THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Improving the Thermostability of Xylanase A from *Bacillus subtilis* by Combining Bioinformatics and Electrostatic Interactions Optimization

Published as part of The Journal of Physical Chemistry virtual special issue "Dave Thirumalai Festschrift".

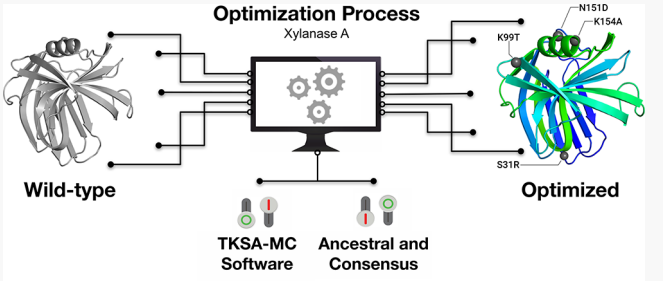
Khoa Ngo, Fernando Bruno da Silva, Vitor B. P. Leite, Vinícius G. Contessoto,* and José N. Onuchic*

Cite This: *J. Phys. Chem. B* 2021, 125, 4359–4367

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The rational improvement of the enzyme catalytic activity is one of the most significant challenges in biotechnology. Most conventional strategies used to engineer enzymes involve selecting mutations to increase their thermostability. Determining good criteria for choosing these substitutions continues to be a challenge. In this work, we combine bioinformatics, electrostatic analysis, and molecular dynamics to predict beneficial mutations that may improve the thermostability of XynA from *Bacillus subtilis*. First, the Tanford–Kirkwood surface accessibility method is used to characterize each ionizable residue contribution to the protein native state stability. Residues identified to be destabilizing were mutated with the corresponding residues determined by the consensus or ancestral sequences at the same locations. Five mutants (K99T/N151D, K99T, S31R, N151D, and K154A) were investigated and compared with 12 control mutants derived from experimental approaches from the literature. Molecular dynamics results show that the mutants exhibited folding temperatures in the order K99T > K99T/N151D > S31R > N151D > WT > K154A. The combined approaches employed provide an effective strategy for low-cost enzyme optimization needed for large-scale biotechnological and medical applications.



1. INTRODUCTION

Advances in protein engineering contribute to a vast and growing number of biotechnology applications. These applications include DNA manipulation, disease treatment, natural and pharmaceutical compound synthesis, and many others. In particular, the bioenergy industry has adopted enzymes for the hydrolysis of lignocellulosic biomass to produce soluble sugars that undergo a fermentation process for bioethanol production.[1] Improving the thermostability of these enzymes would allow them to work longer under the conditions prescribed by the biomass degradation process. Moreover, the ability to use high temperatures for such procedures might not only reduce microbial contamination but also improve reaction rates.[2] There are several strategies for improving the thermostability of enzymes. Directed evolution, for example, mimics the natural evolution cycle under specific laboratory conditions to guide the proteins toward adapting to the new environments through natural selection.[3] In rational design, the prediction of mutations is based on known information about the enzyme sequence and/or structure. By using such information, computational analyses are performed to suggests good candidates to be used in synthetic work. These computational techniques, combined with wet-laboratory approaches, accelerate the protein design and make the protein suitable for different applications.[4]

Despite recent advances in protein engineering, it remains a challenging process to discover beneficial mutations. In this work, we explore different techniques to improve the thermostability of XynA from *Bacillus subtilis*. Xylanases are enzymes that catalyze the hydrolysis of the β-1,4 glycosidic linkages of xylans, present in many plant cell walls.[5] XynA was chosen for our study due to the availability of experimental data in the literature regarding its thermostability and mutant variants; therefore, it is a great test system for our theoretical design approach. We propose to perform rational residue substitution by combining several existing theoretical tools: electrostatic analysis, ancestral sequence reconstruction strategy, and consensus sequence strategy. First, the Tanford–Kirkwood surface accessibility–Monte Carlo (TKSA–MC)[6] method is used to identify candidate residues in the wild-type

Received: February 10, 2021
Revised: April 10, 2021
Published: April 22, 2021

https://doi.org/10.1021/acs.jpcb.1c01253
J. Phys. Chem. B 2021, 125, 4359–4367

sequence to be mutated. Then, the residues suitable for mutation are identified and later mutated with the corresponding residue in the ancestral and consensus sequences (SI text and Figure S1). The ancestral sequence can be reconstructed via statistical techniques that help to infer the most statistically probable amino acid sequence of each node on the phylogenetic tree.[7] Using each of the strategies individually can be daunting since each of them still involves a degree of trial and error to select and test beneficial mutations. For example, in the consensus protein design, only 50% of conserved residues are responsible for improved stability while the remaining residues are either neutral or destabilizing.[8] Depending on the complexity of the data used for ancestral and consensus sequence construction, the two may significantly differ from the sequence of interest, making it even more of a challenge to determine an efficient way to proceed with mutation selection. With the novel combination of strategies presented here, it is possible to reduce the randomness in selecting mutations by first identifying which residue is most suitable to be replaced, and second, by directly replacing those residues with those theoretically predicted to be beneficial. Studies have suggested that over deep evolutionary time, the native state stability of proteins tends to decrease.[9,10] Therefore, restoring some ancestral residues in the reconstructed proteins might help to stabilize their extant counterparts.[2] In the consensus design strategy, the consensus residues are simply those with the highest frequency of occurrence at individual positions in the multiple sequence alignment of extant homologues. The rationale is that residues that are highly conserved through millions of years of evolution are likely to hold the key to the thermodynamic stability and catalytic activity of the protein.[11] If those residues are not already present in the protein of interest, then by restoring them we can potentially produce mutants with higher thermostability and activity than for the wild-type counterpart. In the last step, molecular dynamics simulations were carried out to explore the mutation effects on the protein folding and stability.

2. METHODS

2.1. Selection of Eligible Mutations Based on the Consensus and Ancestral Sequences. By accessing the Protein Data Bank (http://rcsb.org),[12] we obtained the FASTA sequence and the tertiary structures of XynA using PDB ID 1XXN.[1] A search was then performed using the basic local alignment search tool (BLAST) to look for putative GH11s and other homologous protein sequences. The consensus sequence was generated using the consensus finder web server with a set of homologous proteins.[13] The same set of homologues is used to compute a phylogenetic tree in *PhyloBayes*, a Bayesian Monte Carlo Markov Chain (MCMC) sampler.[14] Finally, using the *FastML* web server, the sequence alignment and the consensus tree were used to reconstruct the ancestral sequences of all corresponding proteins. The ModWeb server was used to predict the tertiary structures of the wild-type (WT) protein and of the ancestral and the consensus models.[15] Afterward, the TKSA-MC server was used to evaluate the electrostatic free-energy contribution of each polar-charged residue in the WT native state at a specific pH and temperature.[16] This is achieved by calculating protein charge−charge interactions via the Tanford−Kirkwood surface accessibility model combined with the Monte Carlo method for sampling different protein protonation states.

Destabilizing residues that are candidates to be replaced must fulfill two criteria: first, they must present unfavorable energy values, $\Delta G_{qq} \geq 0$, and second, their side chains must be exposed to the solvent with a solvent-accessible surface area (SASA) ratio of ≥50%. There is a strong correlation between the heat capacity variation (ΔC_p) and ΔSASA.[17] The ΔC_p variation affects the protein free-energy profile (ΔG) and the melting temperature (T_m). The criterion is adopted to avoid significant changes in the protein ΔC_p.[18] It is then expected that the selected mutation will contribute to only a slight variation of the protein structure and will not drastically change the protein hydrophobic core. The ancestral, the consensus, and the wild-type sequences were aligned together. Residues identified to be electrostatically destabilizing by the TKSA-MC method were replaced with those found at the same positions in the ancestral or consensus sequences. However, the method application was expanded, and the mutations are not limited only to destabilizing residues: If a charged residue appears in either the ancestral or the consensus sequence while the residue at the same position in the WT is uncharged, then the uncharged residue is replaced with the charged one. After the mutant structures were modeled, the TKSA-MC method was also applied to optimize the electrostatics and to determine whether the mutation is stabilizing. Five mutants were created (K99T/N151D, K99T, S31R, N151D, and K154A), and their structures were predicted using the ModWeb server.[19] Figure 1 presents the 1D and 3D structures of XynA and highlights the mutated residues on the basis of the criteria presented in the methodology.

2.2. Non-Native Potential Addition to the Structure-Based Cα Model. The protein folding process was investigated in the wild type and the mutants of XynA by molecular dynamics employing the Cα structure-based model (SBM-Cα) with the addition of non-native potentials. The amino acids of the protein in the SBM-Cα are represented by a single bead located in the α-carbon position.[20−22] The contribution of the SBM-Cα potential is given by the first five terms in eq 1, and the last term corresponds to the non-native potential that accounts for the electrostatic interactions among all charged residues. This approach has been widely used in systems that take into account fixed charge.[23−27]

$$
\begin{aligned}
V(\Gamma, \Gamma_o) = & \sum_{\text{bonds}} \epsilon_r (r - r_o)^2 + \sum_{\text{angles}} \epsilon_\theta (\theta - \theta_o)^2 \\
& + \sum_{\text{backbone}} \epsilon_\phi \left\{ [1 - \cos(\phi - \phi_o)] + \frac{1}{2}[1 - \cos(3(\phi - \phi_o))] \right\} \\
& + \sum_{\text{contacts}} \epsilon_C \left[5 \left(\frac{d_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{d_{ij}}{r_{ij}} \right)^{10} \right] + \sum_{\text{noncontacts}} \epsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12} \\
& + \sum_{\text{electrostatic}} K_{\text{electrostatic}} \frac{q_i q_j \exp(-\kappa r_{ij})}{\epsilon_K r_{ij}}
\end{aligned}
\tag{1}
$$

In the SBM-Cα potential, parameters r, θ, and ϕ represent the distance between two subsequent residues and the angles formed by three and four subsequent residues of the protein, respectively. All SBM parameters are obtained from the native

https://doi.org/10.1021/acs.jpcb.1c01253
J. Phys. Chem. B 2021, 125, 4359−4367

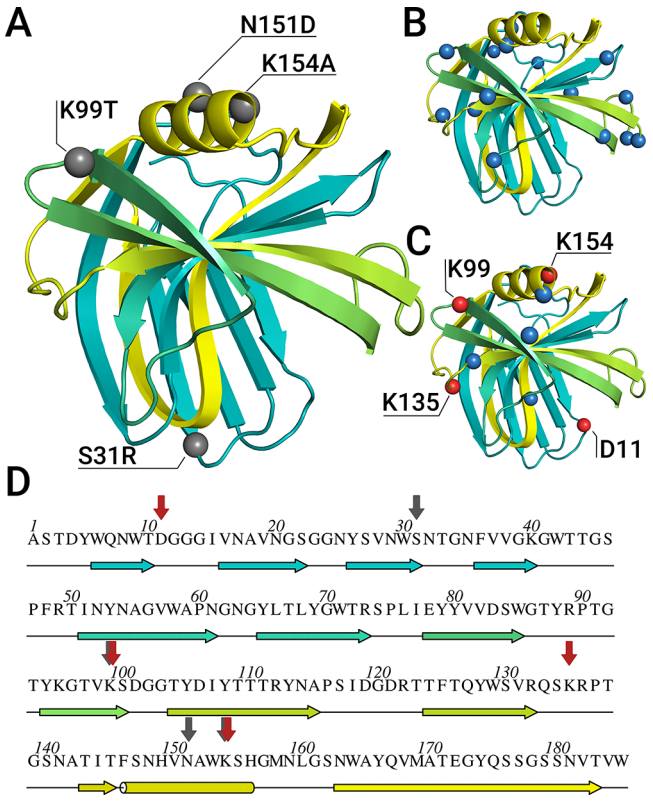


Figure 1. Cartoon representation (A–C) and primary sequence (D) of the wild-type xylanase structure with PDB ID 1XXN. (A) The gray beads correspond to the mutations predicted by TKSA and their structures created by the ModWeb server. (B) The blue beads are the residues with energy favorable to stabilizing the protein, $\Delta G_{qq} < 0$. (C) The red and blue beads are the residues with unfavorable energy, $\Delta G_{qq} \geq 0$, with SASA $\geq$ 50% and SASA < 50%, respectively. (D) The primary sequence of xylanase with gray and red arrows indicating the bead positions represented in A and B.

structure. r_{ij} represents the distance between two noncovalent beads, and d_{ij} is the distance between two beads in the native structure. The contact map is determined by the shadow contact map algorithm.[28] The last term in the SBM-Cα potential represents the nonspecific repulsion to all residue pairs which are not in contact in the native structure. The strength of the bonds, angles, and dihedral angles is described by ϵ_r, ϵ_θ, and ϵ_ϕ, respectively, and parameters $\epsilon_r = 100\epsilon_C$, $\epsilon_\theta = 20\epsilon_C$, $\epsilon_\phi = \epsilon_C$, and $\epsilon_{NC} = \epsilon_C$ in which ϵ_C is equal to 1 unit (in reduced units).

The electrostatic interactions are considered by adding point charges at beads,[23,29−31] which represent the acidic/basic residues (i.e., histidine, lysine, and arginine are positively charged; glutamic acid and aspartic acid are negatively charged). Two charged amino acids interact via the last term in eq 1, in which $K_{electrostatic} = 332$ kcal Å/(mol e^2),[32] q_i and q_j are residue charges i and j, $\epsilon_K = 80$ is the dielectric constant, and κ is the inverse of the Debye length.[25,33]

2.3. Simulation Details. Simulations were performed using the GROMACS (version 2019.2) molecular dynamic package.[34] All files necessary to perform structure-based simulations were obtained from the SMOG web server.[35] The Berendsen thermostat algorithm[36] was employed to maintain coupling to an external bath with a relaxation of 1 ps. The proteins were initialized in an open random configuration and simulated over 2.5 × 10^9 steps (250 ns), and this process was repeated for five simulations. The configurations were saved every 4000 steps. The replica exchange molecular dynamics (REMD)[37,38] technique was employed to allow systems with similar potential energies to sample conformations at different temperatures, thus overcoming energy barriers on the potential surface to better sample the energy landscape. For each protein model, 12 replicas were set to run in a temperature range exponentially distributed around an estimated folding temperature obtained from preliminary testing. The reaction coordinate used to follow the folding events is based on the fraction of the native contacts (Q). A native contact is formed when the distance between two residues with indices i and j, where $j > i + 3$, is less than $1.2d_{ij}$, where d_{ij} is such a distance in the folded state. The thermodynamic free-energy profile for each mutant was obtained using the weighted histogram analysis method (WHAM) after combining multiple simulations for a range of different temperatures. By combining all the three approaches (electrostatics analysis, ancestral sequence reconstruction approach, and the consensus sequence approach) as demonstrated in Figure 2, we are able to identify residues suitable for mutations to increase the enzyme thermostability.

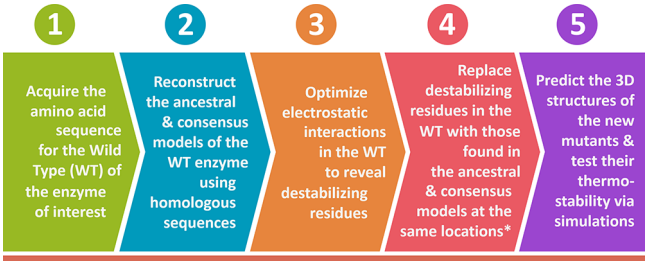


Figure 2. Outline of steps taken for the rational selection of mutants. The process starts by obtaining the amino acid sequences and predicting the corresponding tertiary structures. With this information, the ancestral and consensus sequences are reconstructed. Next, electrostatic interactions are optimized using the TKSA-MC web server to determine destabilizing residues in the WT to serve as candidates to be experimentally mutated. After mutating the destabilizing residues, the mutant structures are predicted and their thermostability is verified by a simulation of the folding process. *As a possible expansion of application in the case of proteins absent of electrostatically destabilizing residues, step 4 may also include mutating uncharged residues in the WT with charged residues located in the ancestral and consensus sequence at the same locations. The resulting mutants are then considered for further analysis if they display more favorable ΔG_{qq} than does the WT per TKSA-MC results.

3. RESULTS AND DISCUSSION

3.1. Mutant Selection and ΔG_{qq} Analysis. First, the consensus sequence was generated by the consensus-finder web server using the closest 2000 homologous proteins. Evolution can introduce new pathways and add roughness to the energy landscape. Since independent pathways are not highly conserved, constructing a consensus of homologues removes them and smoothes the energy landscape.[2] To reconstruct the ancestral sequence, the same sequences were imported into *Jalview*[39] and aligned using MAFFT.[40] Sequences with 95% redundancy or that were significantly longer or shorter than the WT sequence were also selectively removed. The remaining sequences were realigned with MAFFT and fed to *Gblock*[41] to further eliminate regions

https://doi.org/10.1021/acs.jpcb.1c01253
J. Phys. Chem. B 2021, 125, 4359−4367

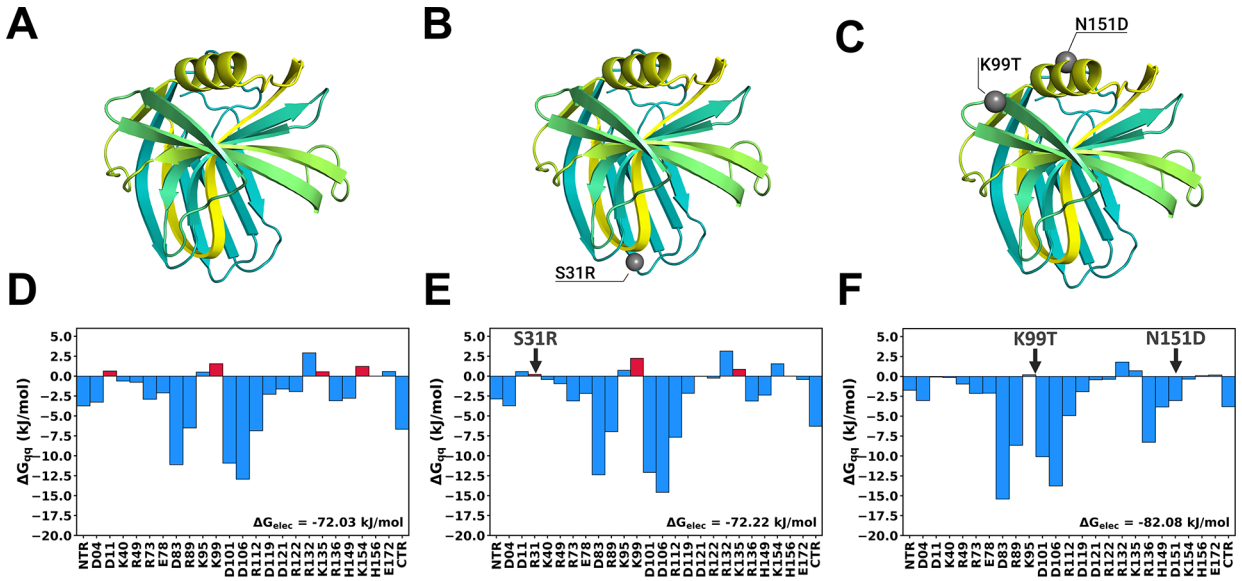


Figure 3. Sample usage of electrostatics analysis to identify candidate residues to be mutated by using the TKSA-MC web server. (A−C) Generated 3D structures of the XynA WT variant and the S31R and K99T/N151D mutants, respectively. (D−F) TKSA-MC results of the corresponding structures above them, indicating the charge−charge energy contribution of each ionizable residue to the native state stability when compared to the unfolded state. Red bars represent destabilizing residues with unfavorable energy values $\Delta G_{qq} \geq 0$. These residues, when mutated, might help stabilize the protein. Bars that have unfavorable energy values but have SASA < 50% are not candidates to be replaced and remain blue.

with weak phylogenetic links. The resulting segments were imported into *ProtTest*[42] to determine the best-fitting evolutionary model prior to MCMC sampling.[43] With all of the necessary information, the ancestral reconstruction process could commence. The sequences were imported into *PhyloBayes*, and two MCMC chains were set to run in parallel, regularly probing for convergence. After 4000 cycles, 10% of the initial trees were discarded as burn-in, and the remaining trees were then used to compute averages. Only trees with a posterior consensus of 0.70 or higher in each node are considered in the final consensus tree. Finally, using the *FastML* web server,[44] the sequence alignment and the consensus tree were used to reconstruct the ancestral sequences of all corresponding proteins. For the next step, the TKSA-MC method was performed on the WT with pH and temperature parameters set at 6.0 and 328 K, respectively, to evaluate the interaction energy among the charges of each ionizable residue in the protein. The values for pH and temperature were chosen because they have been experimentally verified to be optimal for XynA.[16] The mutants were selected on the basis of the TKSA-MC results, as shown in Figure 3. (See Figures S2−S4 in the SI.) Residues D11, K99, K135, and K154 were found to destabilize the native structure by having $\Delta G_{qq} \geq 0$ with SASA ≥ 50%. Although D11 and K135 are candidates to be mutated and increase the protein thermostability, they are not explored because ancestral and consensus sequences have identical residues at 11 and 135 locations. Electrostatics optimization using the TKSA-MC is essential for optimization under different pH and temperature conditions. These features make the method versatile enough to be used for a variety of applications. After aligning all three sequences (WT, ancestral, and consensus) using MAFFT, candidate residues in the WT were mutated with other residues found in the same locations in the ancestral or consensus sequences. The 3D structures of all mutants were generated using ModWeb.[19] (See the RMSD in Table S1 and the structures Figure S5 in the SI.) Another essential aspect to consider is to perform the mutation to improve thermostability while maintaining catalytic activity. The strategy includes locating the enzyme activity site and avoiding mutations in that region. The protein−ligand binding site was predicted using the COACH web server.[45,46] The mutants selected and the reasons behind their selection are as follows:

- S31R: *R* appeared only in the ancestral sequence but nowhere else at the same location in the multiple sequence alignment (MSA).
- K154A: *A* appeared in both the ancestral and consensus sequences at the same location in the MSA.
- K99T: *T* appeared in both the ancestral and consensus sequences at the same location in the MSA.
- N151D: *D* appeared in both the ancestral and consensus sequences at the same location in the MSA. N is uncharged, and N151D exhibits a lower ΔG_{elec} than does the WT.
- K99T/N151D: We combined the previous two point mutations after simulations revealed that both K99T and N151D were stabilizing.

3.2. Molecular Dynamics Simulations with SBM-Cα Models. Simulations were performed in GROMACS[34] using SBM-Cα models, and the thermodynamic properties of all mutants and the WT are displayed in Figure 4. Four mutants exhibited higher folding temperatures than the WT variant, while one exhibited slightly lower folding temperature. Twelve control mutants obtained from experiments were also simulated to validate the simulation results of the new mutants. The folding temperature, T_F, for each model was obtained by the peak of its heat capacity, C_v. In Figure 5A, the changes in the folding temperature displayed a significant improvement over the WT (see Figure S6 in the SI). The K99T mutation displays the highest folding temperature, making this mutation most likely viable for processes that require higher temperatures. Higher T_F is associated with an increase in the native state free-energy stabilization, as shown

https://doi.org/10.1021/acs.jpcb.1c01253
J. Phys. Chem. B 2021, 125, 4359−4367

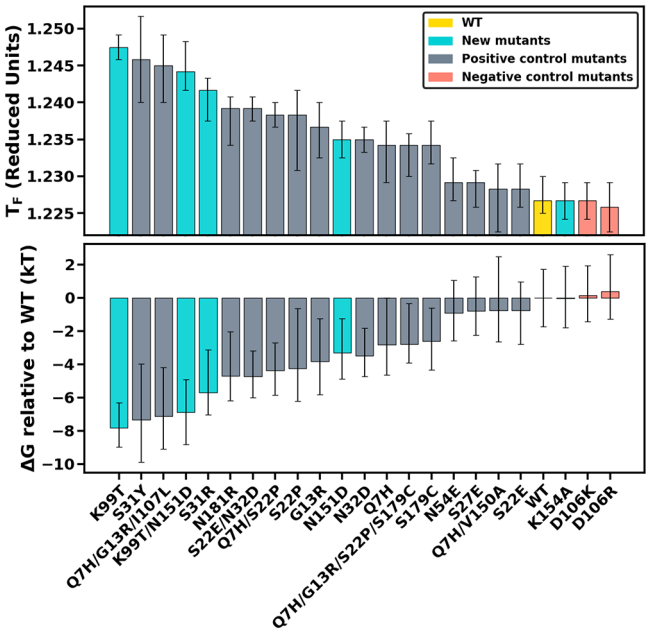


Figure 4. Thermodynamic parameters obtained from simulations for all mutants. K99T and K154A were obtained by mutating an electrostatically destabilizing residue in the WT, identified through optimizing electrostatic interactions using the TKSA-MC method, with a residue located in the ancestral or consensus sequence at the same location. S31R and N151D were obtained by mutating an uncharged residue in the WT with a charged residue located in the ancestral and consensus sequence at the same location and exhibit lower ΔG_{elec} than the WT (ΔG_{elec} = −72.22 and −91.07 kJ/mol, respectively, compared to −72.03 kJ/mol for the WT). K99T/N151D was obtained by mutating an electrostatically destabilizing residue and an uncharged residue in the WT with residues located in the ancestral and consensus sequence at the same locations and was found to exhibit lower ΔG_{elec} than the WT (ΔG_{elec} = −82.08 vs −72.03 kJ/mol, respectively).

in Figure 5B. The Q region from 0 to 0.2 is used to calculate the free energy of the protein unfolded conformations, Q = 0.2 to 0.8 for the transition state and Q = 0.8 or above for the native state. Therefore, the lower the free-energy minimum of the native state, the more stable the protein and consequently the more difficult for the protein to unfold. The substantial decrease in the native state free energy of the mutants relative to the wild type indicates that mutants S31R and K99T were successfully stabilized (Figure S7 in the SI). Therefore, there is a direct connection between the native state free energy and the folding temperature. The results reveal that the mutants stabilized the native state of the protein, allowing them to maintain the catalytic activity at higher temperatures compared to the WT. The mutants' folding temperatures obtained using the approach outlined in this paper were compared against the mutants' folding temperatures investigated by wet-laboratory experiments (Q7H/G13R/I107L, Q7H/G13R/S22P/S179C, Q7H/S22P, Q7H, Q7H/V150A, S22P, G13R, S31Y,[16] S179C,[47] N181R, N32D, S22E/N32D, N54E, S27E, and S22E[48]). Two control mutants were also intentionally destabilized by mutating the most electrostatically stabilizing residue identified in the TKSA-MC result with a random residue of the opposite charge (D106K and D106R). As predicted, these two mutants were observed to exhibit a lower T_F than the WT, as shown in Figure 4. Both mutations D106R and D106K are responsible for a significant unfavorable change in the electrostatic free energy. The values go from −47.91 and −35.53 kJ/mol for D106R and D106K, respectively, to −72.03 kJ/mol for the WT. However, both mutations only decreased the stability of the WT by a small amount. This can be attributed to the fact that the mutated proteins form more native contacts (685 and 683 contacts for D106R and D106K, respectively, compared to 679 contacts for the native WT). The formation of new native contacts slightly offsets the loss of stability due to unfavorable electrostatic changes, resulting in a small decrease in both mutants' stability.[23,31]

At the folding temperature, the contributions of the contact and electrostatic interactions to the folding process were calculated from the simulation trajectory data and are presented in Figure 6. The contact energy function depends on the degree of nativeness of a particular configuration. As the number of native contacts increases, so does the contact energy contribution. This observation comes from the model

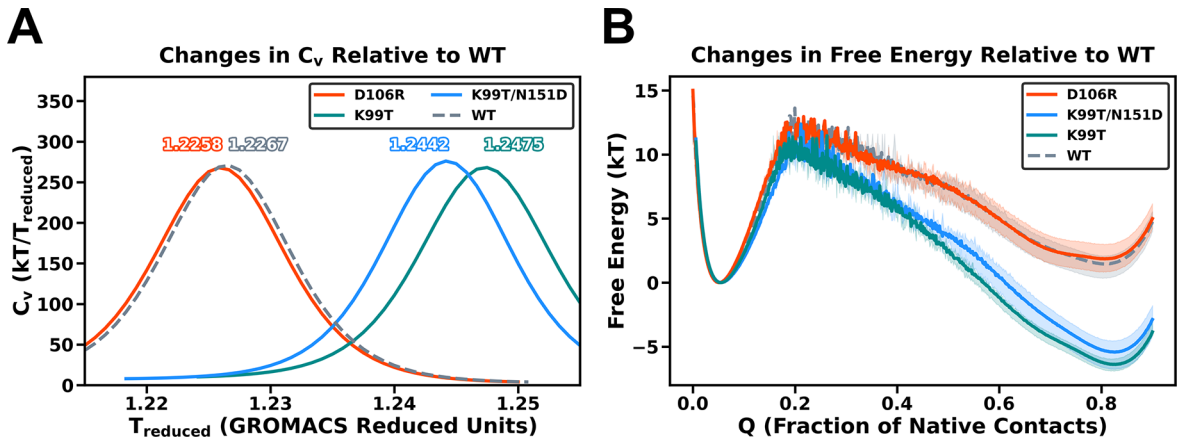


Figure 5. (A) Heat capacity curve. (B) Free-energy profile as a function of the fraction of native contacts at the folding temperature of the WT. Orange lines represent the mutant introduced as the negative control (D106R). Gray lines represent the wild type (WT). Blue and green lines represent the most thermostable mutants (K99T/N151D and K99T, respectively) obtained using the methods described above. In A, the C_v peak of D106R shows a lower folding temperature compared with the WT. In B, the free-energy profile shows that mutation D106R has a less stable native state compared with the WT. In contrast, the C_v of S31R and K99T has a higher folding temperature than D106R and the WT. S31R and K99T free-energy profiles show that both mutations present a more stable native state than the WT and the D106R mutation.

https://doi.org/10.1021/acs.jpcb.1c01253
J. Phys. Chem. B 2021, 125, 4359−4367

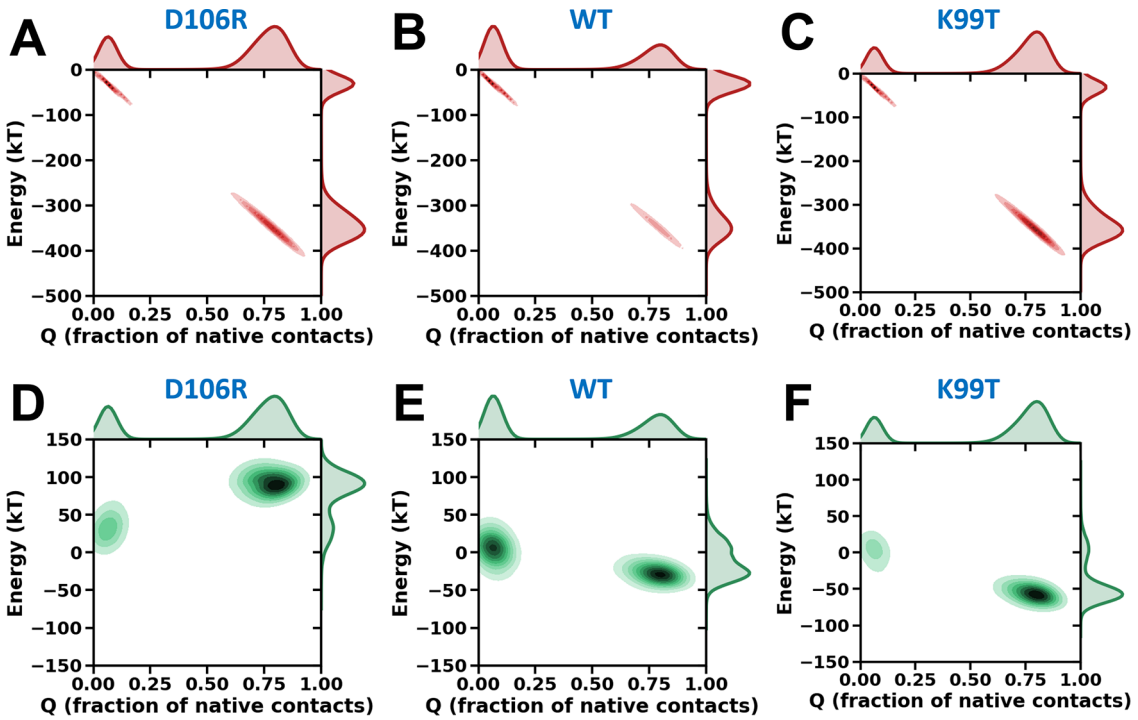


Figure 6. (A−C) Distributions of the contact energy function over reaction coordinate Q at the simulation temperature closest to the folding temperature of D106R, WT, and K99T protein models, respectively. (D−F) Those for the electrostatic potentials.

construction and applies to the cases of D106R, WT, and K99T shown in Figure 6A−C, respectively. On the other hand, the electrostatic interactions are responsible for the most significant variation in the energy plots (Figure 6D−F for D106R, WT, and K99T, respectively) and play an essential role in the folding process and native state stability. The energy variations were expected once the mutants were chosen from the TKSA-MC electrostatics optimization.[6] For the destabilizing mutation (D106R), the electrostatic energy contributions increase once the protein folds to its native state. This positive energy contribution is responsible for making the D106R native state less stable when compared to the WT. The WT analysis presents some variation in the electrostatic energy between the unfolded and the folded state. The electrostatic energy decreases during the folding process, indicating that charge−charge interactions help the formation of the native contacts. A similar effect is observed in the electrostatically optimized mutation K99T. This predicted mutant presents a decrease in the electrostatic energy in the native state compared with the WT. This mutation favors the formation of native state contacts with optimized electrostatic interactions in the protein (Figures S8 and S9 in the SI). This nonnative interaction optimization helps to decrease the native state frustration, which may increase protein stability.[21,22,49−52] The hydrophobic nonnative interactions contribution was also explored in the protein folding simulations, and the results lead to a similar optimum mutation ranking (SI text and Figure S10−S14.)

Among the mutants, S31R was the only instance in which the residue selected to be the substitute (R) appeared in only the ancestral sequence but not in the consensus sequence. In K99T, N151D, and K154A, the substituting residues appear in both the ancestral and consensus sequences. Over time, residues that are important to protein stability and activity tend to be conserved to help the protein function in its respective biological context. Consistent with our findings, if a residue is stabilizing in the ancestral protein, then it is likely to be conserved throughout evolutionary time and appear in the extant proteins. Even though most ancient proteins were highly thermostable, the trend in thermostability is not smooth over time as some studies have shown.[10] As the protein adapts to changes in the environment through the ages, its stability fluctuates in response to environmental changes. For this reason, there might not be a significant trend in stability in recent ancestors without also accounting for even older ancestors. Future studies could incorporate the utilizing sequences from different ancestral nodes in the phylogenetic tree for mutant selection to account for such fluctuations in protein stability.

3.3. Estimations of Folding Temperatures for the Mutant Based on Simulation Results. Both simulation and experimental T_F values of all control mutants are known (except for D106K and D106R in Table 1), and the two temperature data points are plotted against each other and then linearly fitted. Then, the predicted mutants with only simulation results are plotted along the regression line, as shown in Figure 7. A rough estimation of the mutants' theoretical T_F was obtained using extrapolation and is displayed in Table 2. Third-generation mutant Q7H/G13R/S22P/S179C is reported experimentally to increase the enzyme melting temperature by 17.7 °C. The results from simulations did not predict such considerable stabilization. The suggestion for the predictions' underestimation is that this mutant is reported to make a disulfide bond between xylanases.[53] The disulfide bond formation and interactions between different chains are not taken into account in this work.

https://doi.org/10.1021/acs.jpcb.1c01253
J. Phys. Chem. B 2021, 125, 4359−4367

Table 1. Simulation and Experimental T_F of All Mutants and WT

model	simulation T_F (reduced units)	experimental T_F (K)
S31Y	$1.2458^{+0.0058}_{-0.0059}$	332.2[16]
Q7H/G13R/I107L	$1.2450^{+0.0050}_{-0.0042}$	344.6[16]
N181R	$1.2392^{+0.0050}_{-0.0016}$	331.0[48]
S22E/N32D	$1.2392^{+0.0017}_{-0.0016}$	332.4[48]
Q7H/S22P	$1.2383^{+0.0016}_{-0.0017}$	341.8[16]
S22P	$1.2383^{+0.0075}_{-0.0034}$	334.8[16]
G13R	$1.2367^{+0.0042}_{-0.0033}$	335.4[16]
N32D	$1.2350^{+0.0017}_{-0.0017}$	328.3[48]
Q7H	$1.2342^{+0.0050}_{-0.0033}$	335.8[16]
Q7H/G13R/S22P/S179C	$1.2342^{+0.0042}_{-0.0016}$	349.5[16]
S179C	$1.2342^{+0.0025}_{-0.0033}$	333.0[47]
N54E	$1.2292^{+0.0025}_{-0.0033}$	329.0[48]
S27E	$1.2292^{+0.0034}_{-0.0016}$	328.7[48]
Q7H/V150A	$1.2283^{+0.0058}_{-0.0034}$	335.1[16]
S22E	$1.2283^{+0.0025}_{-0.0034}$	330.6[48]
WT	$1.2267^{+0.0017}_{-0.0033}$	331.8,[16] 330.4[48]

4. CONCLUSIONS

Among the variety of computational tools for designing and performing protein engineering, we chose to combine electrostatic analysis, ancestral, and consensus sequences approaches to identify destabilizing residues and predict stabilizing mutations for them. First, the TKSA-MC method reports locations of charged residues contributing to the destabilization of the native state. The ancestral and consensus sequences suggest possible replacements for these destabilizing residues. The most thermostable mutant (K99T) exhibited a higher folding temperature than other mutants and the WT. The set of mutations was investigated by molecular dynamics to identify the energy contributions during the folding process and the native state stability. The comparison of the folding temperature of the predicted stabilizing mutations of XynA from *B. subtilis* with the experimental data from the literature corroborates this approach by combining the presented computational strategies. The method assists in predicting whether a mutation will increase the protein thermostability. Other mutations that have no experimental data are presented in the Supporting Information. Although the predictions in this work are single mutations only, the methodology presented here could be applied to several mutation candidates at once.

Table 2. Folding Temperature Estimations of the Tested Mutants in Kelvin

rank	model	simulation T_F (reduced units)	extrapolated T_F (K)	ΔT_F from WT (K)
1	K99T	$1.2475^{+0.0017}_{-0.0017}$	352.6652	20.87
2	K99T/N151D	$1.2442^{+0.0025}_{-0.0041}$	347.5702	15.77
3	S31R	$1.2417^{+0.0042}_{-0.0016}$	343.7103	11.91
4	N151D	$1.2350^{+0.0025}_{-0.0025}$	333.3657	1.57
5	WT	$1.2267^{+0.0017}_{-0.0033}$	331.8000	0.00
6	D106K	$1.2267^{+0.0025}_{-0.0025}$	320.5508	−11.25
7	K154A	$1.2267^{+0.0025}_{-0.0025}$	320.5508	−11.25
8	D106R	$1.2258^{+0.0033}_{-0.0034}$	319.1613	−12.64

These mutations can help experimental groups to speed up the search for good mutations. While the enzymes' catalytic activity is also worth investigating in addition to just stability, the strategy adopted and tested in this work is promising in improving the thermostability of proteins, enabling them to be used in processes that require higher temperatures than what the wild type can withstand.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.1c01253.

FASTA sequences for the WT, ancestral, and consensus sequences; multiple sequences alignment of the three main sequences; TKSA-MC results for all mutants tested; RMSD values for all mutants with the wild-type protein; superposition of all mutants with the cartoon to wild-type tertiary structure representation; heat capacity curve for each investigated mutant; free-energy profile as a function of the fraction of native contacts Q; changes to contact frequency of the most and least thermostable mutants compared to the WT at different temperatures; simulation results of sample mutants obtained from using only TKSA-MC compared to the TKSA-MC, ancestral, and consensus combined approach; hydrophobic contributions were incorporate into the vanilla C-α structure-based model; thermodynamic parameters from folding simulations with the hydrophobic potential for all mutants; heat capacity curve and the free-energy profile comparing the best mutants, the negative control,

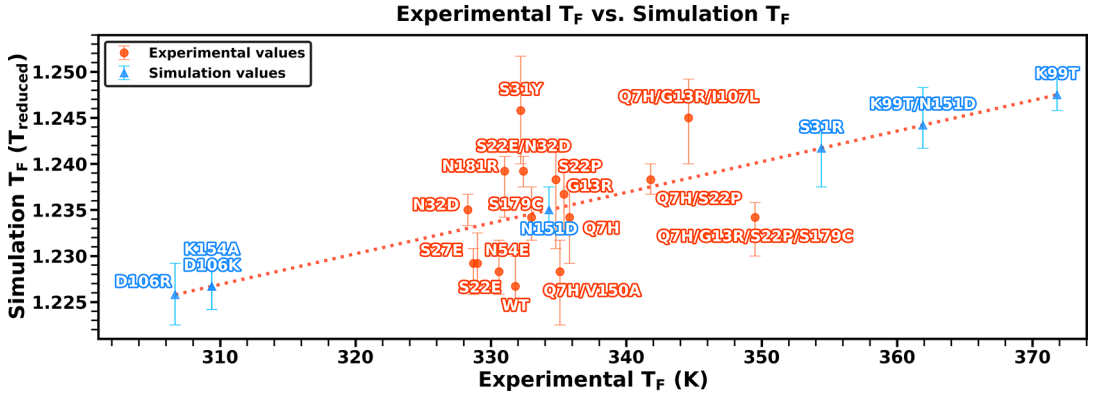


Figure 7. Plot of T_F from simulations in reduced units against T_F from experiments (K). Models labeled in orange have both simulation and experimental T_F known and were used to create a linear regression line (represented as an orange dotted line in the plot). Models in blue have only simulation T_F values known and are plotted along the regression line for estimations of theoretical folding temperatures in Kelvin.

https://doi.org/10.1021/acs.jpcb.1c01253
J. Phys. Chem. B 2021, 125, 4359−4367

and the WT; energy contributions from the native contacts and electrostatic and hydrophobic potentials; heat capacity curve for each investigated mutant with the hydrophobic potential; free-energy profile as a function of the fraction of native contacts Q with the hydrophobic potential (PDF)

AUTHOR INFORMATION

Corresponding Authors

Vinícius G. Contessoto − *Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States; Departamento de Física, Instituto de Biociências, Letras e Ciências Exatas UNESP - Univ. Estadual Paulista, São José do Rio Preto, SP, Brazil;* orcid.org/0000-0002-1891-9563; Email: contessoto@rice.edu

José N. Onuchic − *Center for Theoretical Biological Physics, Department of Physics & Astronomy, Department of Chemistry, and Department of Biosciences, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0002-9448-0388; Email: jonuchic@rice.edu

Authors

Khoa Ngo − *Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States; Department of Physics, University of Houston, Houston, Texas 77004, United States*

Fernando Bruno da Silva − *Departamento de Física, Instituto de Biociências, Letras e Ciências Exatas UNESP - Univ. Estadual Paulista, São José do Rio Preto, SP, Brazil*

Vitor B. P. Leite − *Departamento de Física, Instituto de Biociências, Letras e Ciências Exatas UNESP - Univ. Estadual Paulista, São José do Rio Preto, SP, Brazil;* orcid.org/0000-0003-0008-9079

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.1c01253

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by the Center for Theoretical Biological Physics sponsored by the NSF (grant PHY-2019745). J.N.O. was also supported by the NSF (grant CHE-1614101) and by the Welch Foundation (grant C-1792). J.N.O. is a Cancer Prevention and Research Institute of Texas (CPRIT) Scholar in Cancer Research. V.G.C. is a Robert A. Welch Postdoctoral Fellow and was also funded by FAPESP (São Paulo Research Foundation) and CAPES (Higher Education Personnel Improvement Coordination) grants 2016/13998-8 and 2017/09662-7. V.B.P.L. was supported by the National Council for Scientific and Technological Development (CNPq) and FAPESP grants 2014/06862-7, 2016/19766-1, and 2019/22540-3. F.B.d.S. was supported by the National Council for Scientific and Technological Development (CNPq) grant process no. 141715/2017-0

REFERENCES

(1) Binod, P.; Gnansounou, E.; Sindhu, R.; Pandey, A. Enzymes for second generation biofuels: Recent developments and future perspectives. *Bioresource Technology Reports* **2019**, *5*, 317−325.

(2) Kazlauskas, R. Engineering more stable proteins. *Chem. Soc. Rev.* **2018**, *47*, 9026−9045.

(3) Cobb, R. E.; Chao, R.; Zhao, H. Directed evolution: Past, present, and future. *AIChE J.* **2013**, *59*, 1432−1440.

(4) Hellinga, H. W. Rational protein design: Combining theory and experiment. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 10015−10017.

(5) Alponti, J. S.; Maldonado, R. F.; Ward, R. J. Thermostabilization of Bacillus subtilis GH11 xylanase by surface charge engineering. *Int. J. Biol. Macromol.* **2016**, *87*, 522−528.

(6) Contessoto, V. G.; Oliveira, V. M. D.; Fernandes, B. R.; Slade, G. G.; Leite, V. B. P. TKSA-MC: A web server for rational mutation through the optimization of protein charge interactions. *Proteins: Struct., Funct., Genet.* **2018**, *86*, 1184−1188.

(7) Thornton, J. W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **2004**, *5*, 366−375.

(8) Porebski, B. T.; Buckle, A. M. Consensus protein design. *Protein Eng., Des. Sel.* **2016**, *29*, 245−251.

(9) Gaucher, E. A.; Govindarajan, S.; Ganesh, O. K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **2008**, *451*, 704−707.

(10) Wheeler, L. C.; Lim, S. A.; Marqusee, S.; Harms, M. J. The thermostability and specificity of ancient proteins. *Curr. Opin. Struct. Biol.* **2016**, *38*, 37−43.

(11) Sternke, M.; Tripp, K. W.; Barrick, D. Consensus sequence design as a general strategy to create hyperstable, biologically active proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 11275−11284.

(12) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(13) Jones, B. J.; Lim, H. Y.; Huang, J.; Kazlauskas, R. J. Comparison of Five Protein Engineering Strategies for Stabilizing an /-Hydrolase. *Biochemistry* **2017**, *56*, 6521−6532.

(14) Lartillot, N.; Lepage, T.; Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **2009**, *25*, 2286−2288.

(15) Pieper, U.; Eswar, N.; Davis, F. P.; Braberg, H.; Madhusudhan, M. S.; Rossi, A.; Marti-Renom, M.; Karchin, R.; Webb, B. M.; Eramian, D.; et al. MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* **2006**, *34*, D291−D295.

(16) Ruller, R.; Deliberto, L.; Ferreira, T. L.; Ward, R. J. Thermostable variants of the recombinant xylanase a from Bacillus subtilis produced by directed evolution show reduced heat capacity changes. *Proteins: Struct., Funct., Genet.* **2008**, *70*, 1280−1293.

(17) Myers, J. K.; Nick Pace, C.; Martin Scholtz, J. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **1995**, *4*, 2138−2148.

(18) Gribenko, A. V.; Patel, M. M.; Liu, J.; McCallum, S. A.; Wang, C.; Makhatadze, G. I. Rational stabilization of enzymes by computational redesign of surface charge−charge interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 2601−2606.

(19) Pieper, U.; Eswar, N.; Webb, B. M.; Eramian, D.; Kelly, L.; Barkan, D. T.; Carter, H.; Mankoo, P.; Karchin, R.; Marti-Renom, M. A.; et al. modbase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* **2009**, *37*, D347−D354.

(20) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and Energetic Factors: What Determines the Structural Details of the Transition State Ensemble and "En-Route" Intermediates for Protein Folding? An Investigation for Small Globular Proteins. *J. Mol. Biol.* **2000**, *298*, 937−953.

(21) Contessoto, V. G.; Lima, D. T.; Oliveira, R. J.; Bruni, A. T.; Chahine, J.; Leite, V. B. P. Analyzing the Effect of Homogeneous Frustration in Protein Folding. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 1727−1737.

(22) Mouro, P. R.; de Godoi Contessoto, V.; Chahine, J.; Junio de Oliveira, R.; Pereira Leite, V. B. Quantifying Nonnative Interactions in the Protein-Folding Free-Energy Landscape. *Biophys. J.* **2016**, *111*, 287−293.

(23) Azia, A.; Levy, Y. Nonnative Electrostatic Interactions Can Modulate Protein Folding: Molecular Dynamics with a Grain of Salt. *J. Mol. Biol.* **2009**, *393*, 527−542.
(24) Tripathi, S.; Garcìa, A. E.; Makhatadze, G. I. Alterations of Nonconserved Residues Affect Protein Stability and Folding Dynamics through Charge−Charge Interactions. *J. Phys. Chem. B* **2015**, *119*, 13103−13112.
(25) Tzul, F. O.; Schweiker, K. L.; Makhatadze, G. I. Modulation of folding energy landscape by charge−charge interactions: Linking experiments with computational modeling. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E259−E266.
(26) Bruno da Silva, F.; Contessoto, V. G.; De Oliveira, V. M.; Clarke, J.; Leite, V. B. Non-native cooperative interactions modulate protein folding rates. *J. Phys. Chem. B* **2018**, *122*, 10817−10824.
(27) da Silva, F. B.; de Oliveira, V. M.; Sanches, M. N.; Contessoto, V. G.; Leite, V. B. P. Rational Design of Chymotrypsin Inhibitor 2 by Optimizing Non-Native Interactions. *J. Chem. Inf. Model.* **2020**, *60*, 982−988.
(28) Noel, J. K.; Whitford, P. C.; Onuchic, J. N. The Shadow Map: A General Contact Definition for Capturing the Dynamics of Biomolecular Folding and Function. *J. Phys. Chem. B* **2012**, *116*, 8692−8702.
(29) Contessoto, V. G.; de Oliveira, V. M.; de Carvalho, S. J.; Oliveira, L. C.; Leite, V. B. P. NTL9 Folding at Constant pH: The Importance of Electrostatic Interaction and pH Dependence. *J. Chem. Theory Comput.* **2016**, *12*, 3270−3277.
(30) Coronado, M. A.; Caruso, I. P.; De Oliveira, V. M.; Contessoto, V. G.; Leite, V. B. P.; Kawai, L. A.; Arni, R. K.; Eberle, R. J. Cold Shock Protein A from Corynebacterium Pseudotuberculosis: Role of Electrostatic Forces in the Stability of the Secondary Structure. *Protein Pept. Lett.* **2017**, *24*, 358−367.
(31) de Oliveira, V. M.; de Godoi Contessoto, V.; da Silva, F. B.; Caetano, D. L. Z.; de Carvalho, S. J.; Leite, V. B. P. Effects of pH and salt concentration on stability of a protein G variant using coarse-grained models. *Biophys. J.* **2018**, *114*, 65−75.
(32) Ullner, M.; Woodward, C. E.; Jönsson, B. A Debye-Hückel Theory for Electrostatic Interactions in Proteins. *J. Chem. Phys.* **1996**, *105*, 2056−2065.
(33) Tan, Z.-J.; Chen, S.-J. Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte. *J. Chem. Phys.* **2005**, *122*, 044903.
(34) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.
(35) Noel, J. K.; Levi, M.; Raghunathan, M.; Lammert, H.; Hayes, R. L.; Onuchic, J. N.; Whitford, P. C. SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLoS Comput. Biol.* **2016**, *12*, No. e1004794.
(36) Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; Nola, A. D.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(37) Day, R.; Paschek, D.; Garcia, A. E. Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1889−1899.
(38) Qi, R.; Wei, G.; Ma, B.; Nussinov, R. Replica Exchange Molecular Dynamics: A Practical Application Protocol with Solutions to Common Problems and a Peptide Aggregation and Self-Assembly Example. *Methods Mol. Biol. (N. Y., NY, U. S.)* **2018**, *1777*, 101−119.
(39) Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J. Jalview Version 2−a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189−1191.
(40) Katoh, K.; Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772−780.
(41) Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **2000**, *17*, 540−552.
(42) Darriba, D.; Taboada, G. L.; Doallo, R.; Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **2011**, *27*, 1164−1165.
(43) Le, S. Q.; Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **2008**, *25*, 1307−1320.
(44) Ashkenazy, H.; Penn, O.; Doron-Faigenboim, A.; Cohen, O.; Cannarozzi, G.; Zomer, O.; Pupko, T. FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* **2012**, *40*, W580−584.
(45) Yang, J.; Roy, A.; Zhang, Y. Protein−ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* **2013**, *29*, 2588−2595.
(46) Yang, J.; Roy, A.; Zhang, Y. BioLiP: a semi-manually curated database for biologically relevant ligand−protein interactions. *Nucleic Acids Res.* **2012**, *41*, D1096−D1103.
(47) Silva, S. B.; Pinheiro, M. P.; Fuzo, C. A.; Silva, S. R.; Ferreira, T. L.; Lourenzoni, M. R.; Nonato, M. C.; Vieira, D. S.; Ward, R. J. The role of local residue environmental changes in thermostable mutants of the GH11 xylanase from Bacillus subtilis. *Int. J. Biol. Macromol.* **2017**, *97*, 574−584.
(48) Alponti, J. S.; Fonseca Maldonado, R.; Ward, R. J. Thermostabilization of Bacillus subtilis GH11 xylanase by surface charge engineering. *Int. J. Biol. Macromol.* **2016**, *87*, 522−528.
(49) Ferreiro, D. U.; Komives, E. A.; Wolynes, P. G. Frustration, function and folding. *Curr. Opin. Struct. Biol.* **2018**, *48*, 68−73.
(50) Clementi, C.; Plotkin, S. S. The effects of nonnative interactions on protein folding rates: theory and simulation. *Protein Sci.* **2004**, *13*, 1750−1766.
(51) Sutto, L.; Lätzer, J.; Hegler, J. A.; Ferreiro, D. U.; Wolynes, P. G. Consequences of localized frustration for the folding mechanism of the IM7 protein. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 19825−19830.
(52) Chahine, J.; Oliveira, R. J.; Leite, V. B.; Wang, J. Configuration-dependent diffusion can shift the kinetic transition state and barrier height of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14646−14651.
(53) Wakarchuk, W. W.; Sung, W. L.; Campbell, R. L.; Cunningham, A.; Watson, D. C.; Yaguchi, M. Thermostabilization of the Bacillus circulans xylanase by the introduction of disulfide bonds. *Protein Eng., Des. Sel.* **1994**, *7*, 1379−1386.

https://doi.org/10.1021/acs.jpcb.1c01253
J. Phys. Chem. B 2021, 125, 4359−4367