Born equation and by using the solvent-accessible surface area (SASA), respectively. The conformational entropy $T\Delta S$ was considered negligible and was therefore not considered in our calculations.

We estimated the binding free energy on 100 conformations collected either every nanosecond throughout the 100 ns or between 40 and 50 ns of MD simulations. The standard parameters reported by Miller et al. for calculations with AMBER program *mmpbsa.py*⁵⁷ were used, with the default implemented GBSA method.⁵⁸

Geometric Parameter Analyses. The flexibility of the enzymes and cofactors was assessed during MD simulations by calculating the RMSD of all cofactor or protein backbone atoms, relative to the initial structure, as well as the RMSF of the protein backbone or cofactor heavy atoms. The total number of hydrogen bonds between the cofactor and the protein was also determined during MD simulations.

The distances and dihedral angles between key positions were computed along the MD simulations. For assessing the displacement of the cofactor in state 2 toward its release, we calculated the distance between the C7 atom from the NMN moiety and the N atom from G252 backbone or the N atom from the CG3 thioacyl intermediate. To evaluate the conformational distortions of the NMN moiety of the cofactor in state 1, the distances between the C α from the CG3 and C3 atom from the nicotinamide base of the NAD(P)H cofactor

were measured. Moreover, the dihedral angle between C4 and N1 from NMN and C α and N from CG3 was computed for the NAD(P)H in state 1. Distances between atoms belonging to key residues interacting either with NADH or with each other were also calculated during MD simulations.

Generation of CSR-SALAD Variants. Cofactor specificity reversal and activity recovery mutants were generated by QuickChange PCR using primers containing degenerate codons, as indicated by the CSR-SALAD tool.¹⁷ PCR products were transformed into BL21(DE3) cells by electroporation and plated onto LB agar plates supplemented with kanamycin. 96 deep-well blocks containing 0.5 mL of LB per well were inoculated with one colony per well and grown overnight at 37 °C with shaking (all plate shaking was performed at 900 rpm with a 3 mm orbital throw). The next day, 50 μL from each well was transferred to a fresh block containing 600 μL of LB per well. Cells were grown for 4 h at 37 °C before expression was induced by adding IPTG to a final concentration of 1 mM. Cultures were then grown overnight at room temperature before harvesting by centrifugation and storing at -20 °C.

Testing CSR-SALAD Variants. Cells were lysed by addition of 200 μL of lysis buffer containing 50 mM HEPES pH 7.5, 200 mM NaCl, 1 mg/mL lysozyme, and 25 U/mL Pierce Universal Nuclease, followed by shaking for 1 h at room temperature. Reactions were set up by adding 5 μL of lysate to a 96-well assay plate followed by 195 μL of a reaction mastermix containing 50 mM HEPES pH 7.5, 5 mM KCl, 5 mM MgCl₂, 1 mM NAD(P)⁺, 5 mM FBP, 0.05 mg/mL FBA, and 0.05 mg/mL TIM. Absorbance at 340 nm was recorded every 15 s for up to 15 min on a Molecular Devices SpectraMax 190 plate reader at room temperature.

Cloning and Expression of In Silico Designed Variants. Gene fragments were ordered from Twist biosciences for a 300 bp region around the active site and were cloned into the vector containing the GapN gene by Gibson assembly using NEBuilder HiFi DNA assembly kit from NEB. Mutations at two sites that were not contained within this region (I20 and A36) were generated using NEB KLD mutagenesis. Plasmids were transformed into BL21-(DE3) and plated onto LB agar supplemented with 50 mg/L kanamycin. Colonies were grown overnight in 5 mL of LB supplemented with kanamycin before inoculation of 0.5 mL pre-culture into 50 mL of LB. Cultures were grown with shaking at 225 rpm to an OD₆₀₀ of 0.5. Expression was induced with addition of 1 mM IPTG before overnight incubation at 18 °C. Cells were harvested by centrifugation and stored at -20 °C.

Testing of In Silico Designed Variants. Cells were lysed by incubation with 10 mL of lysis buffer containing 50 mM HEPES pH 7.5, 200 mM NaCl, 1 mg/mL lysozyme, and 25 U/mL Pierce Universal Nuclease, followed by shaking for 1 h at room temperature. Protein was purified on a HisTrap FF 5 mL column (Cytiva) before desalting on a HiPrep 26/100 desalting column (Cytiva). Reactions contained 0.1 mg/mL GapN variant, 50 mM HEPES pH 7.5, 5 mM KCl, 5 mM MgCl₂, 1 mM NAD(P)⁺, 5 mM FBP, 0.05 mg/mL FBA, and 0.05 mg/mL TIM. Absorbance at 340 nm was recorded every 15 s for up to 15 min on a Molecular Devices SpectraMax 190 plate reader at room temperature. Variants were purified and tested in batches of 5–7 with a positive control of wild-type SmGapN with NADP⁺. For the final side-by-side comparison, a lower concentration of GapN variant was used (0.05 mg/mL instead of 0.1), DL-G3P was used as the substrate at a

concentration of 0.4 mM, and the pH of the buffer was changed to the reported optimum of 8.2.¹³

Kinetics with GapN Variants. Protein was expressed as above. Purification was performed using a 5 mL HisTrap FF (Cytiva) followed by size exclusion using Cytiva HiLoad 26/600 Superdex 200 pg. Reactions were set up containing 0.02–20 μ M GapN, 50 mM HEPES pH 8.2, 5 mM MgCl₂, 0.4 mM DL-G3P, and a range of NAD(P)⁺ concentrations. Data was collected on a Horiba Duetta fluorescence spectrometer with excitation/emission wavelengths of 340/440 nm and processed using Graphpad Prism software. Kinetic parameters were calculated using an allosteric sigmoidal model. This fits previous work with homologues,¹¹ though some studies on GapN have used classic Michaelis–Menten models.¹³ See Table S6 for a comparison of R² values for different fits.

F-1,6-BP and G3P were purchased from Sigma; all other chemicals were purchased from Fisher.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscatal.3c01452>.

CSR-SALAD results; details of GapN mutants experimentally tested; second experimental screen and kinetics measurements results; binding free energies of NAD(P)H cofactors in interaction with GapN enzymes; conformations of WT SmGapN with NAD(P)H cofactors during MD simulations; RMSF and Δ RMSF of protein backbone and NAD(P)H cofactors during MD simulations for WT SmGapN; RMSD of enzyme backbone and NAD(P)H and number of hydrogen bonds between cofactor and WT SmGapN during MD simulations; distances between NMN moiety of NAD(P)H cofactors in state 2 and thioacyl intermediate of WT or mutant GapN enzymes during MD simulations; conformations of SM230 mutant with NADH during MD simulations; RMSF and Δ RMSF of protein backbone and NAD(P)H cofactors for mutant GapN compared to WT enzyme; reorientation of residues 178 to 181 for WT and mutants GapN; distances between atoms from key residues during MD simulations for SM217 mutant (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Sophie Barbe – Toulouse Biotechnology Institute (TBI), Université de Toulouse, CNRS, INRAE, INSA, ANITI, 31077 Toulouse, France; orcid.org/0000-0003-2581-5022; Phone: +33 (0)5 61 55 94 93; Email: Sophie.Barbe@insa-toulouse.fr

Yannick J. Bomble – Biosciences Center, National Renewable Energy Laboratory, Golden, Colorado 80401, United States; orcid.org/0000-0001-7624-8000; Phone: +1 303 384 7729; Email: Yannick.Bomble@nrel.gov

Authors

Sam J. B. Mallinson – Biosciences Center, National Renewable Energy Laboratory, Golden, Colorado 80401, United States
Delphine Dessaux – Toulouse Biotechnology Institute (TBI), Université de Toulouse, CNRS, INRAE, INSA, ANITI, 31077 Toulouse, France; orcid.org/0000-0002-6541-1105

Complete contact information is available at: <https://pubs.acs.org/10.1021/acscatal.3c01452>

Author Contributions

[§]S.J.B.M. and D.D. contributed equally to this work. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

This work was authored by Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory for the United States Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding for Y.J.B. and S.J.B.M. was provided by the United States DOE Energy Efficiency and Renewable Energy (EERE) Bioenergy Technologies Office (BETO). This work was also authored and supported by INRAE for D.D. and S.B. and the French ANR grant ANR-19-PI3A-0004 for S.B.

Notes

The authors declare no competing financial interest. The publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for United States Government purposes. The views expressed in the article do not necessarily represent the views of the United States Department of Energy or the United States Government.

■ ACKNOWLEDGMENTS

This work was granted access to the HPC resources of CALMIP supercomputing center under the allocation 2021-P21026.

■ REFERENCES

- (1) Meadows, A. L.; Hawkins, K. M.; Tsegaye, Y.; Antipov, E.; Kim, Y.; Raetz, L.; Dahl, R. H.; Tai, A.; Mahatdejkul-Meadows, T.; Xu, L.; Zhao, L.; Dasika, M. S.; Murarka, A.; Lenihan, J.; Eng, D.; Leng, J. S.; Liu, C.-L.; Wenger, J. W.; Jiang, H.; Chao, L.; Westfall, P.; Lai, J.; Ganesan, S.; Jackson, P.; Mans, R.; Platt, D.; Reeves, C. D.; Saija, P. R.; Wichmann, G.; Holmes, V. F.; Benjamin, K.; Hill, P. W.; Gardner, T. S.; Tsong, A. E. Rewriting Yeast Central Carbon Metabolism for Industrial Isoprenoid Production. *Nature* **2016**, *537*, 694–697.
- (2) Vardon, D. R.; Franden, M. A.; Johnson, C. W.; Karp, E. M.; Guarnieri, M. T.; Linger, J. G.; Salm, M. J.; Strathmann, T. J.; Beckham, G. T. Adipic Acid Production from Lignin. *Energy Environ. Sci.* **2015**, *8*, 617–628.
- (3) Arvay, E.; Biggs, B. W.; Guerrero, L.; Jiang, V.; Tyo, K. Engineering *Acinetobacter Baylyi* ADP1 for Mevalonate Production from Lignin Monomers. *Metab. Eng. Commun.* **2021**, *13*, No. e00173.
- (4) Opgenorth, P. H.; Korman, T. P.; Iancu, L.; Bowie, J. U. A Molecular Rheostat Maintains ATP Levels to Drive a Synthetic Biochemistry System. *Nat. Chem. Biol.* **2017**, *13*, 938–942.
- (5) Brown, A. T.; Wittenberger, C. L. The Occurrence of Multiple Glyceraldehyde-3-Phosphate Dehydrogenases in Cariogenic Streptococci. *Biochem. Biophys. Res. Commun.* **1971**, *43*, 217–224.
- (6) Smillie, R. M.; Fuller, R. C. Further Observations on Glyceraldehyde 3-Phosphate Dehydrogenases in Plants and Photosynthetic Bacteria. *Biochem. Biophys. Res. Commun.* **1960**, *3*, 368–372.
- (7) Boyd, D. A.; Cvitkovitch, D. G.; Hamilton, I. R. Sequence, Expression, and Function of the Gene for the Nonphosphorylating, NADP-Dependent Glyceraldehyde-3-Phosphate Dehydrogenase of *Streptococcus Mutans*. *J. Bacteriol.* **1995**, *177*, 2622–2627.

- (8) Kay, J. E.; Jewett, M. C. A Cell-Free System for Production of 2,3-Butanediol Is Robust to Growth-Toxic Compounds. *Metab. Eng. Commun.* **2019**, No. e00114.
- (9) Brunner, N. A.; Brinkmann, H.; Siebers, B.; Hensel, R. NAD⁺-Dependent Glyceraldehyde-3-Phosphate Dehydrogenase from *Thermoproteus Tenax*: The First Identified Archaeal Member of the Aldehyde Dehydrogenase Superfamily Is a Glycolytic Enzyme with Unusual Regulatory Properties. *J. Biol. Chem.* **1998**, *273*, 6149–6156.
- (10) Matsubara, K.; Yokooji, Y.; Atomi, H.; Imanaka, T. Biochemical and Genetic Characterization of the Three Metabolic Routes in *Thermococcus Kodakarensis* Linking Glyceraldehyde 3-Phosphate and 3-Phosphoglycerate. *Mol. Microbiol.* **2011**, *81*, 1300–1312.
- (11) Lorentzen, E.; Hensel, R.; Knura, T.; Ahmed, H.; Pohl, E. Structural Basis of Allosteric Regulation and Substrate Specificity of the Non-Phosphorylating Glyceraldehyde 3-Phosphate Dehydrogenase from *Thermoproteus Tenax*. *J. Mol. Biol.* **2004**, *341*, 815–828.
- (12) Marchal, S.; Rahuel-Clermont, S.; Branlant, G. Role of Glutamate-268 in the Catalytic Mechanism of Nonphosphorylating Glyceraldehyde-3-Phosphate Dehydrogenase from *Streptococcus Mutans*. *Biochemistry* **2000**, *39*, 3327–3335.
- (13) Marchal, S.; Cobessi, D.; Rahuel-Clermont, S.; Tête-Favier, F.; Aubry, A.; Branlant, G. Chemical Mechanism and Substrate Binding Sites of NADP-Dependent Aldehyde Dehydrogenase from *Streptococcus Mutans*. *Chem.-Biol. Interact.* **2001**, *130–132*, 15–28.
- (14) Cobessi, D.; Tête-Favier, F.; Marchal, S.; Branlant, G.; Aubry, A. Structural and Biochemical Investigations of the Catalytic Mechanism of an NADP-Dependent Aldehyde Dehydrogenase from *Streptococcus Mutans*. *J. Mol. Biol.* **2000**, *300*, 141–152.
- (15) Cobessi, D.; Tête-Favier, F.; Marchal, S.; Azza, S.; Branlant, G.; Aubry, A. Apo and Holo Crystal Structures of an NADP-Dependent Aldehyde Dehydrogenase from *Streptococcus Mutans*. *J. Mol. Biol.* **1999**, *290*, 161–173.
- (16) Rahuel-Clermont, S.; Bchini, R.; Barbe, S.; Boutserin, S.; André, I.; Talfournier, F. Enzyme Active Site Loop Revealed as a Gatekeeper for Cofactor Flip by Targeted Molecular Dynamics Simulations and FRET-Based Kinetics. *ACS Catal.* **2019**, *9*, 1337–1346.
- (17) Cahn, J. K. B.; Werlang, C. A.; Baumschlager, A.; Brinkmann-Chen, S.; Mayo, S. L.; Arnold, F. H. A General Tool for Engineering the NAD/NADP Cofactor Preference of Oxidoreductases. *ACS Synth. Biol.* **2017**, *6*, 326–333.
- (18) Ruffini, M.; Vucinic, J.; de Givry, S.; Katsirelos, G.; Barbe, S.; Schiex, T. Guaranteed Diversity and Optimality in Cost Function Network Based Computational Protein Design Methods. *Algorithms* **2021**, *14*, No. 168.
- (19) Simoncini, D.; Allouche, D.; de Givry, S.; Delmas, C.; Barbe, S.; Schiex, T. Guaranteed Discrete Energy Optimization on Large Protein Design Problems. *J. Chem. Theory Comput.* **2015**, *11*, 5980–5989.
- (20) Traoré, S.; Allouche, D.; André, I.; de Givry, S.; Katsirelos, G.; Schiex, T.; Barbe, S. A New Framework for Computational Protein Design through Cost Function Network Optimization. *Bioinformatics* **2013**, *29*, 2129–2136.
- (21) Vucinic, J.; Simoncini, D.; Ruffini, M.; Barbe, S.; Schiex, T. Positive Multistate Protein Design. *Bioinformatics* **2020**, *36*, 122–130.
- (22) D'Ambrosio, K.; Pailot, A.; Talfournier, F.; Didierjean, C.; Benedetti, E.; Aubry, A.; Branlant, G.; Corbier, C. The First Crystal Structure of a Thioacylzyme Intermediate in the ALDH Family: New Coenzyme Conformation and Relevance to Catalysis. *Biochemistry* **2006**, *45*, 2978–2986.
- (23) Di Costanzo, L.; Gomez, G. A.; Christianson, D. W. Crystal Structure of Lactaldehyde Dehydrogenase from *Escherichia Coli* and Inferences Regarding Substrate and Cofactor Specificity. *J. Mol. Biol.* **2007**, *366*, 481–493.
- (24) Perez-Miller, S. J.; Hurley, T. D. Coenzyme Isomerization Is Integral to Catalysis in Aldehyde Dehydrogenase. *Biochemistry* **2003**, *42*, 7100–7109.
- (25) Stenkamp, R. E.; Le Trong, I.; Amory, J. K.; Paik, J.; Goldstein, A. S. RCSB PDB - 4X2Q: Crystal Structure of Human Aldehyde Dehydrogenase, ALDH1a2. <https://www.rcsb.org/structure/4X2Q>. (accessed October 01, 2023).
- (26) Ahvazi, B.; Coulombe, R.; Delarge, M.; Vedadi, M.; Zhang, L.; Meighen, E.; Vrieling, A. Crystal Structure of the NADP⁺-Dependent Aldehyde Dehydrogenase from *Vibrio Harveyi*: Structural Implications for Cofactor Specificity and Affinity. *Biochem. J.* **2000**, *349*, 853–861.
- (27) Perozich, J.; Kuo, I.; Lindahl, R.; Hempel, J. Coenzyme Specificity in Aldehyde Dehydrogenase. *Chem.-Biol. Interact.* **2001**, *130–132*, 115–124.
- (28) Shortall, K.; Djeghader, A.; Magner, E.; Soulimane, T. Insights into Aldehyde Dehydrogenase Enzymes: A Structural Perspective. *Front. Mol. Biosci.* **2021**, *8*, No. 659550.
- (29) Pailot, A.; D'Ambrosio, K.; Corbier, C.; Talfournier, F.; Branlant, G. Invariant Thr244 Is Essential for the Efficient Acylation Step of the Non-Phosphorylating Glyceraldehyde-3-Phosphate Dehydrogenase from *Streptococcus Mutans*. *Biochem. J.* **2006**, *400*, 521–530.
- (30) Perozich, J.; Kuo, I.; Wang, B. C.; Boesch, J. S.; Lindahl, R.; Hempel, J. Shifting the NAD/NADP Preference in Class 3 Aldehyde Dehydrogenase. *Eur. J. Biochem.* **2000**, *267*, 6197–6203.
- (31) Talfournier, F.; Pailot, A.; Stinès-Chaumeil, C.; Branlant, G. Stabilization and Conformational Isomerization of the Cofactor during the Catalysis in Hydrolytic ALDHs. *Chem.-Biol. Interact.* **2009**, *178*, 79–83.
- (32) Gmelch, T. J.; Sperl, J. M.; Sieber, V. Molecular Dynamics Analysis of a Rationally Designed Aldehyde Dehydrogenase Gives Insights into Improved Activity for the Non-Native Cofactor NAD⁺. *ACS Synth. Biol.* **2020**, *9*, 920–929.
- (33) Zhang, L.; Ahvazi, B.; Szittner, R.; Vrieling, A.; Meighen, E. Change of Nucleotide Specificity and Enhancement of Catalytic Efficiency in Single Point Mutants of *Vibrio Harveyi* Aldehyde Dehydrogenase. *Biochemistry* **1999**, *38*, 11440–11447.
- (34) Pompd: Positive Multistate Protein Design. GitLab. <https://forgemia.inra.fr/thomas.schiex/pompd>. (accessed January 06, 2023).
- (35) Schiex, T.; Fargier, H.; Verfaillie, G. In *Valued Constraint Satisfaction Problems: Hard and Easy Problems*, 14th International Joint Conference on Artificial Intelligence (IJCAI 1995); Proceedings IJCAI; Morgan Kaufmann: Montreal, Canada, 1995; pp 631–637.
- (36) King, E.; Cui, Y.; Aspacio, D.; Nicklen, F.; Zhang, L.; Maxel, S.; Luo, R.; Siegel, J. B.; Aitchison, E.; Li, H. Engineering Embden–Meyerhof–Parnas Glycolysis to Generate Noncanonical Reducing Power. *ACS Catal.* **2022**, *12*, 8582–8592.
- (37) Schrödinger, L.; DeLano, W. PyMol. 2020. <http://www.pymol.org/pymol>.
- (38) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson–Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667.
- (39) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G.-W.; Holst, M. J.; McCammon, J. A.; Baker, N. A. Improvements to the APBS Biomolecular Solvation Software Suite. *Protein Sci.* **2018**, *27*, 112–128.
- (40) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (41) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (42) AMBER parameter database. <http://amber.manchester.ac.uk/>. (accessed April 05, 2023)
- (43) Holmberg, N.; Ryde, U.; Bülow, L. Redesign of the Coenzyme Specificity in L-Lactate Dehydrogenase from *Bacillus Stearothermophilus* Using Site-Directed Mutagenesis and Media Engineering. *Protein Eng., Des. Sel.* **1999**, *12*, 851–856.

(44) Ryde, U. Molecular Dynamics Simulations of Alcohol Dehydrogenase with a Four- or Five-Coordinate Catalytic Zinc Ion. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 40–56.

(45) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(46) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(47) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(48) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(49) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E. L.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. AMBER, 2018 <https://ambermd.org/doc12/Amber18.pdf>.

(50) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

(51) Shapovalov, M. V.; Dunbrack, R. L. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure* **2011**, *19*, 844–858.

(52) Tyka, M. D.; Keedy, D. A.; André, I.; DiMaio, F.; Song, Y.; Richardson, D. C.; Richardson, J. S.; Baker, D. Alternate States of Proteins Revealed by Detailed Energy Landscape Mapping. *J. Mol. Biol.* **2011**, *405*, 607–618.

(53) Pavlovicz, R. E.; Park, H.; DiMaio, F. Efficient Consideration of Coordinated Water Molecules Improves Computational Protein-Protein and Protein-Ligand Docking Discrimination. *PLoS Comput. Biol.* **2020**, *16*, No. e1008103.

(54) Park, H.; Zhou, G.; Baek, M.; Baker, D.; DiMaio, F. Force Field Optimization Guided by Small Molecule Crystal Lattice Data Enables Consistent Sub-Angstrom Protein–Ligand Docking. *J. Chem. Theory Comput.* **2021**, *17*, 2000–2010.

(55) Drew, K.; Renfrew, P. D.; Craven, T. W.; Butterfoss, G. L.; Chou, F.-C.; Lyskov, S.; Bullock, B. N.; Watkins, A.; Labonte, J. W.; Pacella, M.; Kilambi, K. P.; Leaver-Fay, A.; Kuhlman, B.; Gray, J. J.; Bradley, P.; Kirshenbaum, K.; Arora, P. S.; Das, R.; Bonneau, R. Adding Diverse Noncanonical Backbones to Rosetta: Enabling Peptidomimetic Design. *PLoS One* **2013**, *8*, No. e67051.

(56) Renfrew, P. D.; Choi, E. J.; Bonneau, R.; Kuhlman, B. Incorporation of Noncanonical Amino Acids into Rosetta and Use in Computational Protein–Peptide Interface Design. *PLoS One* **2012**, *7*, No. e32637.

(57) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.

(58) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383–394.

Recommended by ACS

Synthetic NAD(P)(H) Cycle for ATP Regeneration

Emma Willett and Scott Banta

JUNE 27, 2023
ACS SYNTHETIC BIOLOGY

READ 

A Hitchhiker's Guide to Supplying Enzymatic Reducing Power into Synthetic Cells

Michele Partipilo, Dirk Jan Slotboom, *et al.*

APRIL 13, 2023
ACS SYNTHETIC BIOLOGY

READ 

Engineering of a Reductive Aminase to Enable the Synthesis of a Key Intermediate to a CDK 2/4/6 Inhibitor

Jeremy Stefflik, Carlos Martinez, *et al.*

JULY 18, 2023
ACS CATALYSIS

READ 

Rational Design of the Substrate Tunnel of β -Ketothiolase Reveals a Local Cationic Domain Modulated Rule that Improves the Efficiency of Claisen Condensation

Lixia Liu, Yu Deng, *et al.*

JUNE 05, 2023
ACS CATALYSIS

READ 

Get More Suggestions >