ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Understanding the Stabilization Mechanism of a Thermostable Mutant of Hygromycin B Phosphotransferase by Protein Sector-Guided Dynamic Analysis

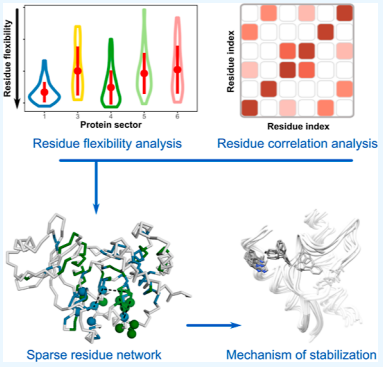
Zixiao Xue* and Shu Quan*

Cite This: *ACS Omega* 2023, 8, 25739−25748

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Point mutations can exert beneficial effects on proteins, including stabilization. The stabilizing effects of mutations are typically attributed to changes in free energy and residue interactions. However, these explanations lack detail and physical insights, which hinder the mechanistic study of protein stabilization and prevent accurate computational prediction of stabilizing mutations. Here, we investigate the physical mechanism underlying the enhanced thermostability of a Hygromycin B phosphotransferase mutant, Hph5. We find that the unpredictable mutation A118V induces rotation of F199, allowing it to establish an aromatic−aromatic interaction with W235. In contrast, the predictable mutation T246A acts through static hydrophobic interactions within the protein core. These discoveries were accelerated by a residue-coevolution-based theory, which links mutational effects to stability-associated local structures, providing valuable guidance for mechanistic exploration. The established workflow will benefit the development of accurate stability prediction programs and can be used to mine a protein stability database for undiscovered physical mechanisms.



INTRODUCTION

Engineering protein stability is crucial for a wide range of applications, from developing biocatalysts to designing therapeutics.[1−3] In general, the stabilization afforded by beneficial mutations can be explained by the changes in free energy and residue interactions caused by the mutations.[4,5] However, these explanations are usually not detailed and rarely take into account protein dynamics. This hinders the exploration of protein folding and the development of high-precision stability prediction programs.[6] To advance the study of protein folding and stability prediction, in-depth physical insights into protein stabilization are required.

Molecular dynamics simulation (MDS) is a promising tool to understand the physical mechanisms of protein stabilization because it describes the atomic movements of proteins using empirical potentials.[7] MDS can reveal the dynamic process by which spatially distant mutations exert nonadditive stabilizing effects.[8] The basic readout of MDS is protein backbone flexibility, which helps uncover targets for stability engineering,[9] based on the observation that high flexibility often associates with low stability.[10] Coupling flexibility analysis with free-energy calculations, researchers reveal mutation-induced low-energy conformations in protein mutants, explaining their stabilizing effects from the perspective of global and secondary structure levels.[11−13] These analyses are often computationally intense (μs scale for global motions). In contrast, ns-scale MDS can already describe residue dynamics.[14] By analyzing the dynamic behavior of residues influenced by amino acid mutations, we may deepen our mechanistic understanding of stabilizing effects afforded by mutations.

To boost the potential of MDS in studying protein stabilization, we propose to use protein sector analysis[15] as a preprocessing step to identify stability-associated residue networks for MDS analysis. This residue-coevolution-based analysis divides residues into sectors (residue groups) based on the similarity of coevolutionary features. It allows definition of residue clusters which are most likely associated with protein stability and has been successfully applied to suggest targets for stability engineering.[16] By combining residue-coevolution with simulation analyses, researchers have identified interfacial residues that maintain the stability of protein complexes[17] and designed protein scaffolds with robust stability and function.[18] These results suggest that the combination of protein sector analysis and MDS could be promising in studying protein stabilization.

In this work, we combined protein sector analysis and MDS to study stabilization mechanisms using a thermostable mutant of Hygromycin-B 4-*O*-kinase (Hph[19]) as the target protein. This mutant, called Hph5,[20] was isolated in a directed evolution experiment and improved the thermotolerance of Hph by

Received: January 18, 2023
Accepted: June 12, 2023
Published: July 13, 2023

approximately 16 °C. It contains five amino acid substitutions: D20G, A118V, S225P, Q226L, and T246A. We focus on Hph5 because not all of its stabilizing mutations can be easily predicted, even a dominant stabilizing mutation, A118V. Thus, we aim to provide a theoretical understanding of the stabilization mechanism of Hph5 from a protein dynamic perspective.

Assisted by protein sector analysis, we defined the sparse residue networks surrounding the two dominant stabilizing mutations of Hph5 and used this information to guide MDS analysis. Through timeline-energy comparisons, we revealed key residue interactions influenced by A118V and T246A and further investigated their roles in stabilizing Hph5 via mutational and dynamic analyses. Our findings suggest that the T246A mutation induces the formation of a hydrophobic network within the protein core, enhancing the connection between α-helices 6 and 7. The A118V mutation induces residue movement of F199, followed by the establishment of an aromatic−aromatic interaction between F199 and W235 of α-helix 6. This dynamic change cannot be identified based upon static structural analysis. This work demonstrates a workflow for concentrated MDS focusing on stability-associated local structures to underlie the effects of point mutations on stability.

MATERIALS AND METHODS

Strains, Plasmids, and Growth Conditions. All primers, strains, and plasmids used in this study are listed in Table S6. *Escherichia coli* strain Trans1 T1 (TransGen Biotech) and BL21 (DE3) were utilized for cloning and protein purification, respectively. Unless otherwise indicated, all strains were grown at 37 °C with shaking (220 rpm) in LB-Lennox medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) supplemented with kanamycin (50 μg/mL). Various mutants of *hph5* (on plasmid pET28b-*6his-sumo-hph5*) were created using QuikChange Site-directed Mutagenesis.

Protein Expression and Purification. For protein expression, overnight cultures of *E. coli* BL21 (DE3) cells harboring a pET28b-based vector were diluted into 600 mL of LB-Lennox medium and grown at 37 °C until an OD_{600} of 0.6 was reached. The cytoplasmic expression of Hph5 mutants was induced by 0.4 mM isopropyl β-D-thiogalactoside (IPTG). The culture was grown further at 16 °C for ∼16 h with shaking (220 rpm). After centrifugation at 4 °C for 30 min, the cell pellet was resuspended in chilled lysis buffer (20 mM Tris, pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$), supplemented with a 1× protein inhibitor cocktail (Sigma).

For protein purification, the cells were mechanically disrupted at 700 bar and the lysate was cleared through centrifugation (10,000*g*) at 4 °C for 1 h. The supernatant was incubated with 1.5 mL of His-Tag purification resin (Roche) at 4 °C for 2 h with gentle rotation. The resin was washed with chilled lysis buffer and resuspended in 10 mL of lysis buffer. To remove the 6His-SUMO tag, protease ULP1 was added for on-resin digestion, and the mixture was incubated at 4 °C for 4 h with gentle rotation. Hph5 mutants were collected in the flow-through, concentrated by diafiltration (Amicon Ultra-15, Millipore) and stored at −80 °C. Protein purity and concentration were determined by SDS-PAGE analysis and absorption at 280 nm, respectively.

Differential Scanning Fluorimetry. For thermally induced protein unfolding and aggregation, various Hph5 mutants were diluted in protein purification buffer (25 mM Tris, pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$) to a final concentration of 1 mg/mL and then filled in glass capillaries (Nanotemper) by the siphon effect. The protein-loaded capillaries were placed in the Prometheus NT.48 instrument (Nanotemper) and subjected to a thermal gradient from 20 to 80 °C (heating rate of 1 °C/min). The tryptophan and tyrosine fluorescence (at 330 and 350 nm) and light scattering (at 385 nm) were monitored continuously. The ratio of fluorescent signals at 330 and 350 nm (F_{350}/F_{330}) and light scattering signals were normalized to the range 0−100% to represent the percentages of unfolded and aggregated proteins, respectively (Figure S1).

Protein Sector Analysis. The multiple sequence alignment (MSA) of Hph contains 1272 sequences (rows) and 341 residue positions (columns). At residue position i (j), we labeled the substitution event that mutates the amino acid in row k to that in row l as X_{ikl} (X_{jkl}). These two substitution events can be scored using the BLOSUM50[21] substitution matrix to obtain the respective substitution scores. Considering substitution events between all rows, the residue correlation coefficient $r_{i,j}$ (representing coevolutionary signal between positions i and j) can be calculated with eq 1[16]

$$r_{i,j} = \left[\frac{2}{M(M-1)}\right]\sum_{k=1}^{M-1}\left(\sum_{l=k+1}^{M}[(X_{ikl} - \langle X_i\rangle)(X_{jkl} - \langle X_j\rangle)]/(\sigma_i\sigma_j)]\right) \quad (1)$$

X_{ikl} (X_{jkl}) is the substitution score between rows k and l at position i (j). $\langle X_i\rangle$ ($\langle X_j\rangle$) is the mean of all substitution scores at position i (j), and σ_i (σ_j) is the standard deviation. M is the number of MSA sequences. The correlation coefficients $r_{i,j}$ were calculated with the MATLAB code, and the resulting values were used to build an RCA matrix of size 341 × 341 (341 is the residue number of Hph). After eigen-decomposition of the RCA matrix, 341 eigenvalues ($\lambda_1, \lambda_2, \lambda_3, ..., \lambda_{341}$) and 341 eigenvectors ($\nu_1, \nu_2, \nu_3, ..., \nu_{341}$) were generated. The eigenvalues are ranked by magnitude and correspond to the order of eigenvectors. Each eigenvector also contains 341 vector values. See Figure S3 for a schematic representation.

Top-ranked eigenvalues and eigenvectors contain effective coevolutionary signals. The eigenvectors (ν_2, ν_3, ν_4) were used as the parameters to assign protein sectors because they can reconstitute an RCA matrix containing principle coevolutionary features (Figure S4, top right). The value of ν_1 was not used because its signals may arise from residue conservation in different protein families rather than coevolution.[15,16] The eigen-decomposition and reconstitution of the RCA matrix were performed by functions in R base.

Only residues that show effective coevolutionary signals were included in the protein sector analysis. To remove residues with weak signals, we set thresholds using value distribution of random eigenvectors (ν_2, ν_3, ν_4). These random eigenvectors were eigen-decomposed from 10 mock RCA matrices, which were built with 10 randomized MSAs (shuffling the order of amino acids at each residue position). The random values of eigenvector ν_2, ν_3, and ν_4 were individually plotted as a histogram and fitted by normal distributions (Figure S5). For a residue with weak coevolutionary signals, its eigenvector values all fall within the boundary of the three normal distributions. Finally, protein sectors were assigned to 230 Hph residues with effective coevolutionary signals using k-mean clustering (R package "factoextra"). The separation of sector residues was examined using a three-dimension scatter plot (Figure 2A).

https://doi.org/10.1021/acsomega.3c00373
ACS Omega 2023, 8, 25739−25748

Molecular Dynamics Simulation (MDS). MDSs were performed on crystal structures[22] of Hph WT (PDB ID: 3w0s) and Hph5 (PDB ID: 3w0n) using the Amber ff99SB-ILDN[23] force field with GROMACS[24] 2020.1. The PDB files were cleaned by removing water and ligands. The protein was placed in a cubic box with a minimal distance of 20 Å from the edges of the box. The box was filled with water molecules (SPC model) to create a solvent environment. Then a small portion of water molecules were replaced by 21 and 20 Na^+ ions to neutralize the charges of Hph WT and Hph5, respectively. About 120 water molecules were further replaced by Na^+ and Cl^- ions to create a 150 mM NaCl environment. Energy minimization was achieved by using the steepest descent algorithm (stop when maximum force <10 kJ/mol, <50,000 steps). LINCS algorithm was applied to constraint H-bonds. The system was first equilibrated in the NVT ensemble (constant atom number N, volume V, and temperature T) for 100 ps to gradually increase the simulation temperature to 410 K and further equilibrated in the NPT ensemble (constant atom number N, pressure P, and temperature T) for 1 ns to equilibrate the pressure at 1 bar. In the NPT ensemble, temperature was regulated using a velocity rescaling thermostat[25] with a coupling constant of 0.1 ps, and pressure was regulated using the Parrinello−Rahman approach[26] with a coupling constant of 2.0 ps. Finally, productive MDSs were run in the NPT ensemble for 30 ns at a simulation temperature of 410 K, with trajectory frames recorded every 5 ps.

The MDS of 410 K was used for dynamic analyses because it can effectively discriminate between the differences in dynamics between Hph WT and Hph5, compared with the MDS of 310 K (Figure S7). The temperature of 410 K did not cause large conformational changes within the ns-scale simulation, as evidenced by secondary structure integrity and residue contact pattern (Figures S8 and S9).

Analysis of MDS Trajectories. Python module MDAnalysis[27] was used for root mean square deviation (RMSD) and root mean square fluctuation (RMSF) analysis. The RMSD and RMSF were calculated from backbone atoms and Cα of individual residues, respectively, with the initial structure of MDS set as the reference for calculations. The average values of RMSD and RMSF from two independent simulations were used to represent the global protein and individual residue flexibility, respectively (Figure S7).

In each MDS trajectory, the protein structures were output at a 0.05 ns interval using MDAnalysis to generate 601 PDB files for subsequent timeline trace and violin plot analyses.

For energy analysis of residue interactions, the interaction energy was calculated using Rosetta Energy Function 2015[28] (REF15) for each PDB file. The energy values were plotted against the simulation time to generate timeline-energy traces (Figures 3, S12, S13). For measurement of Cβ distances (Figure S11), the distances of paired Cβ atoms were calculated using the calcDistance() function in the Python module ProDy.[29] For measurement of dihedral angles of residue F199 (Figure 4B), the angle phi or psi was calculated using the Dihedral() function in MDAnalysis. The four atoms defining angle phi are: C^{198}−N^{199}−Cα^{199}−C^{199}, while angle psi is defined by: N^{199}−Cα^{199}−C^{199}−N^{200}.

Alchemical Free-Energy Calculations. The free-energy calculation was performed on folded structures of Hph WT and Hph5 using the Amber ff99SB-ILDN force field with GROMACS 2021.4. The equilibration run is similar to the MDS workflow, with the differences described below. Mutations in Hph5 were created using pmx[30] with a single topology approach. The single topology can describe the state transition of the mutated amino acid using a parameter λ. λ represents a certain path for the transformation between the wild-type state (represented by $\lambda = 0$) and the mutated state (represented by $\lambda = 1$). The wild-type Hph protein ($\lambda = 0$) was placed in a 150 mM NaCl environment, with a buffer width of 15 Å from the edge of the cubic box. The system was first equilibrated in an NPT ensemble for 1 ns, at a simulation temperature of 298 K and a pressure of 1 bar. The equilibration run lasted 10 ns in the NPT ensemble. The simulation trajectory was output at 0.1 ns intervals to generate 100 snapshot structures. The mutated Hph protein ($\lambda = 1$) was simulated using the same process.

Alchemical transformations for the forward (λ: 0 → 1) and backward directions (λ: 1 → 0) were performed with 100 snapshot structures separately, using the non-equilibration run. The free-energy difference ΔG_{folded} for the folded structure is calculated by Bennett acceptance ratio (BAR) estimator,[31] after using thermodynamic integration to collect work values for the forward and backward transformations. The $\Delta G_{unfolded}$ was collected from the tripeptide database of pmx (http://pmx.mpibpc.mpg.de/tripeptide.html) which uses an ACE-GXG-NME tripeptide (X stands for mutated residue, G stands for amino acid glycine, ACE: acetylation at N-terminus. NME: N-methylation at C-terminus) to approximate the unfolded state of Hph protein. ΔG_{folded} and $\Delta G_{unfolded}$ represent the free-energy difference between the mutated and the wild-type states under folded and unfolded structures, respectively (Figure S15A). Finally, the difference in folding free energy between Hph and the Hph mutant ($\Delta\Delta G_{UN}$) is calculated by $\Delta\Delta G_{UN} = \Delta G_{folded} - \Delta G_{unfolded}$.

Computational Prediction of Stability Effects. The change in protein stability upon point mutations is predicted using a change in folding free energy $\Delta\Delta G_{UN}$, with negative and positive values corresponding to stabilizing and destabilizing effects, respectively.

$\Delta\Delta G_{UN}$ is defined with eqs 2 and 3

$$\Delta G_{UN} = G_{N(Native)} - G_{U(Unfold)} \quad (2)$$

$$\Delta\Delta G_{UN} = \Delta G_{UN(mutant)} - \Delta G_{UN(wild\text{-}type)} \quad (3)$$

The crystal structures of Hph WT (PDB ID: 3w0s) and Hph5 (PDB ID: 3w0n) were used for computational predictions.

Prediction by Rosetta. Initially, the Hph WT/Hph5 structures were energy-minimized using the FastRelax() protocol in Pyrosetta[32] 4.0. The energy-minimized structure was saved as a PDB file, serving as the start structure for computational mutations. Then, point mutations were achieved using the mutate_residue() function. Finally, the folding free energies ΔG_{UN} of Hph WT/Hph5 and mutants were calculated using REF15, and the resulting ΔG_{UN} was used to calculate $\Delta\Delta G_{UN}$. The above steps, beginning from RastRelax(), were repeated 100 times to calculate averages and standard deviations of $\Delta\Delta G_{UN}$.

Prediction by FoldX. Initially, the Hph WT/Hph5 structures were optimized to deal with bad torsion angles or Van der Waals' clashes using the RepairPDB module in FoldX[33] 5.0. Then, the mutation list was generated by a python script. Finally, stability effects of the listed mutations were predicted using the BuildModel module in FoldX. Each mutant was predicted fifteen times (numberOfRuns = 15), following the default setting (pH = 7, temperature = 298 K, ion strength = 0.05 M) to calculate averages and standard deviations of $\Delta\Delta G_{UN}$.

https://doi.org/10.1021/acsomega.3c00373
ACS Omega 2023, 8, 25739−25748

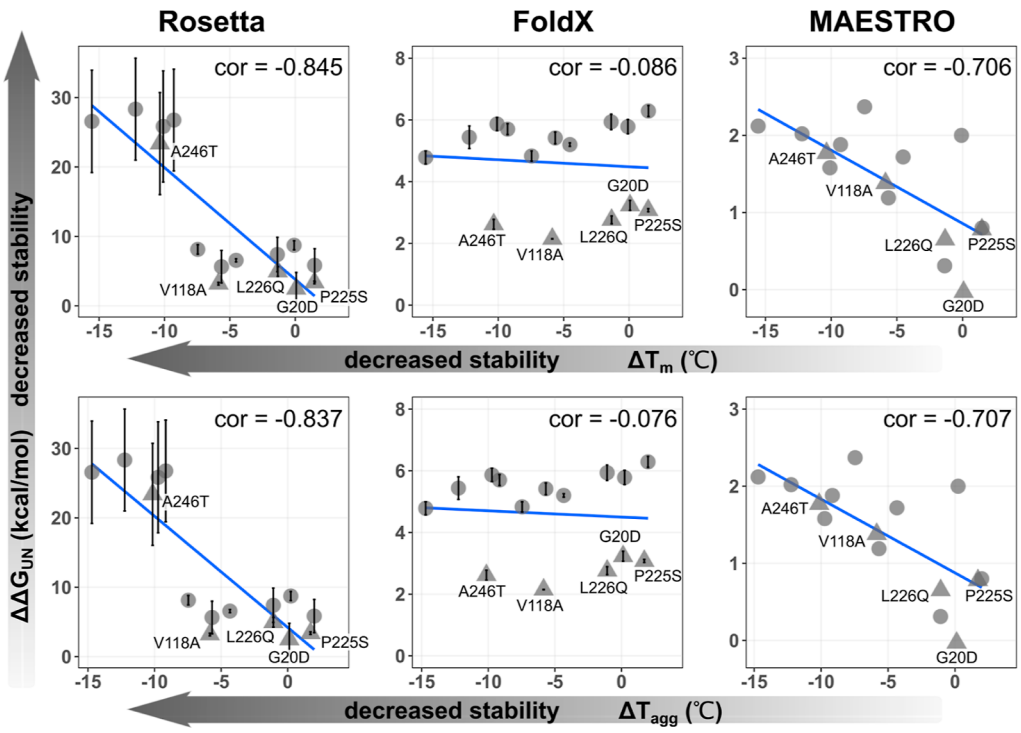


Figure 1. Correlation plots of predicted and experimental stability effects of reverse mutations on Hph5. Changes in folding free energy ($\Delta\Delta G_{UN}$) upon reverse mutagenesis of Hph5, predicted by Rosetta, FoldX, and MAESTRO, plotted against experimentally determined ΔT_m or ΔT_{agg} values. $\Delta G_{UN} = G_{native} - G_{unfolded}$, $\Delta\Delta G_{UN} = \Delta G_{UN(mutant)} - \Delta G_{UN\ (Hph5)}$. Top, $\Delta T_m = T_{m(mutant)} - T_{m(Hph5)}$; Bottom, $\Delta T_{agg} = T_{agg(mutant)} - T_{agg(Hph5)}$. Triangles, single-reverse mutants. Circles, double-reverse mutants. Blue line, linear regression fitting. Cor, Pearson correlation coefficient. Values of ΔT_m and ΔT_{agg} are the average of two independent measurements. The data sources are summarized in Tables S2 and S3.

Prediction by MAESTRO. Stability effects of mutations on Hph WT/Hph5 were predicted using established protocols in MAESTRO[34] 1.2.35 (command line version). Point mutations were manually specified in the option (--evalmut). The $\Delta\Delta G_{UN}$ was calculated using default settings (pH = 7, configuration file = config.xml).

RESULTS AND DISCUSSION

We used three popular computational programs to predict the stabilizing effect of Hph5 mutations. These programs include Rosetta, FoldX, and MAESTRO, based on empirical force field, statistical potential, and machine learning, respectively.[33−35] Changes in folding free energy ($\Delta\Delta G_{UN}$) were obtained to assess the stabilizing effect of each mutation within Hph5, with negative values corresponding to stabilizing mutations. Only FoldX predicted four mutations to have a stabilizing effect (Table S1). The other two programs predicted zero stabilizing mutations, and Rosetta even predicted that A118V and S225P would have strong destabilizing effects. In addition, T246A was consistently predicted as a neutral mutation (Table S1), although it is a mutation repeatedly identified in the first round of directed evolution experiments to obtain Hph5.[20] Because the current methods are more accurate in predicting destabilizing mutations,[36] we wonder whether the effects of Hph5 mutations could be better predicted in the reverse direction (Hph5 to Hph WT).

In predictions for five single-reverse mutations and their pairwise combinations (ten double-reverse mutations), Rosetta, FoldX, and MAESTRO predicted four, five, and two destabilizing mutations, respectively (Table S2). Contrary to the predictions of forward mutations, Rosetta and MAESTRO predicted that reverse mutation A246T had the strongest destabilizing effect. In addition, Rosetta and FoldX considered that the effects of single-reverse mutations were additive (Table S2).

To verify the predictions, we experimentally determined the destabilizing effects of various reverse mutations by differential scanning fluorimetry[37] with decreases in the melting temperature T_m and the aggregation temperature T_{agg} indicating destabilizing mutations (Figure S1). T_m and T_{agg} are the temperatures at which approximately 50% proteins in the population unfold and aggregate, respectively. We found that the single-reverse mutants A246T, V118A, and L226Q exhibited ΔT_m values of −10.31, −5.83, and −1.33 °C, respectively (Table S3). Similarly, the ΔT_{agg} values of these mutants were −10.17, −5.87, and −1.09 °C, respectively. The double-reverse mutants exhibited additive effects of the single-reverse mutants (Figure S2). In conclusion, A118V and T246A are the dominant mutations driving Hph5 stabilization, as their reverse mutations largely destabilize Hph5.

We plotted the predicted stability parameters against measured thermostability values to evaluate the accuracy of computational predictions for the reverse mutations (Figure 1). Rosetta and MAESTRO successfully predicted the destabilizing effect of A246T, showing a good correlation between experimental and predicted values, while FoldX failed to predict the strong destabilizing effects of A246T and V118A mutations. For Rosetta, V118A was also not quantitively predicted to have dominant destabilizing effects. As two of the three computational programs failed to accurately predict the effect of V118A, we next delved into the physical mechanisms behind Hph5 mutations.

Previous research has indicated that the driving force for Hph5 stabilization is hydrophobic interactions.[22] In the Hph5

https://doi.org/10.1021/acsomega.3c00373
ACS Omega 2023, 8, 25739−25748

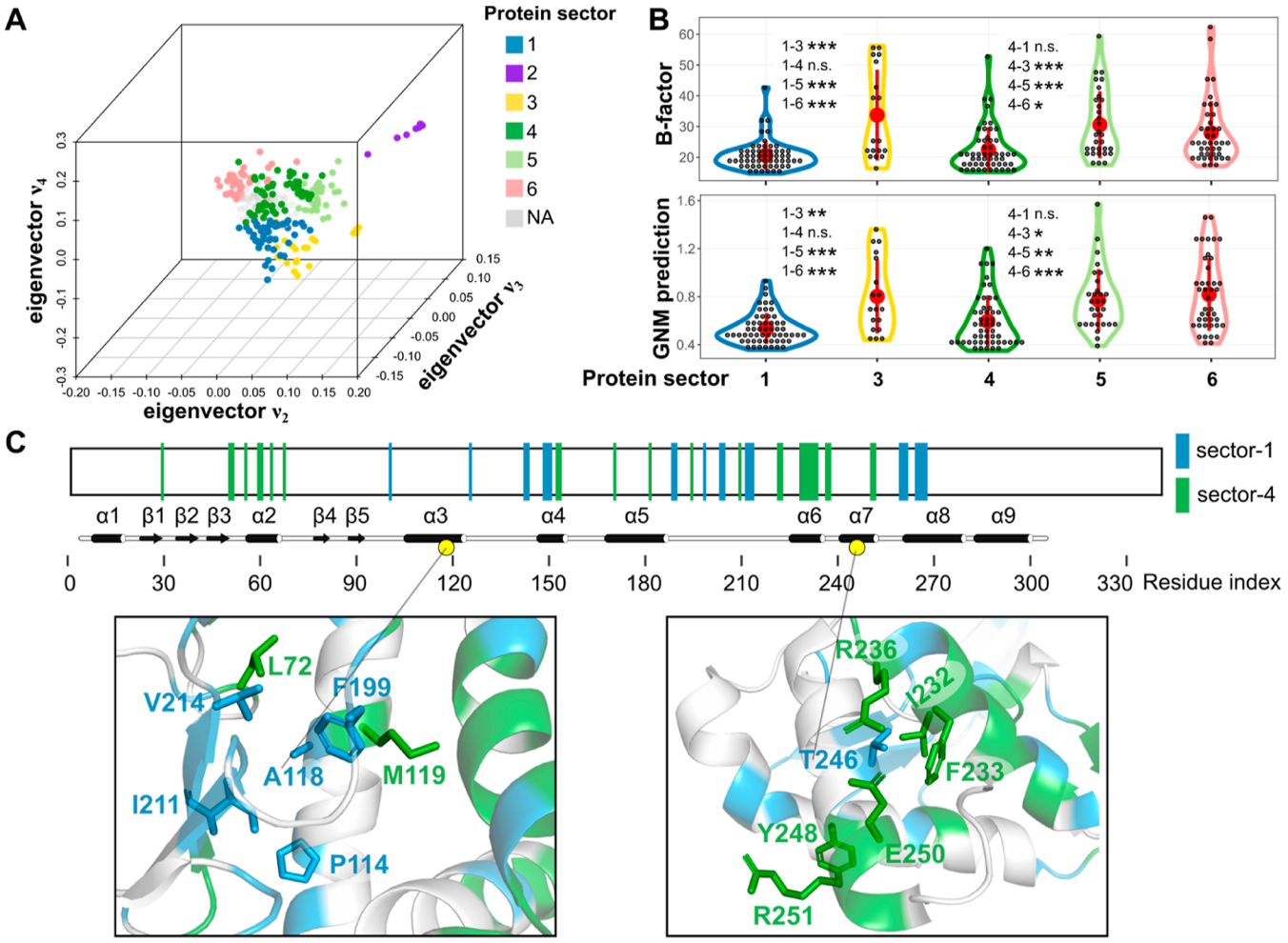


Figure 2. Identification of two sparse residue networks which are associated with Hph5 stabilization. (A) 3-D scatter plot of Hph residues in the space defined by 3 principal eigenvectors. Each dot represents one residue (341 total), colored according to protein sector assignment. NA, no assignment. (B) Violin plots compare residue flexibility between protein sectors. The residue flexibility is indicated by B-factors[39] or calculated by Gaussian Network Model.[40] Red dots and bars denote mean ± standard deviation, individual data points are shown. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, n.s., not significant, one-way ANOVA test (Tukey method, R package "multcomp"). (C) Top: diagram shows the distribution of residues from sector-1 and -4 on Hph primary and secondary structures (α, α-helix as cylinder; β, β-strand as arrow). Yellow ball, residue 118 and 246. Bottom: location of sparse residue network at position 118 (left) and 246 (right). Residues within the network belong to protein sector-1 (blue) or -4 (green) and are within 5 Å of at least one atom of residue 118 or 246. Side chains of the network residues are shown as stick representations.

structure, both residues 118 and 246 are surrounded by hydrophobic residues. However, the stability effect at position 118 was not as predictable as at position 246 (Figure 1). This indicates that there may be additional factors contributing to the stabilizing effect of A118V.

To investigate this further, we used protein sector analysis to determine protein sectors associated with Hph stability. The stability-associated sectors bridge the mutational effect of A118V to its surrounding residue network and further to the stability-associated local structure, providing insights that are not typically available through empirical structure analysis.

We performed protein sector analysis using an established workflow[16] (Figures S3−S5). Briefly, the workflow uses MSA of the Hph sequence to build a coevolutionary matrix, extracts coevolutionary features of each residue from the matrix, and classifies residues based on the similarity of coevolutionary features (see Materials and Methods for details). We divided Hph residues into six protein sectors (Figure 2A). Each sector displayed unique spatial features within the Hph tertiary structure (Figure S6). Residues of the same sector form spatially interconnected networks that link to specific functions,[15] including protein stability.

We propose that stability-related sectors are likely to contain residues with low flexibility. Because in many cases, enzyme thermostability is relevant to the reduced flexibility of the polypeptide chain,[38] and this relevance can be quantitatively analyzed by flexibility indices.[10] Through analyzing the experimental (B-factors[39]) and predicted (Gaussian Network Model[40]) flexibility indices of Hph residues, we identified that residues from sector-1 and sector-4 were significantly less flexible than residues from other sectors (Figure 2B). In the Hph structure, residues from sector-1 and -4 are buried within the protein core, while residues from other sectors are close to the protein surface (Figure S6). Based on these findings, we hypothesized that protein sector-1 and -4 are associated with Hph stability.

We used residues from sector-1 and -4 that are spatially proximal to residues 118 and 246 (within 5 Å) to build the residue networks. As a result, we obtained two sparse residue networks, comprising residues (L72, P114, M119, F199, I211,

https://doi.org/10.1021/acsomega.3c00373
ACS Omega 2023, 8, 25739−25748

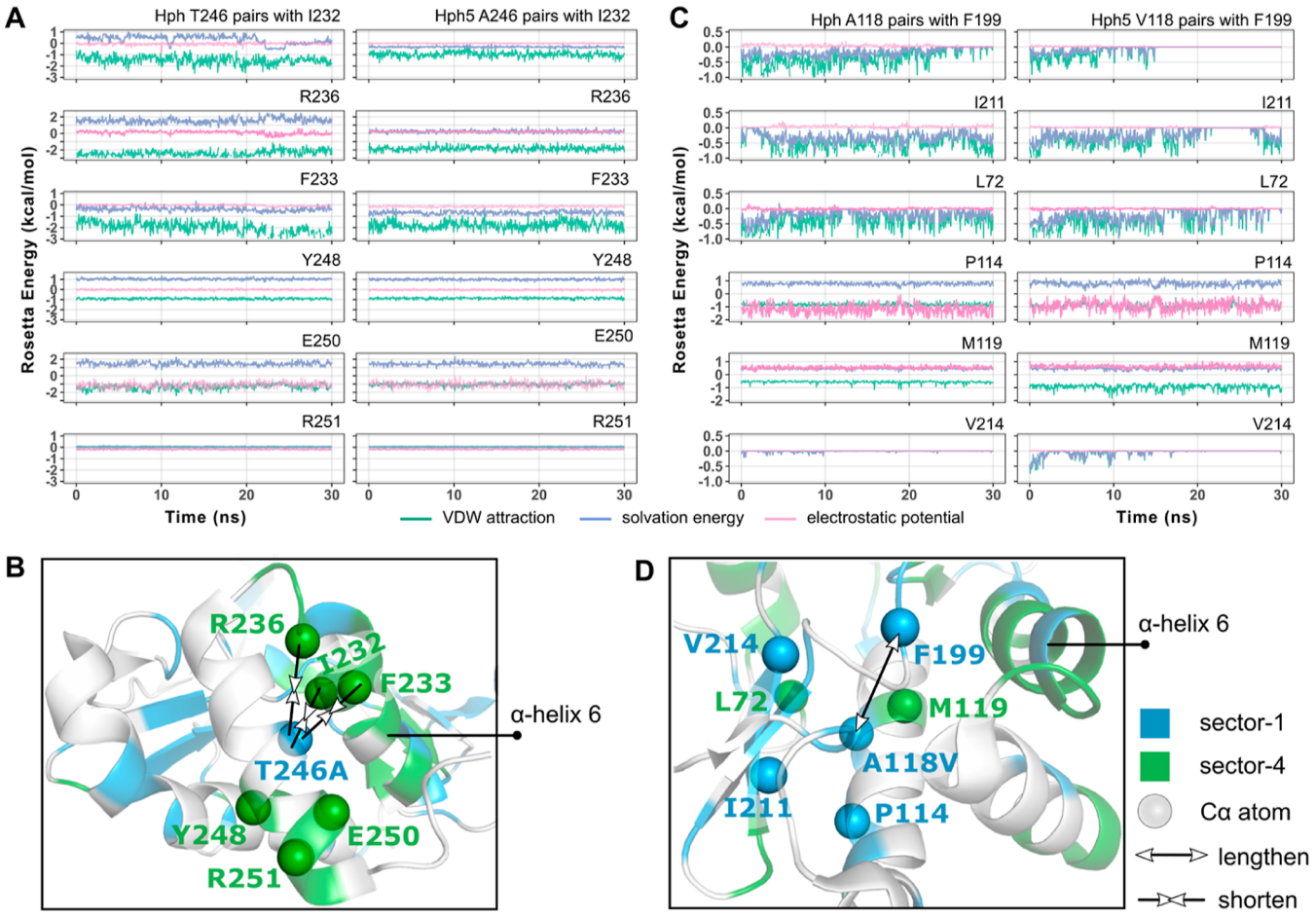


Figure 3. Comparison of residue interactions within the sparse residue networks between Hph WT and Hph5. The interaction energies for residue pairs at positions 246 (A) and 118 (C). The interaction energies were calculated by Rosetta Energy Function 2015:[28] Van Der Waals (VDW) attraction (green), solvation energy (blue), electrostatic potential (magenta). Results are the average of two independent simulations at the simulation temperature of 410 K. Changes in residue interactions at positions 246 (B) and 118 (D) after T246A and A118V mutations.

V214) for position 118 and residues (I232, F233, R236, Y248, E250, R251) for position 246 (Figure 2C). These sparse residue networks bridged residues 118 and 246 to stability-associated local structures, which aided subsequent MDS analyses.

We performed MDS to investigate how A118V and T246A mutations stabilize the protein by influencing residue dynamics in the sparse residue networks. We used an elevated simulation temperature to speed up the simulation process (Figures S7−S9). We plotted residue interaction energies against simulation time to generate energy traces (Figure 3A,C), with low-energy values indicating favorable interactions. By comparing the energy traces of Hph5 and Hph WT, we expected to determine the differences in dynamics of the network residues after mutations.

Within the sparse residue network at position 246, we observed a downshift of solvation energy traces (representing hydrophobic interaction) between I232−A246, R236−A246, and F233−A246 to varying degrees, upon comparison with energy traces of T246 (Figure 3A). This suggests that the hydrophobic interactions of I232 and F233 with residue 246 are enhanced after the T246A mutation. For R236, it is unlikely that the downshift of solvation energy between it and A246 is due to direct hydrophobic interactions between the two residues, as its side chain is hydrophilic. Instead, we speculate that nearby hydrophobic effects may be responsible, specifically the enhanced hydrophobic interactions between I232−A246 and F233−A246, as R236 is spatially close to these two residues (Figure 3B). In other words, the downshift of solvation energy between R236−A246 reflected the enhanced hydrophobicity of the region surrounding A246.

The above analysis does not consider the disruption of side-chain H-bond caused by the T246A mutation. To investigate this further, we compared the mutation-induced changes in H-bond and hydrophobic energies of residue 246. We found that Hph T246 can establish a side-chain H-bond using its hydroxyl group as H-bond donor (main-chain O atom of M242 as H-bond acceptor). This side-chain H-bond exhibited discrete signals during the MDS process (Figure S10A). Loss of the side-chain H-bond due to the T246A mutation resulted in an average $\Delta G_{\text{H-bond}}$ of 0.40 kcal/mol, indicating a destabilizing effect. In contrast, hydrophobic interactions of Hph5 A246 exhibited consistent improvement during MDS, represented by a $\Delta G_{\text{Solvation}}$ of −2.38 kcal/mol (Figure S10B). Comparing between $\Delta G_{\text{H-bond}}$ and $\Delta G_{\text{Solvation}}$, the T246A mutation may produce a net stabilizing effect.

Subsequently, we used energy trace comparison to infer the stabilizing effect of A118V. Compared with the energy traces at position 246, which showed continuous lines, the energy traces at position 118 were intermittent, suggesting that interactions with residue 118 are dynamic. In the case of Hph5 V118−F199 interaction, the energy traces exhibited zero signals beginning from the middle of the timeline (Figure 3C), indicating the loss

https://doi.org/10.1021/acsomega.3c00373
ACS Omega 2023, 8, 25739−25748

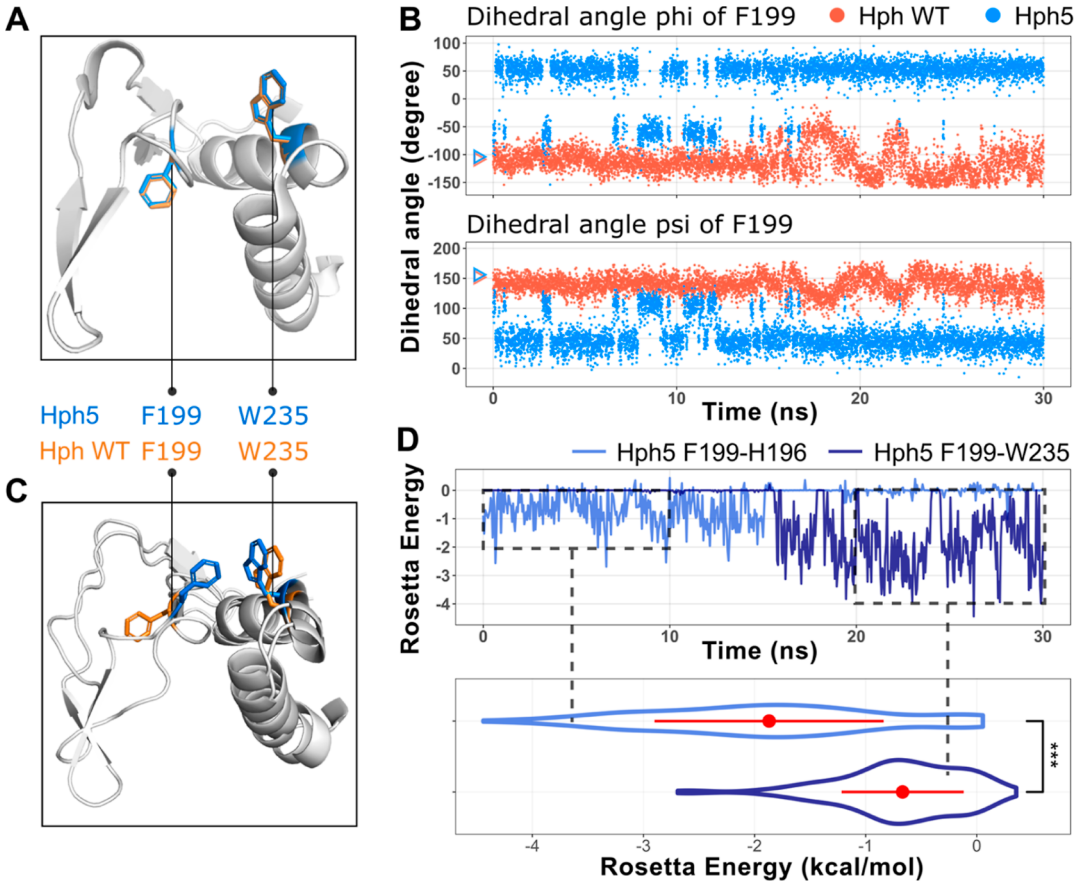


Figure 4. Rotation of F199 side chain and the established interaction between F199−W235 in Hph5. (A) Local structures show side-chain orientations of F199 and W235 in Hph crystal structures. (B) Timeline distributions of dihedral angles phi (top) and psi (bottom) of F199. Triangle at the left, dihedral angles in Hph crystal structures. (C) Local structures show side-chain orientations of F199 and W235 at simulation time of 25 ns. (D) Top, timeline traces of residue interaction energy. Interaction energy is the sum of Van Der Waals attraction, solvation energy, and electrostatic potential. Bottom, violin plot compares residue interaction energy between F199−W235 (20−30 ns) and F199−H196 (0−10 ns). ***$P < 0.001$, two-sided student's *t*-test.

of residue interaction. Unexpectedly, the interactions between Hph5 V118 and other hydrophobic residues did not improve after the A118V mutation. We speculated that the increased side-chain volume of residue 118 after mutation could have a steric hindrance effect, which could prevent it from interacting more strongly with other hydrophobic residues. Alternatively, it is possible that the hydrophobic residues that interact with V118 are not positioned in a way that would allow for increased interaction with the larger valine side chain. Therefore, we hypothesized that the stabilizing effect of A118V is derived from the residue movement of F199 (Figure 3D).

We analyzed the movements of F199 by examining the distribution of Cβ distance. Following the A118V mutation, we observed that F199 moved away from residue 118 and exhibited a broad Cβ distance distribution, indicating strong residue dynamics (Figure S11B). In contrast, prior to the A118V mutation, F199 did not exhibit such strong dynamics. F199 is located at the loop region between α-helices 5 and 6 and is adjacent to α-helix 6 within the tertiary structure (Figure 3D). This loop region and α-helix 6 are rich in residues from stability-associated sectors (Figure 2C). We reasoned that if the movement of F199 establishes interactions with α-helix 6, it may exert a stabilizing effect by connecting two stability-associated local structures.

The aromatic side chain of F199 is able to establish aromatic−aromatic interactions, which can contribute to maintaining protein stability by connecting secondary elements.[41] To verify the role of F199 in maintaining Hph5 thermostability, we performed a mutagenesis experiment and found that the F199A and F199V mutations caused significant decreases in Hph5 T_m with values of −15.9 and −12.1 °C, respectively (Table S4). Notably, α-helix 6 contains four aromatic residues. By scanning the energy traces between these residues and F199, we observed that Hph5 F199−W235 exhibited energy signals beginning from the middle of the timeline (Figure S12), indicating the establishment of residue interaction.

The F199−W235 interaction is not observed in the static structure, in which the side chain of F199 faces away from W235 (Figure 4A). We analyzed the details of the Hph5 F199−W235 interaction using timeline traces of F199 dihedral angles. Interestingly, we observed an obvious side-chain rotation of Hph5 F199, with its dihedral angles phi and psi rotated by ∼150 and ∼110°, respectively (Figure 4B), which turns its side chain toward W235 (Figure 4C). In contrast, F199 does not show side-chain rotation in Hph WT (Figure 4B,C). Before establishing the F199−W235 interaction, F199 maintains a $\pi-\pi$ interaction with H196. Comparing the residue interaction energy between F199−H196 and F199−W235 in Hph5 during MDS, we observed that the energy values of F199−W235 were significantly lower than that of F199−H196 (Figure 4D), suggesting that the establishment of F199−W235 interaction may result in a net stabilizing effect.

https://doi.org/10.1021/acsomega.3c00373
ACS Omega 2023, 8, 25739−25748

The above analysis suggests that the A118V mutation exerts a stabilizing effect by inducing F199 movement rather than by improving hydrophobic interactions. Previous studies have identified A71, L72, and P73 as hydrophobic residues that may interact with Hph5 V118.[22] By scanning the energy traces between residue 118 and residues adjacent to L72, we found that Hph5 V118-A71 showed a stronger hydrophobic interaction than Hph A118-A71 (Figure S13). However, mutation of A71 to amino acids with high hydrophobicity (Ile, Leu, and Val) resulted in only minimal changes in Hph5 thermostability, with $\Delta T_m < -1$ °C (Table S4). While MDS suggested enhanced hydrophobic interaction involving V118, this effect was not significant at the experimental level. Therefore, the data support the hypothesis that improved hydrophobic interaction is not the primary mechanism by which the A118V mutation stabilizes Hph5.

Ultimately, we used superimposed MDS snapshots to illustrate the establishment process of Hph5 F199−W235 interaction. The A118V mutation disrupted the interaction between residue 118 and F199, allowing F199 to rotate its side chain, turn toward W235, and then establish interaction with it (Figure S14A). The Hph5 F199−W235 interaction may stabilize the protein by maintaining low interaction energies (Figure 4D) and connecting two stability-associated local structures. In contrast, F199 in Hph WT had difficulty establishing interaction with W235 because F199−A118 and F199−H196 interactions restricted the side-chain rotation of F199 (Figure S14B).

CONCLUSIONS

In conclusion, protein sector analysis assisted MDS analysis by revealing the sparse residue networks surrounding mutational sites and residue clusters associated with protein stability. The sparse residue network represents the first shell influenced by mutational effects, while residue clusters represent stability-associated local structures. The T246A mutation exerts a stabilizing effect by interacting statically with α-helix 6, and the stability effect of its reverse mutation was effectively predicted using computational programs. In contrast, the A118V mutation stabilizes Hph5 by inducing F199 to interact with α-helix 6 and predicting the stability effect of its reverse mutation was challenging. Protein sector analysis will be a valuable tool for studying the stabilizing effects of mutations within large proteins, as it allows attention to be focused on stability-associated local structures.

Since our workflow identified a stabilizing mechanism that is challenging to obtain based solely on static analysis, we wondered whether MD-based free-energy calculation methods[42,43] can effectively predict the stabilizing effect of A118V. To this end, we used a vigorous MD-based method called alchemical free-energy calculations[44] to predict the free-energy changes associated with Hph5 mutations (Figure S15). The difference in folding free energy between Hph WT and mutant ($\Delta\Delta G_{UN}$) was calculated from the free-energy difference between the reference unfolded states ($\Delta G_{unfolded}$) and the free-energy difference between the native states (ΔG_{folded}). In the forward direction (Hph WT to Hph5), alchemical free-energy calculations correctly predicted the stabilizing effects of A118V, Q226L, and T246A with $\Delta\Delta G_{UN}$ of −1.76, −0.96, and −1.97 kcal/mol, respectively (Table S5). In our detailed energy analysis, the net stabilizing effect of T246A and A118V exhibits −1.98 kcal/mol (Figure S10, comparing the solvation energy and H-bond of residue 246) and −1.20 kcal/mol (Figure 4D, comparing residue−residue interactions between F199−W235 and F199−H196), respectively. The results show that our in-depth energy analysis is well consistent with alchemical free-energy calculations. Thus, we envision that combining our workflow and alchemical free-energy calculations will yield a new tool that simultaneously provides accurate predictions and physical explanations for stabilizing mutations.

This work provided a plausible explanation for the physical mechanism of a mutation that has an unpredictable stability effect. This explanation highlights the importance of considering protein dynamics in improving prediction accuracy. Our dynamic analysis was completed using a computationally light workflow. In the future, we envision to use the established workflow to mine the protein-stability database[45] for more undiscovered physical mechanisms of protein stabilization.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c00373.

Data of stability measurement and prediction, details of protein sector analysis and alchemical free-energy calculations, and extended analyses of molecular dynamics simulation (PDF)

AUTHOR INFORMATION

Corresponding Authors

Zixiao Xue − *State Key Laboratory of Bioreactor Engineering, Shanghai Collaborative Innovation Center for Biomanufacturing (SCICB), East China University of Science and Technology, Shanghai 200237, China;* Present Address: Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou, Zhejiang 310024, China; orcid.org/0000-0002-2685-2389; Email: neoxialulu@163.com

Shu Quan − *State Key Laboratory of Bioreactor Engineering, Shanghai Collaborative Innovation Center for Biomanufacturing (SCICB), East China University of Science and Technology, Shanghai 200237, China; Shanghai Frontiers Science Center of Optogenetic Techniques for Cell Metabolism, Shanghai 200237, China*; Email: shuquan@ecust.edu.cn

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c00373

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are grateful to National Center for Protein Sciences Shanghai for the instrument support for differential scanning fluorimetry. We thank Prof. Yan-Wen Tan for offering MATLAB code of protein sector analysis. We thank Dr. Qi Chen and Dr. Wei He for helpful suggestions in the manuscript preparation. This work was supported by National Natural Science Foundation of China (NSFC) grants 32222049, 31661143021, and 31870054 (to S.Q.).

REFERENCES

(1) Reetz, M. T.; Carballeira, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2007**, *2*, 891−903.

(2) Zheng, Y. C.; Ding, L. Y.; Jia, Q.; Lin, Z.; Hong, R.; Yu, H. L.; Xu, J. H. A High-Throughput Screening Method for the Directed Evolution of Hydroxynitrile Lyase towards Cyanohydrin Synthesis. *Chembiochem* **2021**, *22*, 996−1000.
(3) Traxlmayr, M. W.; Faissner, M.; Stadlmayr, G.; Hasenhindl, C.; Antes, B.; Ruker, F.; Obinger, C. Directed evolution of stabilized IgG1-Fc scaffolds by application of strong heat shock to libraries displayed on yeast. *Biochim. Biophys. Acta* **2012**, *1824*, 542−549.
(4) Kim, D. S.; Song, H. N.; Nam, H. J.; Kim, S. G.; Park, Y. S.; Park, J. C.; Woo, E. J.; Lim, H. K. Directed evolution of human heavy chain variable domain (VH) using in vivo protein fitness filter. *PLoS One* **2014**, *9*, No. e98178.
(5) Brouns, S. J.; Wu, H.; Akerboom, J.; Turnbull, A. P.; de Vos, W. M.; van der Oost, J. Engineering a selectable marker for hyperthermophiles. *J. Biol. Chem.* **2005**, *280*, 11422−11431.
(6) Huang, P.; Chu, S. K. S.; Frizzo, H. N.; Connolly, M. P.; Caster, R. W.; Siegel, J. B. Evaluating Protein Engineering Thermostability Prediction Tools Using an Independently Generated Dataset. *ACS Omega* **2020**, *5*, 6487−6493.
(7) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712−725.
(8) Yu, H.; Dalby, P. A. Coupled molecular dynamics mediate long- and short-range epistasis between mutations that affect stability and aggregation kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E11043−E11052.
(9) Yu, H.; Huang, H. Engineering proteins for thermostability through rigidifying flexible sites. *Biotechnol. Adv.* **2014**, *32*, 308−315.
(10) Reetz, M. T.; Carballeira, J. D.; Vogel, A. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew. Chem., Int. Ed. Engl.* **2006**, *45*, 7745−7751.
(11) Niesen, M. J.; Bhattacharya, S.; Grisshammer, R.; Tate, C. G.; Vaidehi, N. Thermostabilization of the β_1-Adrenergic Receptor Correlates with Increased Entropy of the Inactive State. *J. Phys. Chem. B* **2013**, *117*, 7283−7291.
(12) Lee, S.; Bhattacharya, S.; Tate, C. G.; Grisshammer, R.; Vaidehi, N. Structural dynamics and thermostabilization of neurotensin receptor 1. *J. Phys. Chem. B* **2015**, *119*, 4917−4928.
(13) Bo, L.; Milanetti, E.; Chen, C. G.; Ruocco, G.; Amadei, A.; D'Abramo, M. Computational Modeling of the Thermodynamics of the Mesophilic and Thermophilic Mutants of Trp-Cage Miniprotein. *ACS Omega* **2022**, *7*, 13448−13454.
(14) Srivastava, A.; Nagai, T.; Srivastava, A.; Miyashita, O.; Tama, F. Role of Computational Methods in Going beyond X-ray Crystallography to Explore Protein Structure and Dynamics. *Int. J. Mol. Sci.* **2018**, *19*, 3401.
(15) Halabi, N.; Rivoire, O.; Leibler, S.; Ranganathan, R. Protein sectors: evolutionary units of three-dimensional structure. *Cell* **2009**, *138*, 774−786.
(16) Chang, J.; Zhang, C.; Cheng, H.; Tan, Y. W. Rational Design of Adenylate Kinase Thermostability through Coevolution and Sequence Divergence Analysis. *Int. J. Mol. Sci.* **2021**, *22*, 2768.
(17) Wang, J.; Zhao, Y.; Wang, Y.; Huang, J. Molecular dynamics simulations and statistical coupling analysis reveal functional coevolution network of oncogenic mutations in the CDKN2A-CDK6 complex. *FEBS Lett.* **2013**, *587*, 136−141.
(18) Tian, P.; Louis, J. M.; Baber, J. L.; Aniana, A.; Best, R. B. Co-Evolutionary Fitness Landscapes for Sequence Design. *Angew. Chem., Int. Ed. Engl.* **2018**, *57*, 5953.
(19) Rao, R. N.; Allen, N. E.; Hobbs, J. N., Jr.; Alborn, W. E., Jr.; Kirst, H. A.; Paschal, J. W. Genetic and enzymatic basis of hygromycin B resistance in Escherichia coli. *Antimicrob. Agents Chemother.* **1983**, *24*, 689−695.
(20) Nakamura, A.; Takakura, Y.; Kobayashi, H.; Hoshino, T. In vivo directed evolution for thermostabilization of Escherichia coli hygromycin B phosphotransferase and the use of the gene as a selection marker in the host-vector system of Thermus thermophilus. *J. Biosci. Bioeng.* **2005**, *100*, 158−163.
(21) Henikoff, S.; Henikoff, J. G. Amino Acid Substitution Matrices from Protein Blocks. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 10915−10919.
(22) Iino, D.; Takakura, Y.; Fukano, K.; Sasaki, Y.; Hoshino, T.; Ohsawa, K.; Nakamura, A.; Yajima, S. Crystal structures of the ternary complex of APH(4)-Ia/Hph with hygromycin B and an ATP analog using a thermostable mutant. *J. Struct. Biol.* **2013**, *183*, 76−85.
(23) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950−1958.
(24) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(25) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(26) Abascal, J. L.; Vega, C. eneral purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.
(27) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319−2327.
(28) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; Labonte, J. W.; Pacella, M. S.; Bonneau, R.; Bradley, P.; Dunbrack, R. L.; Das, R.; Baker, D.; Kuhlman, B.; Kortemme, T.; Gray, J. J. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13*, 3031−3048.
(29) Bakan, A.; Dutta, A.; Mao, W.; Liu, Y.; Chennubhotla, C.; Lezon, T. R.; Bahar, I. Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinformatics* **2014**, *30*, 2681−2683.
(30) Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. pmx: Automated protein structure and topology generation for alchemical perturbations. *J. Comput. Chem.* **2015**, *36*, 348−354.
(31) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys. Rev. Lett.* **2003**, *91*, 140601.
(32) Chaudhury, S.; Lyskov, S.; Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **2010**, *26*, 689−691.
(33) Delgado, J.; Radusky, L. G.; Cianferoni, D.; Serrano, L. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics* **2019**, *35*, 4168−4169.
(34) Laimer, J.; Hofer, H.; Fritz, M.; Wegenkittl, S.; Lackner, P. MAESTRO−multi agent stability prediction upon point mutations. *BMC Bioinf.* **2015**, *16*, 116.
(35) Kellogg, E. H.; Leaver-Fay, A.; Baker, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **2011**, *79*, 830−8388.
(36) Pucci, F.; Bernaerts, K. V.; Kwasigroch, J. M.; Rooman, M. Quantification of biases in predictions of protein stability changes upon mutations. *Bioinformatics* **2018**, *34*, 3659−3665.
(37) Desroses, M.; Busker, S.; Astorga-Wells, J.; Attarha, S.; Kolosenko, I.; Zubarev, R. A.; Helleday, T.; Grander, D.; Page, B. D. G. STAT3 differential scanning fluorimetry and differential scanning light scattering assays: Addressing a missing link in the characterization of STAT3 inhibitor interactions. *J. Pharm. Biomed. Anal.* **2018**, *160*, 80−88.
(38) Vihinen, M. Relationship of protein flexibility to thermostability. *Protein Eng.* **1987**, *1*, 477−480.
(39) Sun, Z.; Liu, Q.; Qu, G.; Feng, Y.; Reetz, M. T. Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chem. Rev.* **2019**, *119*, 1626−1665.
(40) Bahar, I.; Atilgan, A. R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding Des.* **1997**, *2*, 173−181.
(41) Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V. Inter-residue and solvent-residue interactions in proteins: A statistical study

https://doi.org/10.1021/acsomega.3c00373
ACS Omega 2023, 8, 25739−25748

on experimental structures. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 139−151.
(42) Steinbrecher, T.; Zhu, C.; Wang, L.; Abel, R.; Negron, C.; Pearlman, D.; Feyfant, E.; Duan, J.; Sherman, W. Predicting the Effect of Amino Acid Single-Point Mutations on Protein Stability-Large-Scale Validation of MD-Based Relative Free Energy Calculations. *J. Mol. Biol.* **2017**, *429*, 948−963.
(43) Seeliger, D.; de Groot, B. L. Protein thermostability calculations using alchemical free energy simulations. *Biophys. J.* **2010**, *98*, 2309−2316.
(44) Aldeghi, M.; de Groot, B. L.; Gapsys, V. Accurate Calculation of Free Energy Changes upon Amino Acid Mutation. *Methods in Molecular Biology*; Humana Press, 2019; Vol. *1851*, pp 19−47.
(45) Stourac, J.; Dubrava, J.; Musil, M.; Horackova, J.; Damborsky, J.; Mazurenko, S.; Bednar, D. FireProtDB: database of manually curated protein stability data. *Nucleic Acids Res.* **2021**, *49*, D319−D324.

https://doi.org/10.1021/acsomega.3c00373
ACS Omega 2023, 8, 25739−25748