

Computational Enzyme Redesign Enhances Tolerance to Denaturants for Peptide C-Terminal Amidation

Published as part of JACS Au virtual special issue "Biocatalysis in Asia and Pacific".

Tong Zhu, Jinyuan Sun, Hua Pang, and Bian Wu*



Cite This: *JACS Au* 2024, 4, 788–797



Read Online

ACCESS |

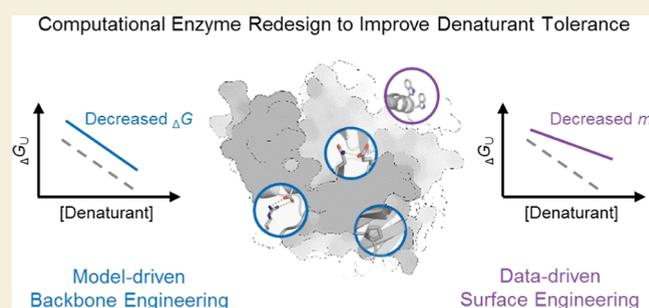
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The escalating demand for biocatalysts in pharmaceutical and biochemical applications underscores the critical imperative to enhance enzyme activity and durability under high denaturant concentrations. Nevertheless, the development of a practical computational redesign protocol for improving enzyme tolerance to denaturants is challenging due to the limitations of relying solely on model-driven approaches to adequately capture denaturant–enzyme interactions. In this study, we introduce an enzyme redesign strategy termed GRAPE_DA, which integrates multiple data-driven and model-driven computational methods to mitigate the sampling biases inherent in a single approach and comprehensively predict beneficial mutations on both the protein surface and backbone. To illustrate the methodology's effectiveness, we applied it to engineer a peptidylamidoglycolate lyase, resulting in a variant exhibiting up to a 24-fold increase in peptide C-terminal amidation activity under 2.5 M guanidine hydrochloride. We anticipate that this integrated engineering strategy will facilitate the development of enzymatic peptide synthesis and functionalization under denaturing conditions and highlight the role of engineering surface residues in governing protein stability.

KEYWORDS: enzyme stability, denaturants, computational redesign, machine learning, peptide C-terminal amidation



1. INTRODUCTION

Growing concerns related to environmental and energy issues have propelled the widespread adoption of biocatalytic approaches, driven by the exceptional selectivity and efficiency of enzymes.^{1,2} However, natural biocatalysts often suffer from inherent fragility, which compromises their activity and durability under demanding conditions such as elevated temperatures and high denaturant concentrations.^{3,4} Over recent decades, substantial interest has arisen in protein engineering strategies aimed at enhancing enzyme tolerance to elevated temperatures.⁵ These conditions can substantially weaken internal hydrophobic interactions, electrostatic interactions, hydrogen bonds, and van der Waals forces, ultimately driving protein unfolding and irreversible aggregation.⁶ To address this challenge, computational tools like Rosetta_ddG⁷ and FoldX⁸ have been developed to identify stabilizing mutations by modeling the physical interactions involved and assessing the energy changes resulting from residue substitutions. These methods have remarkably contributed to numerous successful cases in protein engineering, leading to the creation of enzymes with more robust backbones capable of withstanding elevated temperatures.⁹ Nevertheless, developing a practical computational redesign protocol for bolstering enzyme tolerance to denaturants like urea and

guanidine hydrochloride (Gnd-HCl) remains a continuous endeavor. This ongoing challenge has impeded advancements in optimizing enzymes for protein synthesis, functionalization, and digestion, which are crucial for meeting the demands of applications in pharmaceutical manufacturing, chemical biology, and proteomics research.^{10–12}

The molecular mechanism of denaturant-mediated protein unfolding is still a subject of debate, with two main hypotheses proposed: the indirect mechanism and the direct mechanism.¹³ The indirect mechanism posits that denaturants induce alkane solvation, disrupting the hydrogen bonding network formed by water molecules and diminishing internal hydrophobic interactions.¹⁴ In contrast, recent studies have provided experimental evidence supporting an alternative hypothesis, suggesting a direct mechanism.¹⁵ According to this perspective, denaturants bind to the protein's surface, leading to the formation of a dry molten globule (DMG) state, followed by

Received: December 13, 2023

Revised: January 20, 2024

Accepted: January 22, 2024

Published: February 8, 2024



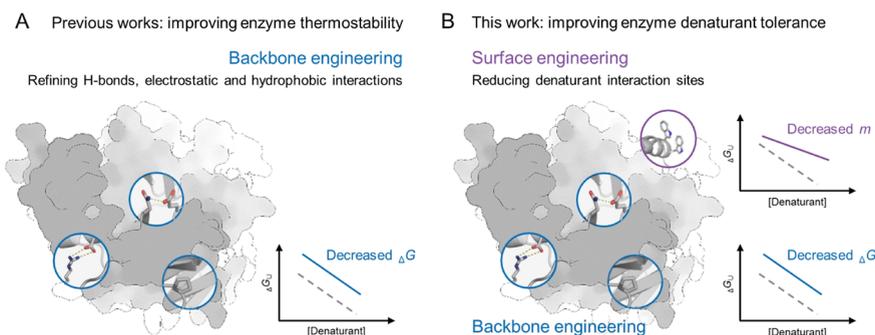


Figure 1. Schematic representation of enzyme engineering to enhance (A) tolerance to elevated temperatures and (B) resistance to high denaturant concentrations.

global structural disruption under the influence of dispersion forces and hydrophobic interactions.¹⁶ The direct mechanism emphasizes the importance of not only reducing backbone flexibility but also minimizing interactions between protein surfaces and denaturing agents to enhance enzyme tolerance to denaturants. Due to the intrinsic chemical and topographical heterogeneity of the protein surface, efficiently depicting protein-denaturant interactions poses a challenge for approaches grounded in physical models.^{17,18} This challenge likely contributes to the observed disparities between the predictions made by Rosetta_ddG and FoldX and the experimentally measured thermodynamic stability changes of surface mutations.¹⁹ To address this restriction, insights from strategies used in protein thermostability engineering may provide valuable guidance. We have previously developed a strategy known as GRAPE, which combines multiple computational tools to counter the sampling biases associated with a single approach for identifying adequate stabilizing mutations to improve enzyme thermostability during the mutation accumulation process.²⁰ Building upon this strategy, we recognized the potential to integrate diverse computational techniques rooted in distinct principles, such as evolutionary insights. This integration may offer effective solutions for identifying beneficial mutations that might otherwise be overlooked by purely model-driven methods.

In this study, we introduce an engineering strategy termed “GRAPE_DA” (GRAPE for improving enzyme tolerance to denaturing agents), which synergistically combines several data-driven and model-driven methods to comprehensively predict stabilizing mutations on both the protein surface and backbone (Figure 1). To demonstrate the effectiveness of this methodology, we selected a peptidylamidoglycolate lyase (PAL) as our exemplar for redesign. Our efforts culminated in the creation of the final mutant, PAL14, which exhibited a remarkable up to 24-fold enhancement in conversion rates under the influence of 2.5 M Gnd·HCl. In addition to mutations that improve the stability of PAL14’s backbone by refining polar and hydrophobic interactions, several polar-to-hydrophobic mutations on the protein surface contribute to the reduction of interaction sites for Gnd·HCl. These refinements in both the surface and internal regions collectively mitigate the effects of denaturants on the enzyme, presenting this strategy as a pragmatic engineering protocol for bolstering the enzyme robustness in the presence of denaturing agents. Within the GRAPE framework, we introduced MSAddG, a machine learning-based module that identified four beneficial mutations involving the substitution of surface polar residues with aromatic ones. This unique preference, not commonly

avored by mainstream computational methods, provides valuable insights into the distinct roles of exposed hydrophobic residues in maintaining protein stability under diverse conditions.

2. RESULTS

2.1. Computational Redesign of a Prokaryotic PAL

PAL catalyses the lysis of peptides possessing a C-terminal hydroxyglycine residue, yielding the corresponding des-glycine peptide amides and glyoxylic acid.²¹ In animals, PAL collaborates with peptidyl-glycine hydroxylating monooxygenase (PHM) to generate C-terminally amidated peptide hormones and toxins.²² Interestingly, several bacterial proteins exhibiting similar activities have been experimentally identified, yet their natural functions and evolutionary trajectories remain unclear.²³ Among these prokaryotic proteins, *Exiguobacterium* sp. PAL (ExiPAL), which can be prepared through recombinant expression in *Escherichia coli*, has been employed as an enzymatic platform for traceless protein synthesis and C-terminal modification.^{24,25} Nonetheless, the limited robustness of ExiPAL curtails its broader utilization for peptides and proteins reliant on Gnd·HCl to maintain solubility.

To address this limitation, we initiated the engineering of ExiPAL to bolster its stability in the presence of denaturing agents. Our initial step involved analyzing the structure of ExiPAL to ascertain the regions amenable to redesign. We generated a structural model of ExiPAL by employing AlphaFold2²⁶ and aligned it with the crystal structure of PAL from *Rattus norvegicus*.²⁷ This analysis revealed that ExiPAL adopts a β -propeller configuration consisting of six blades, each encompassing four antiparallel β -strands (Figure S1A). Within this propeller arrangement, we identified two distinct cavities. One cavity served as the active site, where a zinc ion likely coordinates with H95, H210, and H309. According to the proposed mechanism of RnPAL,²⁷ our scrutiny pinpointed Y178 and R225 as the putative catalytic residues (Figure S1B). In the other cavity, a calcium ion potentially coordinates with the main chains of V43 and I97, as well as the side chain of D310. Our experiments involving ethylene diamine tetraacetic acid (EDTA) demonstrated that EDTA-treated ExiPAL displayed a significant reduction (-21.4 °C, Figure S2) in its apparent melting temperature (T_m). This underscores the pivotal role of divalent metal ions in preserving protein stability. To retain the functional significance of the side chains involved in catalysis and ion binding, we preserved residues H95, H210, H309, Y178, R225, and D310, while the other residues were subjected to engineering.

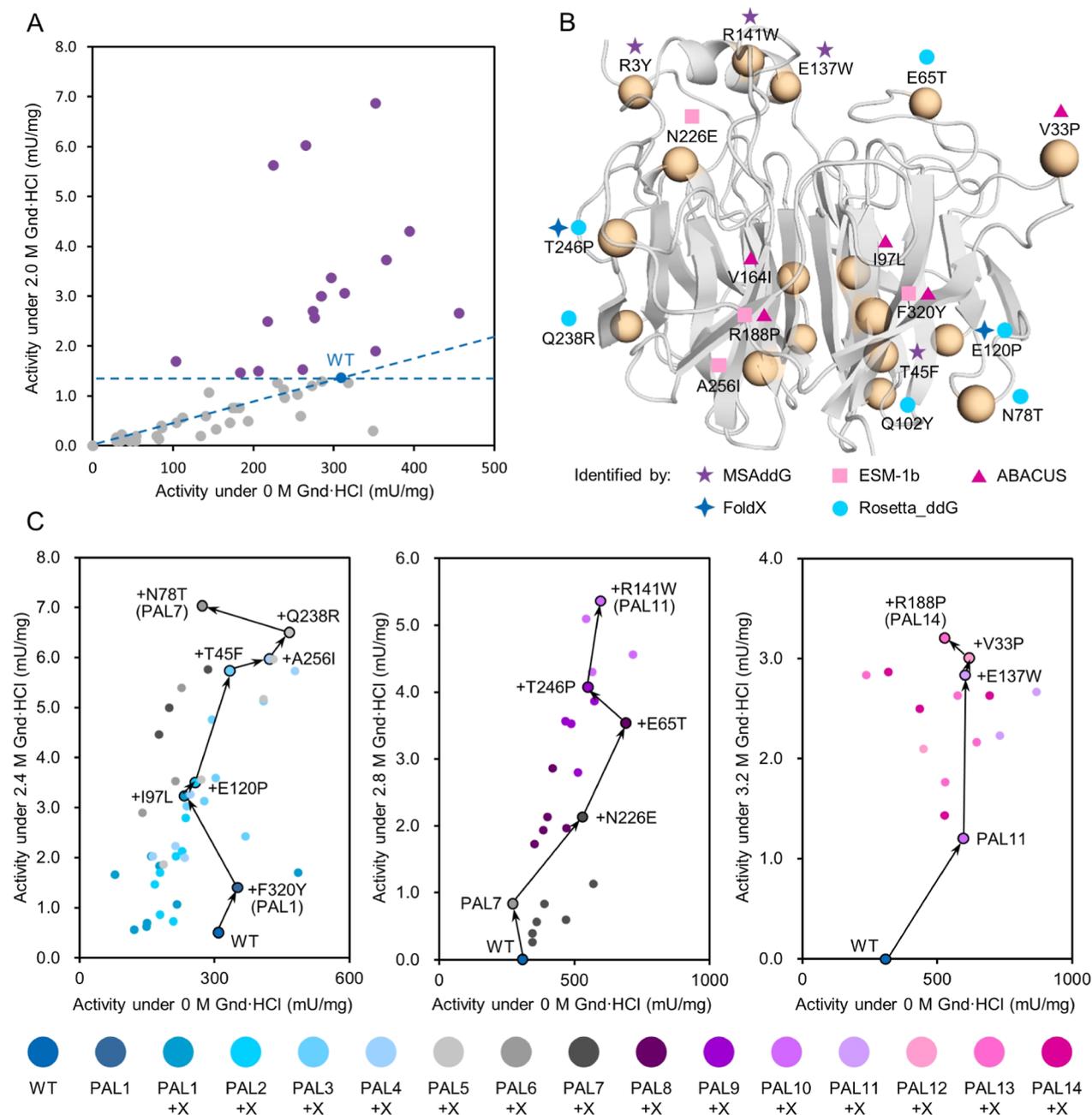


Figure 2. (A) Experimental characterization of the predicted mutations. The lyase activities of the purified enzymes were assessed using α -hydroxyhippuric acid at 0 and 2 M Gnd-HCl, respectively. WT denotes the wild-type ExiPAL (colored in blue). 17 positive candidates are represented in grape circles. (B) Locations and the prediction sources of the 17 mutations. (C) Mutations were systematically accumulated, leading to the generation of PAL14. Each data point represents an accumulation mutant, with colors indicating the respective accumulation round.

Having obtained the structural information on ExiPAL, we delved into leveraging evolutionary insights, another pivotal resource for enhancing enzyme robustness. Due to the intrinsic trade-off between function and stability, enzymes often undergo mutations that increase their activity while sacrificing robustness during evolution.²⁸ In this context, many data-driven methods have been devised to unearth stabilizing mutations by harnessing these accumulated evolutionary changes. One such method is ABACUS, a structure-based sequence redesign software grounded in a statistical energy function.²⁹ ABACUS has demonstrated its utility in improving the thermostability of enzymes like xylanase³⁰ and glucose oxidase.³¹ Given the potential benefits of finding mutations

that reduce backbone flexibility, ABACUS remains a valuable tool in our efforts to enhance ExiPAL's denaturant tolerance. Additionally, strategies such as consensus analysis³² and ancestral sequence reconstruction³³ have been employed to generate enzyme mutants with improved thermostability. However, these methods heavily rely on a deep understanding of the gene family and the intricate selection of relevant sequences, presenting challenges when applied to enzymes with obscure evolutionary trajectories, as is the case with ExiPAL.¹⁷

Consequently, we explored alternative avenues for harnessing evolutionary information by capitalizing on the rapid advancements in machine learning within the field of protein

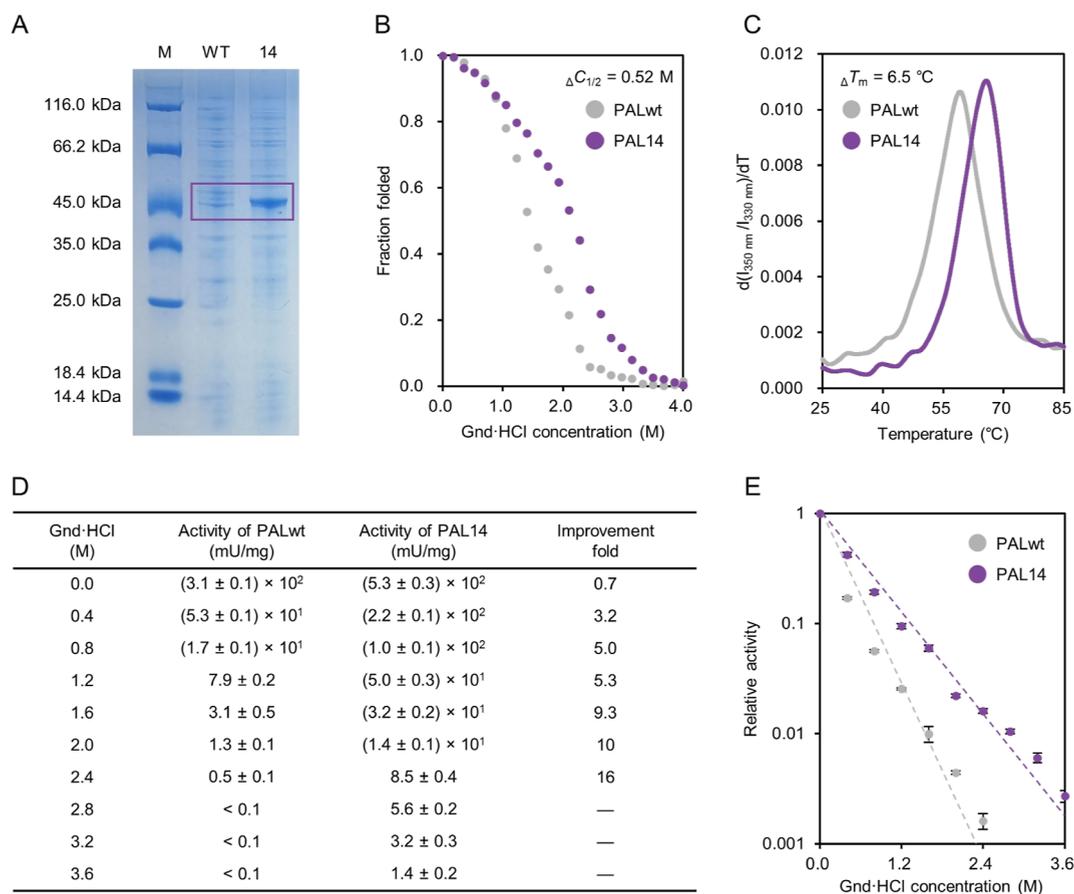


Figure 3. (A) Gel electrophoresis analysis of cell extracts of PALwt and PAL14. M: molecular weight marker. (B) Equilibrium Gnd·HCl-induced unfolding analysis of PALwt and PAL14. (C) Measurement of the apparent T_m values for PALwt and PAL14. (D) Determination of lyase activities of PALwt and PAL14 using α -hydroxyhippuric acid as the substrate at various concentrations of Gnd·HCl. Assays were performed in triplicate. Due to the detection limit of high-performance liquid chromatography (HPLC), the minimum measurable activity is 0.1 mU/mg. (E) The relative activities of PALwt and PAL14 are shown. Activity in the absence of Gnd·HCl was set as 1.

chemistry.³⁴ This has opened up the possibility of more sophisticated integration of evolutionary knowledge and experimental data. Drawing upon the extensive repository of stability parameters stemming from equilibrium denaturant titrations for single-site mutations in the ProThermDB database,³⁵ we developed a machine learning-based method called MSAddG (Figure S5A). MSAddG leverages evolutionary information derived from multiple sequence alignments (MSAs) of homologous proteins. The network architecture of MSAddG incorporates two 7-dimensional vectors representing evolutionary and physicochemical embeddings, respectively (Table S5). After rigorous training using the designated dataset, MSAddG demonstrated impressive fitting outcomes on the training (Figure S5B) and testing (Figure S5C) datasets. To further evaluate the complementary significance of MSAddG, we conducted a comparative analysis with several computational methods, including Rosetta_ddG, FoldX, ABACUS, and ESM-1b, another model that leverages the logit discrepancy between mutated and wild-type amino acids from MSAs to predict mutation effects.³⁶ To prevent data leakage, we employed another dataset containing the ΔT_m values of single-site mutations from previous experimental research. The five approaches demonstrated comparable overall prediction performance, while MSAddG stood out by uniquely identifying 16 stabilizing mutations that were overlooked by the other four methods, representing the

highest count of such identifications (Figure S6 and Table S6). Notably, MSAddG has also proven effective for inherently unstable proteins with T_m values below 45 °C. In addition, ESM-1b independently pinpointed 15 stabilizing mutations, further highlighting the substantial potential of these data-driven methods in mitigating the sampling bias in model-driven approaches.

Subsequently, we employed a combination of data-driven and model-driven techniques to assess the impact of all single-site mutations within ExiPAL on stability. Mutations yielding scores surpassing the respective cutoff values underwent thorough structural scrutiny to detect any potential biophysical pitfalls. In previous efforts to enhance enzyme thermostability, mutations that led to the creation of internal cavities, the disruption of hydrogen-bonding or electrostatic interactions, and the exposure of hydrophobic residues on the protein surface were typically excluded.^{6,20,37} Driven by the direct mechanism of denaturant-induced protein unfolding, we intentionally excluded mutations introducing internal cavities or disturbing internal polar interactions while permitting all residue substitutions on the enzyme's surface. After eliminating duplicate mutations (those of similar types identified by one or more methods at the same site), we experimentally characterized 62 mutants using α -hydroxyhippuric acid as a substrate (Table S1). Mutants displaying simultaneous increases in activity under 2 M Gnd·HCl, along with a higher

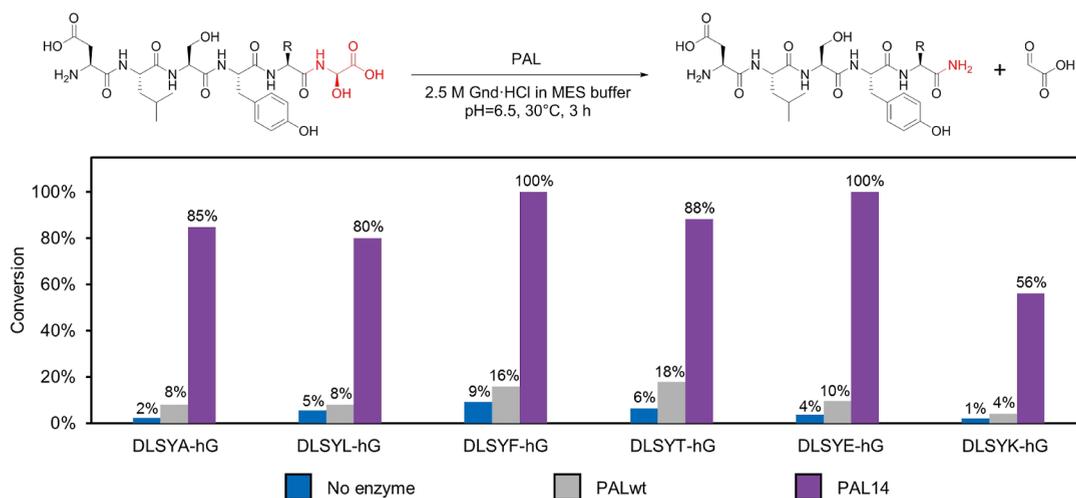


Figure 4. C-terminal amidation of the peptides by PALwt and PAL14 in the presence of 2.5 M Gnd-HCl. The peptide substrates were incubated with PAL at 30 °C and pH 6.5 for 3 h. Conversions were determined by integrating the peak areas of the peptidyl-hydroxyglycine and the produced peptide amide monitored by HPLC at 220 nm.²⁴ As the lysis reaction can proceed without an enzyme at a relatively slower rate, we included a control group with no enzyme added to correct for background product formation. Therefore, the improvement folds were calculated using the following formula: $(C_{\text{PAL14}} - C_{\text{PALwt}}) / (C_{\text{PALwt}} - C_{\text{no enzyme}})$, where C represents the conversion.

ratio of activities under 2 M compared with 0 M Gnd-HCl, were singled out as positive candidates (Figure 2A). In total, 17 beneficial mutations were identified. MSAddG, ESM-1b, ABACUS, Rosetta_ddG, and FoldX offered 4, 4, 5, 6, and 2 mutations, respectively, with some overlapping predictions (Figure 2B). Remarkably, MSAddG identified four beneficial mutations (R3Y, T45F, E137W, and R141W) that involved the replacement of exposed polar residues with aromatic ones, which was not commonly favored by the other four methods (Table S2). This underscores the significance of incorporating complementary approaches to identify mutation types that might be overlooked by prevailing computational tools.

In our pursuit of a more robust variant incorporating multiple mutations, we initially attempted to construct a mutant that encompassed all 17 mutations, but this variant was expressed as inclusion bodies. To overcome the obstacles posed by negative epistatic interactions, we turned to our previously proposed greedy accumulation strategy.²⁰ The mutations were divided into two distinct clusters based on their positions within the protein structure. Cluster 1 comprised 10 residues located within the region comprising the β -sheets and β -turns, while cluster 2 included the remaining 7 residues situated on the loops and α -helix around the active center. To initiate the accumulation process, we selected mutant F320Y (PAL1), which exhibited the highest activity under 2 M Gnd-HCl, as the starting point (Figure 2C). The subsequent rounds of accumulation involved systematically combining mutations from cluster 1 with F320Y. This process yielded the best hit, F320Y/I97L, which displayed enhanced activity under 2.4 M Gnd-HCl. This mutant served as the template for the second round of accumulation. By repeating this iterative process and introducing mutations E120P, T45F, A256I, Q238R, and N78T, we created the PAL7 mutant. The lyase activity of PAL7 under 2.4 M Gnd-HCl could not be further enhanced by the inclusion of additional mutations from cluster 1. At this point, we integrated the mutations associated with cluster 2 into PAL7, with the Gnd-HCl concentration for characterizing mutants raised to 2.8 M. This process spanned four additional rounds of accumulation, involving mutations N226E, E65T, T246P, and R141W,

resulting in the creation of the PAL11 mutant. Continuing our efforts, we further increased the Gnd-HCl concentration and introduced mutations E137W, V33P, and R188P, culminating in the generation of the PAL14 mutant. PAL14 exhibited lyase activity of 528 mU/mg under 0 M Gnd-HCl and 3.2 mU/mg under 3.2 M Gnd-HCl conditions. Despite additional attempts to incorporate the remaining mutations into PAL14, we observed a reduction in lyase activities under the elevated Gnd-HCl concentrations. Consequently, PAL14 was designated as the final mutant in our accumulation process.

2.2. Characterizing the Stability and Catalytic Durability of PAL14

With the optimized PAL14 variant in our possession, we conducted a comprehensive series of experiments to assess its stability. PAL14 exhibited an exceptional expression yield within *E. coli* cells, producing approximately 150 mg of purified enzyme per gram of dry cell weight, surpassing PALwt by over 4-fold. This substantial increase in yield, as evident in the gel electrophoresis analysis of the cell extracts, clearly indicated its improved stability in vivo (Figure 3A). Following purification, both PAL14 and PALwt underwent equilibrium denaturant titration analyses using Gnd-HCl and urea (Figures 3B and S3). In comparison to PALwt, PAL14 demonstrated notable enhancements in denaturant tolerance, with the half concentration ($C_{1/2}$) values increasing by 0.52 M for Gnd-HCl and 0.79 M for urea, respectively. Utilizing data from equilibrium urea titrations, we calculated the Gibbs free energy difference of folding (ΔG) for PALwt and PAL14. The ΔG value decreased from -1.3 to -1.9 kcal/mol, indicating improved thermodynamic stability of PAL14. Moreover, PAL14 exhibited a 6.5 °C higher apparent T_m compared to PALwt, suggesting its elevated kinetic stability (Figure 3C). The enhanced lyase activity of PAL14 under conditions of up to 3.6 M Gnd-HCl also aligned with the improved stability parameters (Figure 3D,E). Additionally, PAL14 demonstrated increased activity and total turnover numbers by more than 3-fold when subjected to other denaturants, including 6 M urea and 0.6 M thiourea (Table S4). These results collectively

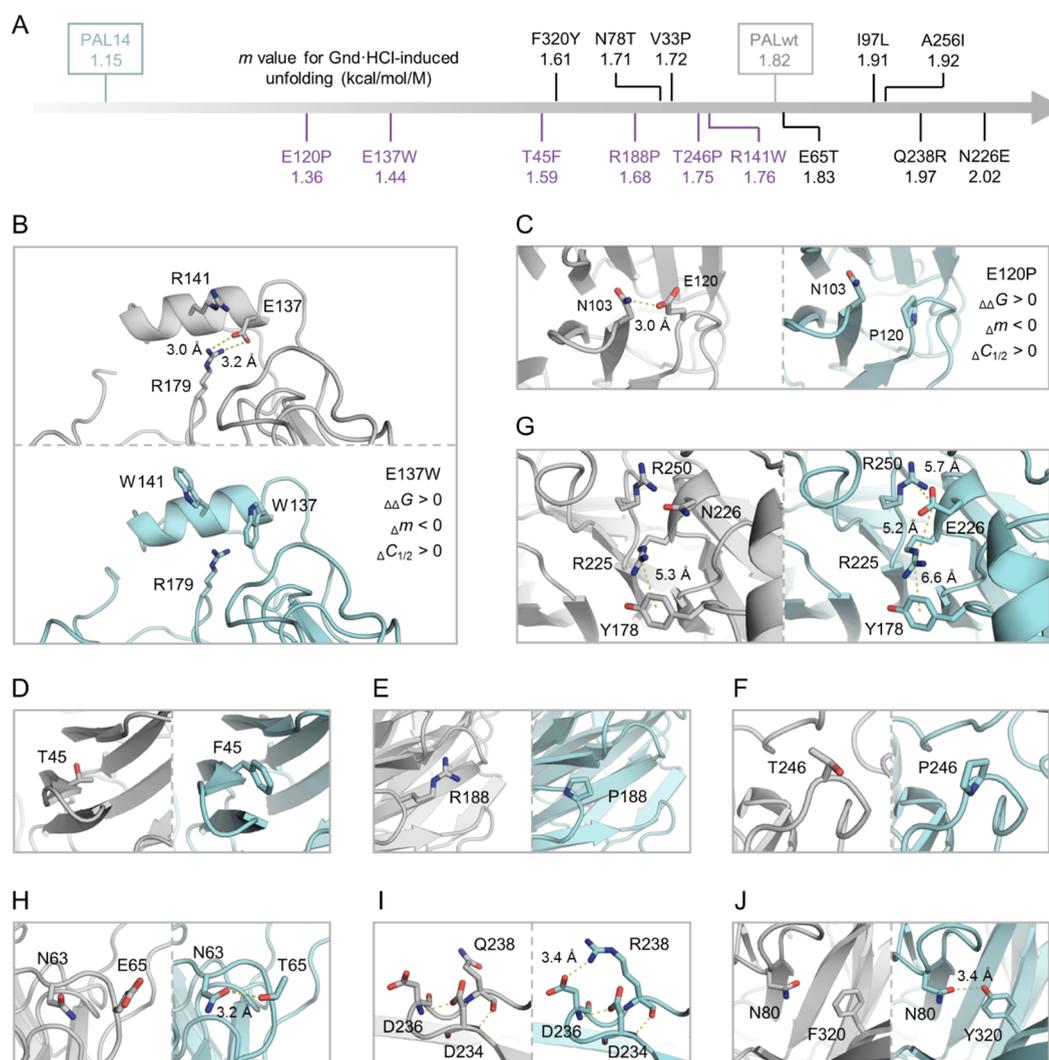


Figure 5. (A) m values of the fitted ΔG_U versus Gnd-HCl concentration lines for PAL variants. (B–J) Structural insights into the impacts of the introduced mutations in PAL14. Structural models of PALwt (grey) and PAL14 (cyan) were generated using AlphaFold2. Yellow dashed lines were used to represent hypothesized electrostatic and cation– π interactions in Figure 5G, whereas in the other figures, they denoted proposed hydrogen bonding interactions. The prediction confidence levels are illustrated in Figure S4.

underscore the significant increase in the structural robustness of ExiPAL achieved through the optimization process.

Furthermore, we conducted an extensive assessment of the catalytic durability of PAL14 using the peptide library DLSYX-hG (hG represents hydroxyglycine) as the substrate (Figure 4). In the presence of 2.5 M Gnd-HCl, a condition used in enzymatic protein conjugation and functionalization,²⁴ PALwt demonstrated only modest enhancement in peptide amide production compared to a control group without the enzyme. In contrast, PAL14 displayed a remarkable improvement in enzymatic conversion rates, achieving up to a 24-fold increase. These findings highlight PAL14's immense potential for facilitating the C-terminal functionalization of peptides that necessitate denaturing agents to maintain solubility. Moreover, we anticipate that the engineering strategy presented in this work will serve as a blueprint for developing enzymes more resistant to high denaturant concentrations, thereby promoting the application of enzymatic biomacromolecule synthesis and modification technologies.

2.3. Impacts of Mutations on PAL14's Denaturant Tolerance

In the context of the direct mechanism of denaturant-induced protein unfolding, the initial step involves the tight binding of denaturants to the protein surface through interactions with side chains and main-chain amides. Subsequently, denaturants may penetrate the protein's core, engaging with internal residues and facilitating protein unfolding.¹⁶ This mechanism emphasizes the importance of minimizing denaturant binding to the protein surface, in addition to the commonly studied changes in ΔG values, for improving the enzyme's tolerance to denaturants. In light of this, we delved into exploring biochemical parameters that could reflect the interaction between the enzyme and denaturants. Previous studies have reported that during equilibrium denaturant titrations, alterations in the slope (m value) of the fitted line of ΔG_U versus denaturant concentration correlate with changes in denaturant accessibility of proteins under varying pH conditions.^{38,39} To investigate this further, we performed equilibrium Gnd-HCl titrations for wild-type PAL and PAL mutants and calculated their m values (Figure S3 and Table

S2). The results indicate that m values decreased for a group of surface mutations, shifting from polar to hydrophobic, including T45F, E120P, E137W, R141W, R188P, and T246P (Figure SA–F). Remarkably, the E120P and E137W mutations substantially reduced the m value by 0.46 and 0.38 kcal/mol/M, respectively. Although these mutations exhibit an increase in ΔG values due to disruption of the native hydrogen bonds, they reduce the interactions between Gnd^+ ions and the protein surface, ultimately enhancing the $C_{1/2}$ values of $\text{Gnd}\cdot\text{HCl}$ -induced unfolding. In contrast, the introduction of a negatively charged residue in the Zn^{2+} cavity (N226E) increased the m value by 0.20 kcal/mol/M. This aligns consistently with an electrostatic theory, which suggests that electrostatic interactions between Gnd^+ ions and negatively charged residues may represent the predominant mechanism behind $\text{Gnd}\cdot\text{HCl}$ -induced destabilization.^{40,41} These results underscore the significance of surface engineering, particularly the reduction of surface negatively charged residues and the substitution of exposed polar residues with hydrophobic ones in enhancing enzyme tolerance to denaturing agents. Regarding the N226E mutation, structural analysis showed that the introduced side-chain carboxylate group may interact with R225 and R250 through electronic interactions (Figure 5G). The pK_a of the catalytic residue Y178 is proposed to be affected by R225 via cation– π interaction;²⁷ thus, the N226E mutation might impact the properties of catalytic residues rather than improving structural robustness.

In addition to surface engineering, several factors that are in line with protein thermostability engineering have also contributed to the heightened robustness of PAL14. The introduction of new hydrogen bonds (E65T, Q238R, and F320Y) likely induces a more constrained state within the loop regions (Figure 5H–J). Furthermore, the introduction of proline residues on the loops, attributed to the V33P, E120P, R188P, and T246P mutations, might play a pivotal role in reducing unfolding entropy, thereby favoring the native folding state of the enzyme.^{42,43} Additionally, the I97L and A256I mutations appear to refine the surrounding steric hindrance and hydrophobic interactions, resulting in a reduction in ΔG values. These findings highlight the collaborative efforts between model-driven and data-driven tools in redesigning both the surface and internal regions of the enzyme, ultimately enhancing PAL14's tolerance to high denaturant concentrations.

3. DISCUSSION

Over more than half a century, $\text{Gnd}\cdot\text{HCl}$ and urea have served as fundamental reagents for assessing protein thermodynamic stability, amassing a vast dataset that includes mutation-related stability changes.^{44,45} While these data have played a pivotal role in refining parameters within various computational tools, the intricate molecular interactions in the realm of protein chemistry extend beyond the scope of a concise energy function. Consequently, researchers have adopted strategies that combine various computational methods to mitigate the limitations of each approach. Alongside well-established model-driven methods, data-driven approaches leveraging evolutionary insights have attracted significant attention. For instance, EmCAST, which relies on an empirical potential related to C_α dihedral angle preferences, has been developed to predict stabilizing mutations, particularly those involving surface hydrophilic residues in monomeric proteins.⁴⁶ However, accurately assessing the interactions between surface

hydrophobic residues, especially aromatic ones, and solvent molecules remains a challenging task in computation protein redesign. It's imperative to acknowledge that urea and Gnd^+ ions can interact not only with hydrophilic but also hydrophobic side chains.⁴⁷ Recent computational modeling studies have advanced the notion that Gnd^+ ions may engage in water-assisted cation– π interactions with individual aromatic amino acids.⁴⁸ Nonetheless, due to the intrinsic chemical and topographical heterogeneity of the protein surface, these interactions are notably influenced by the hydration patterns of Gnd^+ ions and the relative orientations between Gnd^+ ions and the aromatic rings.^{49,50} In the case of ExiPAL, among 11 attempts involving the substitution of exposed hydrophilic residues with aromatic ones by Rosetta_ddG, FoldX, ABACUS, and ESM-1b, only one beneficial mutation (Q102Y) was identified. Regrettably, this mutation did not synergize with other substitutions to further enhance the enzyme robustness. In contrast, MSAddG pinpoints four beneficial mutations (R3Y, T45F, E137W, and R141W) of this type in 11 attempts, highlighting its effectiveness. Although this black box model may not directly provide a detailed energy function describing the interactions between aromatic residues and denaturants, it does shed light on unique insights into surface engineering patterns. Specifically, it emphasizes the usefulness of introducing exposed hydrophobic residues to reduce denaturant–protein interactions and enhance enzyme denaturant tolerance by decreasing the m value. Moreover, this work highlights the significance of collaboration between data-driven and model-driven methods in the comprehensive evaluation of all mutations, on both the protein surface and backbone, during enzyme stability engineering.

We should note that while PAL14 exhibits improvements in both thermodynamic and kinetic stability compared to wild-type PAL, a subset of accumulated mutations, such as T45F and E137W, results in an enhanced $C_{1/2}$ value and a decreased T_m value simultaneously. This observation aligns with previous research findings, which indicate that mutations can exert diverse effects on enzyme apparent robustness under varying conditions due to differences in underlying mechanisms.¹⁹ Consequently, to develop specialized machine learning methods tailored to enhance enzyme robustness across demanding conditions beyond high denaturant concentrations and elevated temperatures, the availability of experimental data under specific conditions becomes paramount. These conditions worthy of study may include extreme pH environments as well as high concentrations of salts, organic solvents, and surfactants. This highlights the critical need to establish high-throughput experimental platforms capable of consistently generating standardized data on protein mutants' stability and activity for model training. We anticipate that with high-quality data and advanced machine learning technologies, the integrated engineering strategy used in this study may be seamlessly extended to redesign proteins with enhanced functionality for a multitude of other challenging scenarios encountered in biocatalysis and beyond. Moreover, aside from single-site mutation prediction, artificial intelligence models may have the potential to optimize the accumulation paths, thereby reducing engineering time and expediting the development of robust biocatalysts.

4. METHODS

4.1. Software and Computational Methods

Chemical structures were crafted by using ChemBioDraw Ultra (version 14.0). To generate structure models of PAL variants, AlphaFold 2 (version 2.0.0) was employed with the settings as previously described,²⁶ and the resulting models were visualized using PyMOL (version 1.7). The final predictions were ranked based on their pLDDT values, and the top-ranked structural model was selected for further analysis. Energy calculations were conducted using Rosetta_ddG (version 2021.16.61629),⁷ FoldX (version 5),⁸ and ABACUS (first version),²⁹ with settings as described by Cui et al.²⁰ To utilize MSAddG, the MSA of PALwt was generated by searching the UniRef30 database and applying the same settings employed during the training of MSAddG. More details on the employment of MSAddG and ESM-1b are described in the [Supporting Information](#). The selection criteria for mutations chosen for structural inspection in each method were as follows: ABACUS energy < -3 a.e.u., ESM-1b score < -1.5, predicted $\Delta\Delta G$ < -1.5 kcal/mol (Rosetta_ddG and FoldX) or -1.0 kcal/mol (MSAddG).

4.2. Site-Directed Mutagenesis

The plasmids for the mutants were generated using a QuickChange site-directed mutagenesis kit (Agilent). After DNA amplification, the PCR products were subjected to *DpnI* (New England Biolabs) treatment to digest the original DNA template. These products were subsequently transformed into *E. coli* TOP10 competent cells. The mutated protein sequences were confirmed through DNA sequencing.

4.3. Expression and Purification of PAL

The expression strain *E. coli* BL21(DE3), which harbored the vector derived from pET28b-PALwt, was cultured in 50 mL of LB broth medium supplemented with 50 mg/L kanamycin sulfate at 37 °C and 180 rpm. When the optical density at 600 nm reached approximately 0.8, protein expression was induced by adding 0.5 mM IPTG. Subsequently, the cells were cultured at 30 °C and 180 rpm for 16 h. The cells were then harvested via centrifugation (14,000g, 10 min, 4 °C), resuspended in a lysis buffer (containing 50 mM Tris, 200 mM NaCl, and 20 mM imidazole, pH 8.0), and subjected to sonication for cell lysis. After centrifugation (14,000g, 60 min, 4 °C), the cell extract was loaded onto a 5 mL HisTrap HP column (GE Healthcare). His-tagged PAL was eluted using an elution buffer (containing 50 mM Tris, 200 mM NaCl, and 300 mM imidazole, pH 8.0). The buffer was subsequently exchanged with a storage buffer (50 mM Tris, 0.1 mM ZnCl₂, and 0.1 mM CaCl₂, pH 8.0) via ultrafiltration (10 kDa Amicon centrifugal filter, Millipore). The PAL variants were concentrated to a final concentration of 4 mg/mL and stored at -20 °C until use.

4.4. Determination of Thermodynamic and Kinetic Stability of PAL

The experiments were conducted by using the Prometheus NT.48 system (NanoTemper Technologies, Munich, Germany). The fraction of folded and unfolded proteins was determined based on the ratio of fluorescence changes at 350 and 330 nm. To determine the T_m values, a sample of the protein solution at 0.2 mg/mL (prepared by diluting the storage protein solution with 200 mM MES buffer, pH 6.5) was subjected to a heating process ranging from 25 °C to 85 °C at a heating rate of 1.0 °C/min. For equilibrium denaturation titrations, 24 samples of the protein solution at 0.2 mg/mL (prepared by diluting the storage protein solution with 200 mM MES buffer containing 0–4 M Gnd-HCl or 0–8 M urea, pH 6.5) were incubated at 4 °C for 18 h equilibrium, and then fluorescence changes were measured at 25 °C.

4.5. Determination of Lyase Activity with α -Hydroxyhippuric Acid

In a 1.5 mL tube, 80 μ L of a reaction buffer consisting of 200 mM MES and varying concentrations of denaturants (0 to 5 M Gnd-HCl, 0 to 7.5 M urea, or 0 to 0.75 M thiourea) at pH 6.5, along with 10 μ L of a 10 mM α -hydroxyhippuric acid solution (prepared in 200 mM MES buffer, pH 6.5), and 10 μ L of a 4 mg/mL PAL solution, were

combined and mixed thoroughly. The enzymatic reactions were carried out at 30 °C for 60 min and then terminated by adding 100 μ L of a 2 M HCl solution. The formation of benzamide was monitored by HPLC at 254 nm. One unit (U) of enzyme activity was defined as the amount of the enzyme capable of producing 1 μ mol of benzamide per minute.

4.6. Determination of Lyase Activity with Peptides

Initially, the peptides DLSYKG, DLSYEG, DLSYTG, DLSYFG, DLSYLG, and DLSYAG were hydroxylated using PHM, following established protocols.²⁴ Then, in a 1.5 mL tube, 100 μ L of a reaction buffer composed of 200 mM MES and 5 M Gnd-HCl at pH 6.5, along with 80 μ L of a 0.125 mM DLSYX-(hG) solution (prepared in 200 mM MES buffer at pH 6.5) and 20 μ L of a 4 mg/mL PAL solution were combined and mixed thoroughly. The reactions were conducted at 30 °C for 3 h, and the samples were quenched by adding an equal volume of 2 M HCl solution. Conversion rates were estimated by integrating the peak areas of DLSYX-NH₂ and DLSYX-(hG) monitored by HPLC at 220 nm.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacsau.3c00792>.

Comparative structural analysis, apparent melting temperatures, equilibrium denaturation titrations, prediction confidence levels, experimental characterization of mutants with single-site mutations, locations, predicted $\Delta\Delta G$ or scores, stability parameters of equilibrium Gnd-HCl titrations, and ΔT_m values of 17 mutations selected for accumulation, experimental characterization of mutants with accumulated mutations, activity and TTN of PAL, computer hardware for model training and software for mutation screening, construction of the dataset for training and testing, training of a machine learning model MSAddG, employment of a machine learning model ESM-1b, comparative analysis of MSAddG with several computational methods, code and data availability, chemical reagents, synthetic peptides, genes, vectors, and strains, protein sequences, and detection of reactions (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Bian Wu – AIM Center, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China; orcid.org/0000-0002-6524-2049; Email: wub@im.ac.cn

Authors

Tong Zhu – AIM Center, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China; orcid.org/0000-0001-7857-8564

Jinyuan Sun – AIM Center, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China; orcid.org/0000-0003-3456-0332

Hua Pang – AIM Center, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

Complete contact information is available at: <https://pubs.acs.org/10.1021/jacsau.3c00792>

Author Contributions

B.W. conceived and supervised the study. T.Z. conducted the biochemical experiments. H.P. was responsible for constructing

and expressing the mutants. J.S. developed the MSAddG software. T.Z. drafted the manuscript, which was reviewed, edited, and approved by all authors.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (2021YFC2103900), the National Natural Science Foundation of China (31822002, 31870055, and 32170033), the Strategic Priority Research Program (XDA24020101 and XDPB1801), the Key Research Program of Frontier Sciences (ZDBS-LY-SM014), the Biological Resources Program (KFJ-BRP-009 and KFJ-BRP-017-58) from the Chinese Academy of Sciences, the Informatization Plan of Chinese Academy of Sciences (CAS-WX2021SF-0111), the Youth Innovation Promotion Association CAS (2022086), and the Postdoctoral Fellowship Program of CPSF (GZC20232929). We extend our sincere gratitude to Zheng Fan from the Public Technology Service Center at the Institute of Microbiology, CAS, for providing guidance in equilibrium denaturant titrations.

REFERENCES

- (1) Wu, S.; Snajdrova, R.; Moore, J. C.; Baldenius, K.; Bornscheuer, U. T. Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angew. Chem., Int. Ed.* **2021**, *60*, 88–119.
- (2) Sheldon, R. A.; Woodley, J. M. Role of Biocatalysis in Sustainable Chemistry. *Chem. Rev.* **2018**, *118*, 801–838.
- (3) Goldenzweig, A.; Fleishman, S. J. Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* **2018**, *87*, 105–129.
- (4) Bednar, D.; Beerens, K.; Sebestova, E.; Bendl, J.; Khare, S.; Chaloupkova, R.; Prokop, Z.; Brezovsky, J.; Baker, D.; Damborsky, J. FireProt: Energy- and Evolution-based Computational Design of Thermostable Multiple-point Mutants. *PLoS Comput. Biol.* **2015**, *11*, No. e1004556.
- (5) Kazlauskas, R. Engineering More Stable Proteins. *Chem. Soc. Rev.* **2018**, *47*, 9026–9045.
- (6) Wijma, H. J.; Floor, R. J.; Jekel, P. A.; Baker, D.; Marrink, S. J.; Janssen, D. B. Computationally Designed Libraries for Rapid Enzyme Stabilization. *Protein Eng., Des. Sel.* **2014**, *27*, 49–58.
- (7) Park, H.; Bradley, P.; Greisen, P.; Liu, Y.; Mulligan, V. K.; Kim, D. E.; Baker, D.; DiMaio, F. Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* **2016**, *12*, 6201–6212.
- (8) Guerois, R.; Nielsen, J. E.; Serrano, L. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. *J. Mol. Biol.* **2002**, *320*, 369–387.
- (9) Cui, Y.; Sun, J.; Wu, B. Computational Enzyme Redesign: Large Jumps in Function. *Trends Chem.* **2022**, *4*, 409–419.
- (10) Nguyen, H. L. T.; Nguyen, T. T.; Vu, Q. T.; Le, H. T.; Pham, Y.; Trinh, P. L.; Bui, T. P.; Phan, T. N. An Efficient Procedure for the Expression and Purification of HIV-1 Protease from Inclusion Bodies. *Protein Expression Purif.* **2015**, *116*, 59–65.
- (11) Toplak, A.; Nuijens, T.; Quaedflieg, P. J. L. M.; Wu, B.; Janssen, D. B. Peptilgase, an Enzyme for Efficient Chemoenzymatic Peptide Synthesis and Cyclization in Water. *Adv. Synth. Catal.* **2016**, *358*, 2140–2147.
- (12) Poulsen, J. W.; Madsen, C. T.; Young, C.; Poulsen, F. M.; Nielsen, M. L. Using Guanidine-Hydrochloride for Fast and Efficient Protein Digestion and Single-step Affinity-purification Mass Spectrometry. *J. Proteome Res.* **2013**, *12*, 1020–1030.
- (13) Canchi, D. R.; García, A. E. Cosolvent Effects on Protein Stability. *Annu. Rev. Phys. Chem.* **2013**, *64*, 273–293.
- (14) Frank, H. S.; Franks, F. Structural Approach to Solvent Power of Water for Hydrocarbons: Urea as a Structure Breaker. *J. Chem. Phys.* **1968**, *48*, 4746–4757.
- (15) O'Brien, E. P.; Dima, R. I.; Brooks, B.; Thirumalai, D. Interactions between Hydrophobic and Ionic Solutes in Aqueous Guanidinium Chloride and Urea Solutions: Lessons for Protein Denaturation Mechanism. *J. Am. Chem. Soc.* **2007**, *129*, 7346–7353.
- (16) Jha, S. K.; Marqusee, S. Kinetic Evidence for a Two-Stage Mechanism of Protein Denaturation by Guanidinium Chloride. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 4856–4861.
- (17) Musil, M.; Konegger, H.; Hon, J.; Bednar, D.; Damborsky, J. Computational Design of Stable and Soluble Biocatalysts. *ACS Catal.* **2019**, *9*, 1033–1054.
- (18) Gazi, R.; Maity, S.; Jana, M. Conformational Features and Hydration Dynamics of Proteins in Cosolvents: A Perspective from Computational Approaches. *ACS Omega* **2023**, *8*, 2832–2843.
- (19) Jones, B. J.; Lim, H. Y.; Huang, J.; Kazlauskas, R. J. Comparison of Five Protein Engineering Strategies for Stabilizing an α/β -Hydrolase. *Biochemistry* **2017**, *56*, 6521–6532.
- (20) Cui, Y.; Chen, Y.; Liu, X.; Dong, S.; Tian, Y.; Qiao, Y.; Mitra, R.; Han, J.; Li, C.; Han, X.; Liu, W.; Chen, Q.; Wei, W.; Wang, X.; Du, W.; Tang, S.; Xiang, H.; Liu, H.; Liang, Y.; Houk, K. N.; Wu, B. Computational Redesign of a PETase for Plastic Biodegradation under Ambient Condition by the GRAPE Strategy. *ACS Catal.* **2021**, *11*, 1340–1350.
- (21) Prigge, S. T.; Mains, R. E.; Eipper, B. A.; Amzel, L. M. New Insights into Copper Monooxygenases and Peptide Amidation: Structure, Mechanism and Function. *Cell. Mol. Life Sci.* **2000**, *57*, 1236–1259.
- (22) Delgado-Prudencio, G.; Possani, L. D.; Becerril, B.; Ortiz, E. The Dual α -Amidation System in Scorpion Venom Glands. *Toxins* **2019**, *11*, 425.
- (23) Shaw, A. C. Peptidyl α -Hydroxyglycine α -Amdating Lyases. U.S. Patent 9,096,843 B2.
- (24) Li, R.; Schmidt, M.; Zhu, T.; Yang, X.; Feng, J.; Tian, Y.; Cui, Y.; Nuijens, T.; Wu, B. Traceless Enzymatic Protein Synthesis without Ligation Sites Constraint. *Natl. Sci. Rev.* **2022**, *9*, nwab158.
- (25) Zhu, T.; Cui, Y.; Geng, W.; Liu, G.; Jiang, H.; Li, R.; Wu, B. Creating an Unusual Glycine-Rich Motif in a Peptide Amidase Leads to Versatile Protein C-Terminal Traceless Functionalization. *ACS Catal.* **2022**, *12*, 8019–8026.
- (26) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
- (27) Chufán, E. E.; De, M.; Eipper, B. A.; Mains, R. E.; Amzel, L. M. Amidation of Bioactive Peptides: the Structure of the Lyase Domain of the Amidating Enzyme. *Structure* **2009**, *17*, 965–973.
- (28) Tokuriki, N.; Tawfik, D. S. Stability Effects of Mutations and Protein Evolvability. *Curr. Opin. Struct. Biol.* **2009**, *19*, 596–604.
- (29) Xiong, P.; Wang, M.; Zhou, X. Q.; Zhang, T. C.; Zhang, J. H.; Chen, Q.; Liu, H. Y. Protein Design with a Comprehensive Statistical Energy Function and Boosted by Experimental Selection for Foldability. *Nat. Commun.* **2014**, *5*, 5330.
- (30) Bu, Y.; Cui, Y.; Peng, Y.; Hu, M.; Tian, Y. E.; Tao, Y.; Wu, B. Engineering Improved Thermostability of the GH11 Xylanase from *Neocallimastix patriciarum* via Computational Library Design. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 3675–3685.
- (31) Mu, Q.; Cui, Y.; Tian, Y. E.; Hu, M.; Tao, Y.; Wu, B. Thermostability Improvement of the Glucose Oxidase from *Aspergillus Niger* for Efficient Gluconic Acid Production via Computational Design. *Int. J. Biol. Macromol.* **2019**, *136*, 1060–1068.
- (32) Wheeler, T. J.; Clements, J.; Finn, R. D. Skylign: A Tool for Creating Informative, Interactive Logos Representing Sequence

Alignments and Profile Hidden Markov Models. *BMC Bioinf.* **2014**, *15*, 7.

(33) Trudeau, D. L.; Kaltenbach, M.; Tawfik, D. S. On the Potential Origins of the High Stability of Reconstructed Ancestral Proteins. *Mol. Biol. Evol.* **2016**, *33*, 2633–2641.

(34) Greener, J. G.; Kandathil, S. M.; Moffat, L.; Jones, D. T. A Guide to Machine Learning for Biologists. *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 40–55.

(35) Nikam, R.; Kulandaisamy, A.; Harini, K.; Sharma, D.; Gromiha, M. M. ProThermDB: Thermodynamic Database for Proteins and Mutants Revisited after 15 Years. *Nucleic Acids Res.* **2021**, *49*, D420–D424.

(36) Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; Fergus, R. Biological Structure and Function Emerge from Scaling Unsupervised Learning to 250 Million Protein Sequences. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2016239118.

(37) Wu, B.; Wijma, H. J.; Song, L.; Rozeboom, H. J.; Poloni, C.; Tian, Y.; Arif, M. I.; Nuijens, T.; Quaedflieg, P. J. L. M.; Szymanski, W.; Feringa, B. L.; Janssen, D. B. Versatile Peptide C-terminal Functionalization via A Computationally Engineered Peptide Amidase. *ACS Catal.* **2016**, *6*, 5405–5414.

(38) Scholtz, J. M.; Grimsley, G. R.; Pace, C. N. Chapter 23 Solvent Denaturation of Proteins and Interpretations of the m Value. *Methods Enzymol.* **2009**, *466*, 549–565.

(39) Amsdr, A.; Noudeh, N. D.; Liu, L.; Chalikian, T. V. On Urea and Temperature Dependences of m-Values. *J. Chem. Phys.* **2019**, *150*, 215103.

(40) Meuzelaar, H.; Panman, M. R.; Woutersen, S. Guanidinium-Induced Denaturation by Breaking of Salt Bridges. *Angew. Chem., Int. Ed.* **2015**, *54*, 15255–15259.

(41) Negi, I.; Jangra, R.; Gharu, A.; Trant, J. F.; Sharma, P. Guanidinium-Amino Acid Hydrogen-bonding Interactions in Protein Crystal Structures: Implications for Guanidinium-induced Protein Denaturation. *Phys. Chem. Chem. Phys.* **2023**, *25*, 857–869.

(42) Fu, H.; Grimsley, G. R.; Razvi, A.; Scholtz, J. M.; Pace, C. N. Increasing Protein Stability by Improving Beta-turns. *Proteins* **2009**, *77*, 491–498.

(43) Watanabe, K.; Suzuki, Y. Protein Thermostabilization by Proline Substitutions. *J. Mol. Catal. B: Enzym.* **1998**, *4*, 167–180.

(44) Tanford, C. Isothermal Unfolding of Globular Proteins in Aqueous Urea Solutions. *J. Am. Chem. Soc.* **1964**, *86*, 2050–2059.

(45) Tanford, C.; Pain, R. H.; Otchin, N. S. Equilibrium and Kinetics of the Unfolding of Lysozyme (Muramidase) by Guanidine Hydrochloride. *J. Mol. Biol.* **1966**, *15*, 489–504.

(46) Rothfuss, M. T.; Becht, D. C.; Zeng, B.; McClelland, L. J.; Yates-Hansen, C.; Bowler, B. E. High-Accuracy Prediction of Stabilizing Surface Mutations to the Three-Helix Bundle, UBA(1), with EmCAST. *J. Am. Chem. Soc.* **2023**, *145*, 22979–22992.

(47) Ajayi, S.; Asakereh, I.; Rezasoltani, H.; Davidson, D.; Khajehpour, M. Does Urea Preferentially Interact with Amide Moieties or Nonpolar Sidechains? A Question Answered Through a Judicious Selection of Model Systems. *ChemPhysChem* **2023**, *24*, No. e202200731.

(48) Balamurugan, K.; Prakash, M.; Subramanian, V. Theoretical Insights into the Role of Water Molecules in the Guanidinium-Based Protein Denaturation Process in Specific to Aromatic Amino Acids. *J. Phys. Chem. B* **2019**, *123*, 2191–2202.

(49) Heyda, J.; Okur, H. I.; Hladílková, J.; Rembert, K. B.; Hunn, W.; Yang, T.; Dzubiella, J.; Jungwirth, P.; Cremer, P. S. Guanidinium can both Cause and Prevent the Hydrophobic Collapse of Biomacromolecules. *J. Am. Chem. Soc.* **2017**, *139*, 863–870.

(50) Cui, D.; Ou, S.; Patel, S. Protein Denaturants at Aqueous-Hydrophobic Interfaces: Self-Consistent Correlation between Induced Interfacial Fluctuations and Denaturant Stability at the Interface. *J. Phys. Chem. B* **2015**, *119*, 164–178.