JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2020, 16, 765−772

pubs.acs.org/JCTC

pH and Charged Mutations Modulate Cold Shock Protein Folding and Stability: A Constant pH Monte Carlo Study

Vinícius M. de Oliveira,[†,∥] Daniel L. Z. Caetano,[‡,∥] Fernando B. da Silva,[‡] Paulo R. Mouro,[‡] Antonio B. de Oliveira, Jr.,[‡] Sidney J. de Carvalho,[‡] and Vitor B. P. Leite*[,‡,§]

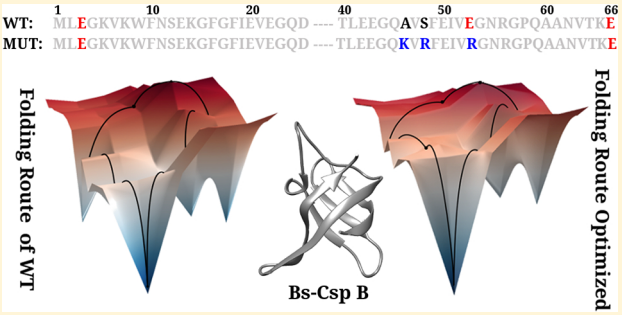
[†]Brazilian Biosciences National Laboratory, National Center for Research in Energy and Materials, LNBio/CNPEM, Campinas, São Paulo, 13083-970, Brazil

[‡]Department of Physics, São Paulo State University (UNESP), Institute of Biosciences, Humanities and Exact Sciences, São José do Rio Preto, São Paulo, 15054-000, Brazil

[§]Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States

Supporting Information

ABSTRACT: The folding and stability of proteins is a fundamental problem in several research fields. In the present paper, we have used different computational approaches to study the effects caused by changes in pH and for charged mutations in cold shock proteins from *Bacillus subtilis* (Bs-CspB). First, we have investigated the contribution of each ionizable residue for these proteins to their thermal stability using the TKSA-MC, a Web server for rational mutation via optimizing the protein charge interactions. Based on these results, we have proposed a new mutation in an already optimized Bs-CspB variant. We have evaluated the effects of this new mutation in the folding energy landscape using structure-based models in Monte Carlo simulation at constant pH, SBM-CpHMC. Our results using this approach have indicated that the charge rearrangements already in the unfolded state are critical to the thermal stability of Bs-CspB. Furthermore, the conjunction of these simplified methods was able not only to predict stabilizing mutations in different pHs but also to provide essential information about their effects in each stage of protein folding.



1. INTRODUCTION

The protein folding process, although theoretically characterized by the presence of an infinity of possible partially folded structures, can exhibit some conformations which have a higher probability of being formed during the transition stages, known as folding nuclei.[1−3] The preference for certain structures shows that the folding of a protein can be governed by the sequence of amino acid residues[4] and by the environment in which that protein is folded.[5−9] Local and nonlocal interactions between amino acid residues can optimize the folding of a protein, with high folding rate, and/or giving the native structure greater thermostability.[10−12] Both the amino acid residue sequence and the environment are responsible for assigning variations in the protein charge distribution, and this variation can be considered one of the causes in the change of the folding process mechanism.[9]

The protein charge distribution depends on its conformational state and the protonation states of the amino acid residues that compose it. Variations in ionic strength and the pH of the environment also influence the charge distribution.[13−18] According to the charge distribution, the electrostatic interactions between amino acid residues are responsible for influencing both the stability of the native protein and its function.[19−24] Given the importance of charged residues, such interactions have been massively studied in the folding process, in both theory and in vitro experiments.[15,25,26] The search for more thermostable enzymes with potential application in the industry has attracted interest for the folding process of variant proteins. Through point mutations, it is sought to change the charge distribution and to optimize protein stability without affecting its function.[27,28] Moreover, stability changes upon mutations are crucial features for genomic diseases such as cancers.[29−31] Therefore, the understanding of this process can provide useful information for medicine and the development of new cancer therapies.

Computational simulations that take into account the energy-landscape theory have been successful in explaining several aspects regarding the protein folding process.[3,17,32−35] Although electrostatic interactions are directly related to how proteins fold, there are few computational works that study the folding process in different pH conditions, allowing a protonation/deprotonation process during the folding.[13,17] In this work, we discuss the effects caused by pH variations in the

Received: September 6, 2019
Published: November 22, 2019

Downloaded via RICE UNIV on June 2, 2021 at 13:56:13 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

© 2019 American Chemical Society

DOI: 10.1021/acs.jctc.9b00894
J. Chem. Theory Comput. 2020, 16, 765−772

folding of the *Bacillus subtilis* Bs-CspB variant protein,[36,37] proposing mutations capable of increasing the thermostability of this protein and elucidating how pH modulates the folding mechanism of this protein, its variant, and mutated forms.

The Bs-CspB variant protein, which is the target of this work, is a globular protein with 66 amino acid residues and PDB ID 2I5M.[37] The variant character of this protein is due to the presence of two mutations that add charged residues in the protein: A46K and S48R. A cartoon of the Bs-CspB variant is shown in Figure 1. The high stability, even in the absence of

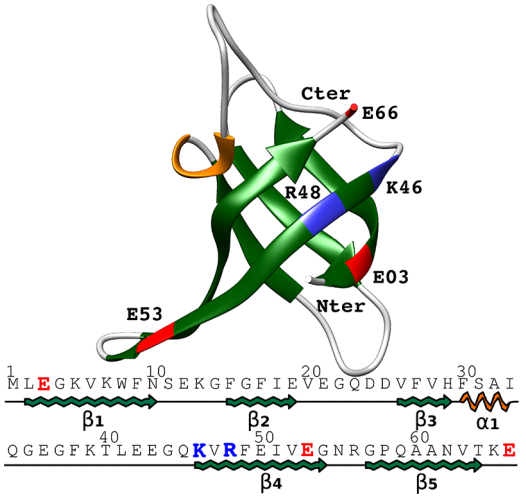


Figure 1. Structure of the *B. subtilis* cold shock protein Bs-CspB variant A46K S48R (PDB ID 2I5M). The mutated residues K46 and R48 are highlighted in blue. The negatively charged residues E03, E53, and E66 are highlighted in red. Below, the primary structure of Bs-CspB variant is presented and the secondary structures are indicated.

bound metals, coupled with the fact that Bs-CspB is a small and well in vitro characterized protein, makes the Bs-CspB variant an ideal model for the study of the relationship between electrostatic interactions and thermostability.[24,37−40]

The objective of this study is to explore and compare the thermodynamic folding aspects of the Bs-CspB and its mutants using a Cα coarse-grained model, in different pH conditions. It is known that the A46K and S48R mutations are capable of optimizing the stability of the Bs-CspB protein.[36,37] However, in this work, we propose a new mutation capable of optimizing the thermostability of this Bs-CspB variant and present evidence of the effect that electrostatic rearrangement has on the folding pathway and stability of this protein in its wild and variant forms. We use the TKSA-MC server[41] to seek rational mutations capable of optimizing the thermostability of the folded protein. Also, a new approach to simulate the folding at constant pH using a Monte Carlo (MC) algorithm is introduced—CpHMC—so we can analyze the effect of mutations on this folding process. The energy landscape of this process has been explored using the Energy Landscape Visualization Method (ELViM),[42] and the differences present in each pH case were shown to be fundamental to protein stability. Thus, the combination of these methodologies can be used as a viable and fast alternative for protein engineering and to understand how rational mutations can change the folding and stability in different pH environments.

2. METHODS

2.1. The TKSA-MC Server. The TKSA-MC is a Web server developed by our research group and is used to calculate the electrostatic interactions between all ionizable residues of globular proteins via the Tanford−Kirkwood method, which includes a solvent accessibility correction.[41] The all-atoms coordinate of the native structure of the protein is used as an input of this program, and the Metropolis Monte Carlo algorithm samples the protonation states of ionizable atoms. The server provides quantitative information about the importance of each residue to protein thermal stability, and it has been a robust tool to find rational mutations that improve thermal stability in different proteins.[17,41,43] This tool was used to select a specific residue to mutate and make the Bs-CspB variant even more thermostable.

2.2. Structure-Based Model in Monte Carlo Simulation at Constant pH (SBM-CpHMC). The wild-type Bs-CspB and its variants were modeled according to the structure-based model (SBM), where the residues of the protein are represented by single beads centered in the Cα position.[13,34,44,45] These beads are connected by harmonic springs and a nonelectrostatic interaction energy; the effective potential energy $U(\Gamma,\Gamma_0)$ has its global minimum associated with its native conformation Γ_0, and is given by

$$U(\Gamma,\Gamma_0) = \sum_{\text{bonds}} \epsilon_r (r - r_0)^2 + \sum_{\text{angles}} \epsilon_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \epsilon_\phi \left\{ [1 - \cos(\phi - \phi_0)] + \frac{1}{2}[1 - \cos(3(\phi - \phi_0))] \right\} + \sum_{\text{contacts}} \epsilon_c \left[5 \left(\frac{d_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{d_{ij}}{r_{ij}} \right)^{10} \right] + \sum_{\text{noncontacts}} \epsilon_{nc} \left(\frac{\sigma_{nc}}{r_{ij}} \right)^{12} + \sum_{\text{electrostatics}} U_{\text{res}}^{\text{el}} \quad (1)$$

where r, θ, and ϕ are defined by the distance between two consecutive residues, the angles formed by three consecutive residues, and the dihedral angles formed by four consecutive residues, respectively. The values of r_0, θ_0, and ϕ_0 were obtained from the native structure of Bs-CspB. The parameter d_{ij} is the native distance between the pair contact i and j ($i < j - 3$) determined by a CSU contact map (Contacts of Structural Units).[46] The parameters $\epsilon_r = 100\epsilon_c$, $\epsilon_\theta = 20\epsilon_c$, and $\epsilon_\phi = \epsilon_{nc} = \epsilon_c$ are defined as a function of the Lennard-Jones 10−12 parameter ϵ_c. In this study, we define $\epsilon_c = 3.14$ kJ/mol. Finally, we have $\sigma_{nc} = 4.0$ Å,[34,44] whereas the last term corresponds to the electrostatic interaction energy between any charged residues and is given by the screened Coulomb potential

$$U_{\text{res}}^{\text{el}}(r_{ij}) = \frac{z_i z_j e^2}{4\pi\epsilon_0\epsilon_s} \frac{\exp(-\kappa r_{ij})}{r_{ij}} \quad (2)$$

where z_i and z_j are the respective valencies of residues i and j, r_{ij} is the radial distance between them, ϵ_0 is the permittivity of free space, $\epsilon_s = 80.0$ is the dielectric constant, and $\kappa = 0.0734$ Å^{-1} is the inverse of the Debye length,[47] which corresponds to a salt concentration of approximately 50 mM.

DOI: 10.1021/acs.jctc.9b00894
J. Chem. Theory Comput. 2020, 16, 765−772

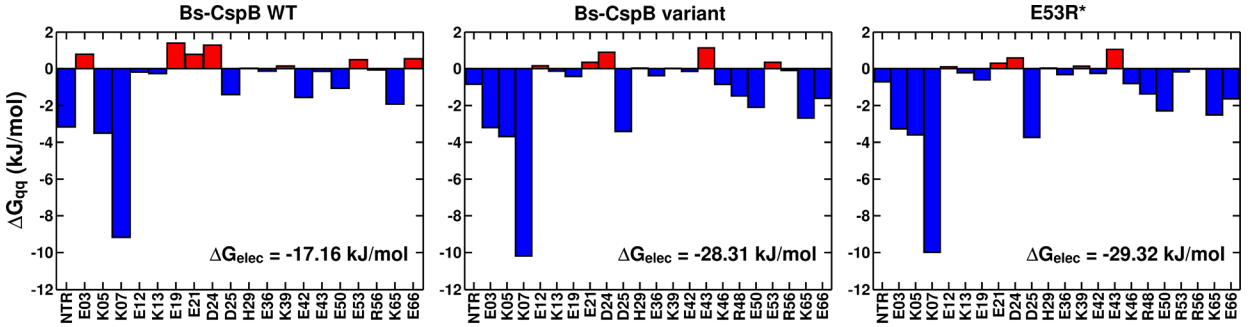


Figure 2. Charge−charge interaction energy ΔG_{qq} calculated by the TKSA-MC model for each ionizable residue at pH 7.0. The three energy profiles for the Bs-CspB WT, Bs-CspB variant, and E53R* are presented. Red bars indicate the residues with the side chain exposed to solvent with SASA ≥ 50% and positive energy contribution to native state stability.

The protein was confined within a simulation cell large enough to eliminate confining effects on its conformation. In order to sample the protein configurations, we have performed Metropolis Monte Carlo simulations, in canonical ensemble, using three movements: (1) translational displacement of a single residue randomly chosen, (2) pivot rotation, and (3) crankshaft movement.[48,49] In each Monte Carlo step, one of these movements is randomly chosen with the same probability. In addition to these movements, one tries to change the protonation state of a given titratable residue during the simulation with a probability of 0.2. More specifically, a titratable residue is randomly chosen, and its protonation state is changed or not according to the Metropolis criterion with energy variation given by eq 3:[13,17,50]

$$\Delta U^{\mathrm{tit}} = \Delta U^{\mathrm{el}} \pm \chi[k_{\mathrm{B}}T(\mathrm{pH} - \mathrm{p}K_{\mathrm{a}}^{0}) \ln 10] \quad (3)$$

where k_B is the Boltzmann constant, T is the temperature, and $\chi = 1$ or -1 if a basic or acid residue is selected, respectively. The positive and negative signs are used for protonation and deprotonation, respectively. The first term in eq 3 corresponds to the electrostatic energy variation (eq 2); the second term corresponds to the contribution due to the proton bath at a given concentration determined by pH; and the last one is the contribution of the compound model. The values of pK_a^0 of the titratable residues adopted in this study are 4.0 for aspartic acid, 4.5 for glutamic acid, 6.3 for histidine, 10.6 for lysine, and 12.0 for arginine.[17] Briefly, for each temperature considered (we have used a set of 40 different temperatures ranging from 190 to 490 K), we start the simulations with the protein in its native state with all titratable residues charged; that is, the basic residues are protonated while the acidic ones are deprotonated. Then, the protein configurations are sampled using three movements (translational displacement of a single residue, pivot rotation, and crankshaft) and changing the state of protonation of the titratable residues. The equilibration process was carried out with 10^7 Monte Carlo steps, and the average properties were calculated using 10^7 statistically uncorrelated configurations of the protein.

2.3. Folding Routes. In the protein folding process, the fraction of native contacts is used as the reaction coordinate to detail the folding. $Q \approx 1.0$ represents the protein in its native state, whereas with $Q \approx 0$ it is in its unfolded state.[34] The native contacts define the folded and unfolded states, but information about its folding pathway is not provided by the reaction coordinate. To fulfill this lack of information, we have used the folding route measure, $R(Q)$, which quantifies the specificity of the folding pathway among all possible pathways that lead the protein to its native state:[51]

$$R(Q) = \sum_{i=1}^{N} \frac{\langle(\langle Q_i\rangle_Q - Q)^2\rangle_Q}{NQ(1-Q)} \quad (4)$$

in which the $R(Q)$ value ranges from 0 to 1. An $R(Q)$ value of near 0 means that the specificity is low; thus any specific contact tends to present the same probability of all other native contacts. On the other hand, $R(Q) = 1$ means that the protein presents only one populated structure that indicates a particular path emerges, while the rest is unreachable. In eq 4, the total number of native contacts is represented by N. Q_i is equal to 1 when the contact i is formed; otherwise Q_i is zero. The average $\langle Q_i\rangle_Q$ is calculated over all configurations that have the same Q.

2.4. Energy Landscape Visualization Method. In order to visualize biological macromolecule landscapes, we address this problem using a metric based on internal distances between amino acids to describe the differences between any two conformations. Using an effective projection method, we are able to go beyond the usual one-dimensional representation and, thus, visualize landscapes in two dimensions. The main concept is based on the assumption that energy states are highly correlated with their molecular conformations. We refer to the Energy Landscape Visualization Method as ELViM.[42]

The visualization method is based on four steps: (i) an ensemble of structures which, in general, are obtained through simulations is generated; (ii) the dissimilarity matrix is calculated by applying a metric throughout the simulated trajectory; (iii) a data processing procedure is carried out to cluster very similar structures into a single conformation; (iv) a multidimensional projection is performed and, then, the dissimilarity matrix is transformed to a two-dimensional (2D) projection.[42]

In summary, ELViM can automatically detect pathways and competing transition state ensembles without the need for an a priori reaction coordinate. Instead of having to intuitively design coordinates, which can involve visual inspection of large numbers of transition events, ELViM provides a means for efficiently mapping structurally distinct pathways during elaborate conformational rearrangement in multidomain, or multicomponent, assemblies.

3. RESULTS AND DISCUSSION

3.1. Stabilizing Mutations. We used the TKSA-MC calculations to understand the electrostatic effects of charged

DOI: 10.1021/acs.jctc.9b00894
J. Chem. Theory Comput. 2020, 16, 765−772

mutations on the stability of the native state of Bs-CspB. The electrostatic profiles of each ionizable residue of the Bs-CspB wild type (WT) and two mutants at pH 7.0 are presented in Figure 2. We can identify the charged residues that are acting to stabilize the protein folded structure (amino acids with negative values of ΔG_{qq}), and those that tend to destabilize this structure, which have positive values of ΔG_{qq}. Red bars mean that the residue has a positive ΔG_{qq} and its side chain is highly exposed to the solvent (solvent-accessible surface area (SASA) > 50%). For these residues, mutations with opposite charge tend to stabilize the native state of the protein, whereas the change of residues with strongly negative ΔG_{qq} usually promotes a destabilization.[17,43]

The Bs-CspB WT (PDB ID 1CSP) presents a ΔG_{elec} (sum of all ΔG_{qq}) of −17.16 kJ/mol. The mutations A46K and S48R present in Bs-CspB variant (Figure 1) promote a decrease of 11.15 kJ/mol, resulting in a ΔG_{elec} = −28.31 kJ/mol. This free energy variation is consistent with experimental data that indicates a stabilization of 11.1 kJ/mol in Bs-CspB variant in comparison with the WT one.[37] The insertion of positive charges in positions 46 and 48 tends to stabilize the C- and N-terminal neighborhood of Bs-CspB, mostly residues E03 and E66 (both are highlighted in Figure 1). The unstable region formed by negative residues between E19 and D24 in WT protein is also stabilized by these mutations. Thus, the Bs-CspB variant becomes 16.9 K more thermostable than its WT.[37]

Although the Bs-CspB variant is already optimized, there still are residues with positive ΔG_{qq} which are potential mutation targets to improve its stability. Here we presented one of these targets: the mutation E53R* (Bs-CspB variant with a new mutation E53R). The electrostatic profile of E53R* presents a ΔG_{elec} that decreases by 1.0 kJ/mol in comparison with the Bs-CspB variant. The electrostatic effect of this mutation is also tested in a pH range 1.0−12.0 (Figure 3). From pH 4.0 to

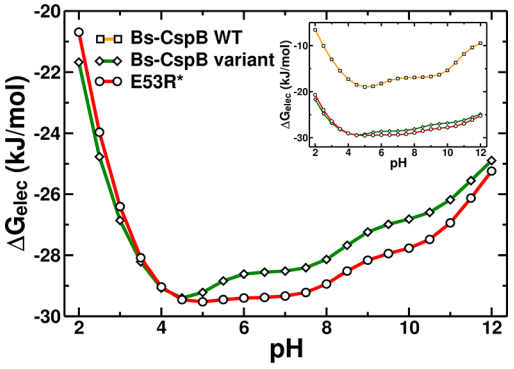


Figure 3. Electrostatic energy contribution to free energy native state stability ΔG_{elec} (sum of all ΔG_{qq} values) as a function of pH in kJ/mol for Bs-CspB WT (yellow, inset figure), variant (green), and E53R* (red). The values were calculated from TKSA-MC simulations in a pH range 2.0−12.0.

12.0, E53R* tends to be more stable than CspB variant, presenting lower values of ΔG_{elec}. On the other hand, the Bs-CspB variant becomes more stable at low pH values. For all the pH range, the Bs-CspB WT has a lower stability than its variant and E53R*. The TKSA-MC method provided possible target mutations to improve thermal stability; however, this method is limited to calculating the contribution of electrostatic interaction in the native state of the protein. For that reason, in sections 3.2 and 3.3, we concentrate our efforts to understand how E53R* can also affect the entire folding process using SBM-CpHMC simulations at three different pH conditions (2.0, 7.0, and 11.0).

3.2. Thermodynamic Properties. The heat capacities C_v of the Bs-CspB WT, variant, and E53R*, for three pH conditions, are presented in Figure 4, parts A, B, and C,

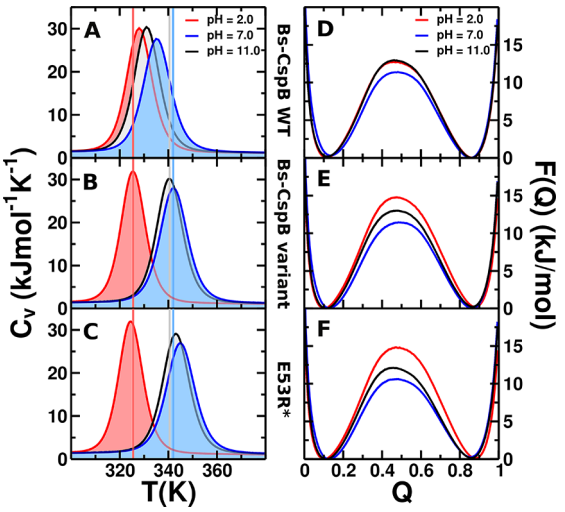


Figure 4. Heat capacity C_v as a function of temperature for the Bs-CspB WT (A), variant (B), and mutant E53R* (C). Free energy $F(Q)$ as a function of fraction of formed native contacts Q for the Bs-CspB WT (D), variant (E), and mutant E53R* (F). Black, blue, and red curves mean pH 11.0, 7.0, and 2.0, respectively. Vertical lines in (A), (B), and (C) are used as a guide to the eye placed on the C_v peaks of Bs-CspB variant protein.

respectively. These results are in agreement with the TKSA-MC calculations, where Bs-CspB is more stable at pH 7.0, followed by pH 11.0 and 2.0, for the WT protein and the mutant cases. In pH 7.0, there is an increase near 7.0 K in the T_M (the temperature at the C_v peak) of the Bs-CspB variant in comparison to the WT one, whereas experimental results show a difference of 16.9 K in thermostability of these proteins. This quantitative difference between experimental and computational data was expected, since the Cα models are too simple at this level of detailed comparison. However, the qualitative agreement is maintained. Still for pH 7.0, the proposed mutant E53R* has its T_M increased close to 3.0 K compared to the protein variant and 10.0 K when compared with Bs-CspB WT.

The Bs-CspB variant and E53R* are considerably less stable at low pH compared to pH 7.0 and 11.0 (almost 20.0 K). Moreover, in pH 2.0, these two proteins become more unstable than Bs-CspB WT. This behavior is unexpected, considering the results of ΔG_{elec} from the TKSA-MC, which indicate that the WT protein would be less stable even in low pH. These result differences indicate the importance of the analysis of the entire folding for the determination of the protein stability. Once the entire folding is analyzed, we show in section 3.3 that the addition of positively charged residues on this protein mainly affects the folding in its unfolded stage, an aspect that the TKSA-MC method does not take into account this state.

Figure 4D−F shows the free energy of the folding process of Bs-CspB and its mutants in the T_M of each case. The free energy was calculated via the weighted histogram analysis

DOI: 10.1021/acs.jctc.9b00894
J. Chem. Theory Comput. 2020, 16, 765−772

method (WHAM) using a set of simulations in 40 different temperatures around T_M.[52] Bs-CspB is a two-state folding protein and presents free energy curves with a minimum in the unfolded state ($Q < 0.3$), a folding barrier in the transition state (TS, for $0.4 < Q < 0.6$), and a folded/native state ($Q > 0.7$). The energy barrier for the three proteins is lower at pH 7.0 than at pH 11.0 and 2.0. The difference in the height of the free energy barriers for each pH varies depending on the mutations. These results indicate a correlation between the free energy barrier of Bs-CspB and its thermostability, since the lower the barrier, the greater T_M. This effect in the free energy barrier can also be associated with the differences in the possible paths covered by the protein during the folding process in each pH, and it is discussed in detail in section 3.3.

3.3. The Folding Process. To understand the charge effect in the protein stabilities, we explored each folding stage of theirs. Figure 5 presents the route measures of Bs-CspB WT, variant, and E53R* at pH 2.0, 7.0, and 11.0. $R(Q)$ indicates the specificity of the routes that lead the unfolded protein to its native state. The higher $R(Q)$, the greater the specificity; i.e., fewer paths are accessed by the protein during the folding. For the case of Bs-CspB WT, the pH changes do not alter the $R(Q)$ curves significantly. In the three pHs presented, the folding specificity is high in the early TS, with Q between 0.3 and 0.4, and abruptly decreases already in the TS ($Q \approx 0.5$). This behavior indicates a difficulty for the protein to find nonfrustrated paths in its unfolded state; once this ensemble is reached, the folding routes become nonspecific. A similar effect occurs in the two mutated proteins presented here, that is, very low specificity in folding paths after TS, so the most significant difference between them lies in the early TS and the TS.

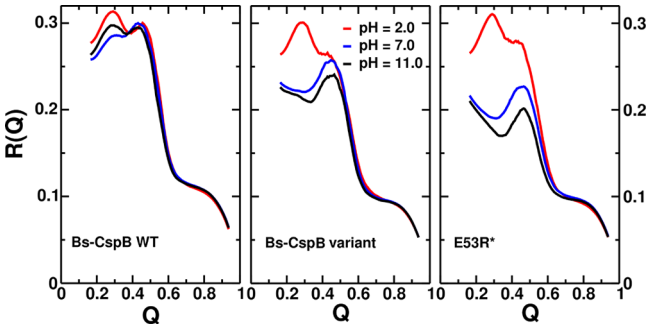


Figure 5. Folding $R(Q)$ as a function of fraction of native contacts Q for Bs-CspB WT, variant, and E53R*. pH 11.0, 7.0, and 2.0 are represented by black, blue, and red curves, respectively.

Unlike the case of Bs-CspB WT, the pH changes affect the folding routes of its mutants considerably. While for pH 2.0 the $R(Q)$ for the mutants follows a behavior similar to the one of the WT protein, for pH 7.0 and 11.0 there is a decrease of $R(Q)$ in the unfolded state. Also, for these two pHs, a peak in the route values arises in the TS. For E53R*, the route measurements at pH 7.0 and 11.0 in the unfolded state are even lower than those for the Bs-CspB variant. These results indicate a correlation between the folding routes of Bs-CspB at the early TS with thermal stability; high values of $R(0.3) \approx 0.3$ are present in the proteins with lower stability, whereas the decrease of $R(Q)$ in this folding stage is related to a thermal stabilization (the cases of Bs-CspB variant and E53R* at pH 7.0 and 11.0). Thus, we have evidence of the importance of the charge rearrangement already in the early TS for the protein native state stability. These results are in agreement with the findings of Tzul and collaborators.[53] Their studies showed that the increase in thermostability of four different proteins is most associated with changes in the unfolded state and the TS stage

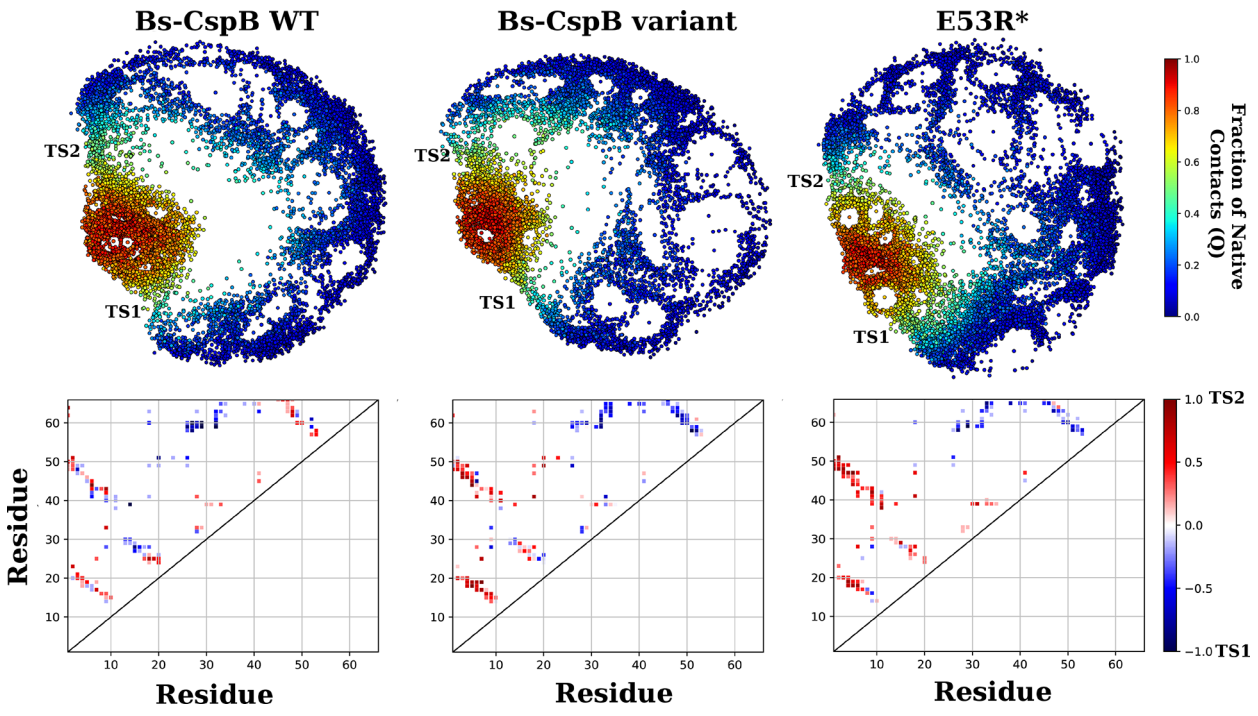


Figure 6. Folding trajectory of Bs-CspB mapped into 2D phase space using ELViM. The Euclidean distance between conformations (where each point represents a unique structure) is relative to structural differences. The color scheme is associated with fraction of native contacts (Q). The differences between the contact maps of TS1 and TS2 ensembles are shown below. Red points represent contact formation preferably in TS2, and blue points are preferably formed in TS1.

DOI: 10.1021/acs.jctc.9b00894
J. Chem. Theory Comput. 2020, 16, 765−772

of the protein, and some charged mutations can decrease the frustration in the energy landscape, promoting higher compactness of the unfolded protein.

The folding energy landscape of Bs-CspB WT and its mutants was evaluated by ELViM. This energy landscape visualization tool provides information about the most significant folding pathways, the protein structures in each stage, and the energies involved. In Figure 6, the visualizations of Bs-CspB WT, variant, and E53R* are presented at their folding temperatures, all of them at pH 7.0 (the cases of pH 2.0 and 11.0 are presented in Figure S1 of the Supporting Information). The Euclidean distance in the 2D map between the points represents their structural differences; i.e., the higher the distance, the greater the dissimilarity between their structures. In these visualization maps, the color means the fraction of native contacts (Q), with blue and red corresponding to unfolded and folded states, respectively. Light blue points in this map correspond to the early TS region, and for the Bs-CspB WT, they occur less often compared to the mutant proteins. This feature seems to explain the high values of $R(Q)$ for the WT in its unfolded state. Also, there is a large gap between the unfolded and folded ensembles for Bs-CspB WT. This gap tends to decrease when the mutations are made, suggesting the ensemble of unfolded states is more compact for the mutants.

In all three cases, two dominant paths lead the protein from its unfolded state to its native structure: the TS1 and TS2 regions, as indicated in Figure 6. The differences between the contact maps of TS1 and TS2 regions are presented in the lower part of Figure 6, where blue points represent the frequently formed contacts in TS1 and red ones represent those in TS2. For Bs-CspB WT, the contacts formed in the region near the N-terminal and the C-terminal have a mix of red and blue points. This means that the WT protein can fold both regions during TS1 and TS2. However, for the two mutants, the contacts formed near the N-terminal are significantly preferable in the TS2 (red points) and there is a tendency for the formation of β-sheets 1, 2, and 3, whereas the contacts in the C-terminal region are made mostly in TS1 (blue points), with β-sheets 4 and 5 tending to be formed. This effect can explain the appearance of the peak in the $R(Q)$ curves in TS for Bs-CspB variant and E53R*. Some representative structures of TS1 and TS2 are present in Figure S2.

The average total electrostatic energy $\langle E^{el}_{tot}\rangle$ and the average specific contribution of residue 53 for this energy $\langle E^{el}_{53}\rangle$, i.e., the electrostatic interaction of residue 53 with all other charged residues, are presented in Figure 7 for each folding stage at pH 2.0, 7.0, and 11.0. The electrostatic energy tends to increase during the Bs-CspB folding at pH 2.0 due to the excess of positive charges approaching in this process. The insertion of an arginine in position 53 also increases the net charge at low pH, thereby causing destabilization of the protein in this environment. Also, the high values of $R(Q)$ presented in Figure 5 for E53R* at pH 2.0 is caused by this excess of positive charges, which promotes the increase of frustration by repulsive electrostatic interactions and, consequently, the possible folding paths are reduced.

At pH 7.0, Bs-CspB WT, variant, and E53R* presented a negative electrostatic energy value for all folding stages. However, for the WT protein, these values are near zero. The most negative total electrostatic energy is for E53R* in this pH condition; also in this condition, the highest stability is found. Regarding the contribution of residue 53 for Bs-CspB WT, its electrostatic energy is positive in all protein stages and tends to increase during the folding. In the case of Bs-CspB variant, the glutamic acid stabilizes the unfolded state and destabilizes the TS and the folded one, whereas the arginine of E53R* promotes precisely the opposite effect. The specific contribution of arginine 53 to stabilize the native state of this protein is even higher at pH 11.0 than at pH 7.0. However, the total electrostatic energy is more positive than at pH 7.0. That is an effect caused by the deprotonation tendency of the basic residues at high pHs.

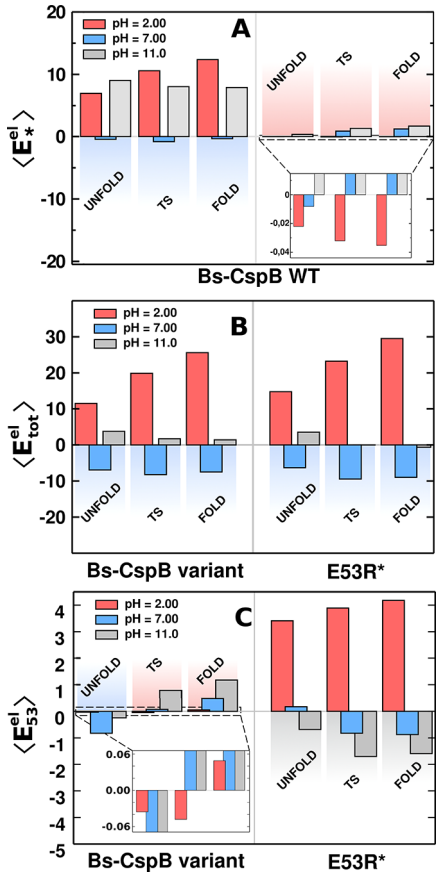


Figure 7. (A) Average of the electrostatic energy between all charged residues $\langle E^{el}_{tot}\rangle$ (left) and average of electrostatic energy between residue 53 and all other charged residues $\langle E^{el}_{53}\rangle$ (right). Results for Bs-CspB WT. (B) $\langle E^{el}_{tot}\rangle$ for Bs-CspB variant and E53R*. (C) $\langle E^{el}_{53}\rangle$ for Bs-CspB variant and E53R*. Red, blue, and gray bars for pH 2.0, 7.0, and 11.0, respectively, in kJ/mol.

4. CONCLUSIONS

In this paper, we combined two Monte Carlo methods to understand the changes promoted by charge mutations in the folding and stability of Bs-CspB. Using the TKSA-MC method, we were able to identify the ionizable residues responsible for stabilizing and destabilizing the native state of the Bs-CspB WT and its optimized variant (A46K and S48R). Furthermore, we proposed a new stabilizing mutation (E53R*) in this variant and tested it at different pH conditions. While TKSA-MC was effective to quantify the charge−charge interactions for the stability of the native structure, SBM-CpHMC can explore these interactions in the entire folding process. The thermal stability results obtained from these simulations are in qualitative agreement with experimental ones. Moreover, the

DOI: 10.1021/acs.jctc.9b00894
J. Chem. Theory Comput. 2020, 16, 765−772

application of ELViM using the trajectories from SBM-CpHMC provided us with a three-dimensional view of the energy landscape, showing differences in folding routes caused by each mutation in alkaline, acid, and neutral pHs. Based on these results, we found that interchain electrostatic interactions during the early transition state seem to be crucial to Bs-CspB stability. Finally, the specific effect of the mutation E53R in Bs-CspB variant was explored in each folding stage, which reinforced the importance of the charge rearrangement already in the unfolded state for protein stability. This rearrangement enhances the formation of favorable contacts in the early TS and the TS and provides a less frustrated energy landscape between these two states. The present work introduces the SBM-CpHMC tool to study the pH and charge mutation effects on proteins. The simulations using this simplified method can provide valuable information about folding thermodynamic properties. Its use in conjunction with TKSA-MC and ELViM can be effective to predict stabilizing mutations and understand how they can modulate the protein energy landscape.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.9b00894.

Folding trajectory of Bs-CspB mapped into 2D phase space using ELViM for pH 2.0 and 11.0; examples of structures from E53R* in transition states TS1 and TS2, respectively (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: vitor.leite@unesp.br.

ORCID

Vinícius M. de Oliveira: 0000-0003-0927-3825
Daniel L. Z. Caetano: 0000-0002-0476-3115
Fernando B. da Silva: 0000-0002-0285-8700
Vitor B. P. Leite: 0000-0003-0008-9079

Author Contributions

‖V.M.O. and D.L.Z.C.: These authors contributed equally to this work.

Funding

V.M.O. was supported by São Paulo Research Foundation (FAPESP), Grant No. 2018/11614-3. D.L.Z.C. thanks the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES), Finance Code 001 for financial support. F.B.S. was supported by the National Council for Scientific and Technological Development (CNPq, Grant Process No. 141715/2017-0). S.J.C. is grateful to the São Paulo Research Foundation (FAPESP), Process Grant No. 2018/01841-2. V.B.P.L. was supported by the São Paulo Research Foundation (FAPESP), Grant Nos. 2018/18668-1 and 2016/19766-1, and was also partially supported by the Center for Theoretical Biological Physics sponsored by the National Science Foundation NSF Grant PHY-1427654.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

All simulations were performed at the Center for Scientific Computing (NCC/GridUNESP) of the São Paulo State University (UNESP).

■ REFERENCES

(1) Martinez, J. C.; Pisabarro, M. T.; Serrano, L. Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat. Struct. Mol. Biol.* **1998**, *5*, 721−729.
(2) Grantcharova, V. P.; Riddle, D. S.; Santiago, J. V.; Baker, D. Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat. Struct. Mol. Biol.* **1998**, *5*, 714−720.
(3) Chavez, L. L.; Onuchic, J. N.; Clementi, C. Quantifying the roughness on the free energy landscape: Entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc.* **2004**, *126*, 8426−8432.
(4) Mirny, L.; Shakhnovich, E. Evolutionary conservation of the folding nucleus. *J. Mol. Biol.* **2001**, *308*, 123−129.
(5) O'Brien, E. P.; Brooks, B. R.; Thirumalai, D. Effects of pH on Proteins: Predictions for Ensemble and Single-Molecule Pulling Experiments. *J. Am. Chem. Soc.* **2012**, *134*, 979−987.
(6) Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci.* **2008**, *18*, 247−251.
(7) Kukić, P.; Farrell, D.; Sondergaard, C. R.; Bjarnadottir, U.; Bradley, J.; Pollastri, G.; Nielsen, J. E. Improving the analysis of NMR spectra tracking pH-induced conformational changes: Removing artefacts of the electric field on the NMR chemical shift. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 971−984.
(8) Krittanai, C.; Johnson, W. C., Jr The relative order of helical propensity of amino acids changes with solvent environment. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 132−141.
(9) Goto, Y.; Calciano, L. J.; Fink, A. L. Acid-induced folding of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 573−577.
(10) Abkevich, V.; Gutin, A.; Shakhnovich, E. Impact of Local and Non-local Interactions on Thermodynamics and Kinetics of Protein Folding. *J. Mol. Biol.* **1995**, *252*, 460−471.
(11) Govindarajan, S.; Goldstein, R. A. Optimal local propensities for model proteins. *Proteins: Struct., Funct., Genet.* **1995**, *22*, 413−418.
(12) Muñoz, V.; Serrano, L. Local versus nonlocal interactions in protein folding and stability − an experimentalist's point of view. *Folding Des.* **1996**, *1*, R71−R77.
(13) 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.
(14) Zarrine-Afsar, A.; Zhang, Z.; Schweiker, K. L.; Makhatadze, G. I.; Davidson, A. R.; Chan, H. S. Kinetic consequences of native state optimization of surface-exposed electrostatic interactions in the Fyn SH3 domain. *Proteins: Struct., Funct., Genet.* **2012**, *80*, 858−870.
(15) Lindman, S.; Xue, W.-F.; Szczepankiewicz, O.; Bauer, M. C.; Nilsson, H.; Linse, S. Salting the Charged Surface: pH and Salt Dependence of Protein G B1 Stability. *Biophys. J.* **2006**, *90*, 2911−2921.
(16) Stewart, A. K.; Kurschat, C. E.; Vaughan-Jones, R. D.; Shmukler, B. E.; Alper, S. L. Acute regulation of mouse AE2 anion exchanger requires isoform-specific amino acid residues from most of the transmembrane domain. *J. Physiol.* **2007**, *584*, 59−73.
(17) Martins de Oliveira, V.; 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.
(18) Coronado, M.; Caruso, I.; De Oliveira, V.; Contessoto, V.; Leite, V.; Arni, R.; Eberle, R.; Kawai, L. A. 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.

DOI: 10.1021/acs.jctc.9b00894
J. Chem. Theory Comput. 2020, 16, 765−772

(19) Davis, S. J.; Davies, E. A.; Tucknott, M. G.; Jones, E. Y.; van der Merwe, P. A. The role of charged residues mediating low affinity protein−protein recognition at the cell surface by CD2. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 5490−5494.
(20) Givaty, O.; Levy, Y. Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.* **2009**, *385*, 1087−1097.
(21) 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.
(22) Su, J. G.; Chen, W. Z.; Wang, C. X. Role of electrostatic interactions for the stability and folding behavior of cold shock protein. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 2157−2169.
(23) Ganguly, D.; Zhang, W.; Chen, J. Electrostatically accelerated encounter and folding for facile recognition of intrinsically disordered proteins. *PLoS Comput. Biol.* **2013**, *9*, e1003363.
(24) Gribenko, A. V.; Makhatadze, G. I. Role of the Charge−Charge Interactions in Defining Stability and Halophilicity of the CspB Proteins. *J. Mol. Biol.* **2007**, *366*, 842−856.
(25) 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.
(26) Khandogin, J.; Chen, J.; Brooks, C. L. Exploring atomistic details of pH-dependent peptide folding. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 18546−18550.
(27) Roca, M.; Messer, B.; Hilvert, D.; Warshel, A. On the relationship between folding and chemical landscapes in enzyme catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 13877−13882.
(28) Huber, K.; Olek, K.; Müller, A.; Tan, C.; Bennett, K.; Colinge, J.; Superti-Furga, G. Proteome-wide drug and metabolite interaction mapping by thermal-stability profiling. *Nat. Methods* **2015**, *12*, 1055−1057.
(29) Li, M.; Kales, S. C.; Ma, K.; Shoemaker, B. A.; Crespo-Barreto, J.; Cangelosi, A. L.; Lipkowitz, S.; Panchenko, A. R. Balancing protein stability and activity in cancer: a new approach for identifying driver mutations affecting CBL ubiquitin ligase activation. *Cancer Res.* **2016**, *76*, 561−571.
(30) Li, Y.; Wang, L.; Liu, J.; Zhang, P.; An, M.; Han, C.; Guan, X.; Zhang, K.; Li, Y. O-GlcNAcylation modulates Bmi-1 protein stability and potential oncogenic function in prostate cancer. *Oncogene* **2017**, *36*, 6293.
(31) Stine, Z. E.; Walton, Z. E.; Altman, B. J.; Hsieh, A. L.; Dang, C. V. MYC, metabolism, and cancer. *Cancer Discovery* **2015**, *5*, 1024−1039.
(32) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein-folding - A synthesis. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167−195.
(33) Fersht, A. R.; Matouschek, A.; Serrano, L. The folding of an enzyme: I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **1992**, *224*, 771−782.
(34) 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.
(35) Bruno da Silva, F.; Contessoto, V. G.; de Oliveira, V. M.; Clarke, J.; Leite, V. B. P. Non-Native Cooperative Interactions Modulate Protein Folding Rates. *J. Phys. Chem. B* **2018**, *122*, 10817−10824.
(36) Willimsky, G.; Bang, H.; Fischer, G.; Marahiel, M. A. Characterization of cspB, a Bacillus subtilis inducible cold shock gene affecting cell viability at low temperatures. *J. Bacteriol.* **1992**, *174*, 6326−6335.
(37) Max, K. E.; Wunderlich, M.; Roske, Y.; Schmid, F. X.; Heinemann, U. Optimized Variants of the Cold Shock Protein from in Vitro Selection: Structural Basis of Their High Thermostability. *J. Mol. Biol.* **2007**, *369*, 1087−1097.
(38) Dominy, B. N.; Perl, D.; Schmid, F. X.; Brooks, C. L. The Effects of Ionic Strength on Protein Stability: The Cold Shock Protein Family. *J. Mol. Biol.* **2002**, *319*, 541−554.
(39) Makhatadze, G. I.; Loladze, V. V.; Gribenko, A. V.; Lopez, M. M. Mechanism of Thermostabilization in a Designed Cold Shock Protein with Optimized Surface Electrostatic Interactions. *J. Mol. Biol.* **2004**, *336*, 929−942.
(40) Garcia-Mira, M. M.; Boehringer, D.; Schmid, F. X. The Folding Transition State of the Cold Shock Protein is Strongly Polarized. *J. Mol. Biol.* **2004**, *339*, 555−569.
(41) Contessoto, V. G.; de Oliveira, V. M.; 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.
(42) Oliveira, A. B.; Yang, H.; Whitford, P. C.; Leite, V. B. Distinguishing biomolecular pathways and metastable states. *J. Chem. Theory Comput.* **2019**, *15*, 6482.
(43) Nguyen, C.; Young, J. T.; Slade, G. G.; Oliveira, R. J.; McCully, M. E. A Dynamic Hydrophobic Core and Surface Salt Bridges Thermostabilize a Designed Three-Helix Bundle. *Biophys. J.* **2019**, *116*, 621−632.
(44) Koga, N.; Takada, S. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Gō-like model. *J. Mol. Biol.* **2001**, *313*, 171−180.
(45) Noel, J. K.; Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.* **2010**, *38*, W657−W661.
(46) Sobolev, V.; Sorokine, A.; Prilusky, J.; Abola, E. E.; Edelman, M. Automated analysis of interatomic contacts in proteins. *Bioinformatics* **1999**, *15*, 327.
(47) Ullner, M.; Woodward, C. E.; Jönsson, B. A Debye-Hückel theory for electrostatic interactions in proteins. *J. Chem. Phys.* **1996**, *105*, 2056−2065.
(48) Binder, K. *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*; Oxford University Press: 1995; Vol. *119*.
(49) Attig, N.; Binder, K.; Grubmuller, H.; Kremer, K. *Computational Soft Matter: From Synthetic Polymers to Proteins*; John von Neumann Institute for Computing (NIC): Juelich, Germany, 2004; Vol. *22*.
(50) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; Macmillan: 1999.
(51) Lammert, H.; Noel, J. K.; Onuchic, J. N. The dominant folding route minimizes backbone distortion in SH3. *PLoS Comput. Biol.* **2012**, *8*, e1002776.
(52) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.* **1995**, *16*, 1339−1350.
(53) 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.

DOI: 10.1021/acs.jctc.9b00894
J. Chem. Theory Comput. 2020, 16, 765−772