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

pubs.acs.org/JPCB

Article

Probing the Energy Landscape of Spectrin R15 and R16 and the Effects of Non-native Interactions

Published as part of The Journal of Physical Chemistry virtual special issue "Jose Onuchic Festschrift".

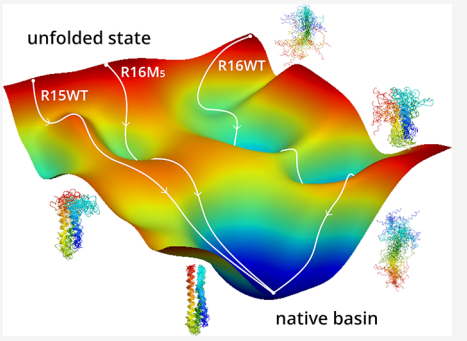
Fernando Bruno da Silva, Vinícius Martins de Oliveira, Antonio Bento de Oliveira Junior, Vinícius de Godoi Contessoto, and Vitor B. P. Leite*

Cite This: *J. Phys. Chem. B* 2023, 127, 1291−1300

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Understanding the details of a protein folding mechanism can be a challenging and complex task. One system with an interesting folding behavior is the α-spectrin domain, where the R15 folds three-orders of magnitude faster than its homologues R16 and R17, despite having similar structures. The molecular origins that explain these folding rate differences remain unclear, but our previous work revealed that a combined effect produced by non-native interactions could be a reasonable cause for these differences. In this study, we explore further the folding process by identifying the molecular paths, metastable states, and the collective motions that lead these unfolded proteins to their native state conformation. Our results uncovered the differences between the folding pathways for the wild-type R15 and R16 and an R16 mutant. The metastable ensembles that speed down the folding were identified using an energy landscape visualization method (ELViM). These ensembles correspond to similar experimentally reported configurations. Our observations indicate that the non-native interactions are also associated with secondary structure misdocking. This computational methodology can be used as a fast, straightforward protocol for shedding light on systems with unclear folding or conformational traps.



INTRODUCTION

The energy landscape theory is crucial for our understanding of protein folding, describing the energy of a protein as a function of its accessible conformations.[1,2] This theory, therefore, encodes the relative stabilities and the energy barriers between molecular events that lead a protein from the unfolded to its native state through multiple metastable conformations. The large number of accessible conformations during protein folding makes it challenging to build computational models that can well sample transition states and capture realistic energy landscapes. Nevertheless, computer simulation methods based on the energy landscape have proven useful in characterizing the folding process in quantitative details and have successfully filled up the gap between theory and experiment.[3−5] In the past decade, protein folding has been studied in depth using various approaches, from minimalist to highly complex models,[6] to answer general questions in molecular and computational biology. Among several developed techniques, computer simulation with a coarse-grained model appears as an outstanding class to study the protein folding process,[7−9] protein−protein interaction,[10] thermodynamics,[11] kinetics,[12] and free energy landscape.[13] Each folding mechanism is associated with its protein native structure, which explains why structure-based models (SBM) describe well general folding properties.[14−18] These models have shown that folding is significantly influenced by different aspects, such as pH, salt concentration, and amino acid substitution. Different approaches have been implemented in it to try to solve specific problems.[19−21] One of these puzzles that remains unclear includes three domains from α-spectrin.[22−28] The R15 domain folds three orders of magnitude faster than its homologues R16 and R17, even though these three domains have similar structures and stabilities.[26] Their folding time difference is associated with the landscape roughness, which is considered the main source for the slower folding and unfolding of R16 and R17 domains.[12,27]

The folding time differences in the R15, R16, and R17 α-spectrin domains have been extensively studied by experimental and computational groups.[12,29] Recent studies have shown that internal collision within the protein, which can be interpreted as an energy dissipation mechanism, does not

Received: August 29, 2022
Revised: December 9, 2022
Published: February 1, 2023

https://doi.org/10.1021/acs.jpcb.2c06178
J. Phys. Chem. B 2023, 127, 1291−1300

contribute to its folding process.[30] In addition, the folding time differences may be associated with different aspects, such as the non-native salt bridges between the helices,[31] the formation of hydrogen bonds,[32] the dispersion interactions,[33] and the dihedral barrier crossing events.[34] In our previous study,[12] we explored the folding time difference by addressing molecular dynamics simulations with the addition of non-native potentials, electrostatic and hydrophobic, into the structure-based Cα model (SBM-Cα). Based on these results, we had reasonable evidence that the folding time differences emerge from non-native cooperative effects given by the hydrophobic (HP^{MJ}) and electrostatic (Elec) interactions.

A computational model was proposed in our previous study,[12] which combined different aspects based on models discussed in the literature.[29,35−38] Computational folding rates were calculated for the wild-type proteins and for the specific mutations that increased the folding rate of the R16 and R17 domains. Their differences displayed reasonable correlations with experimental results. These observations were results of specific amino acid substitutions, E14F, E15D, I18L, K21V, and V25L (also denominated by M_5), that play an essential role in reducing the roughness of the energy landscape profile. These mutations allow an increase in the folding rates of R16 and R17, making them as high as those of R15. Furthermore, for these specific mutations, the experimental data showed that the folding mechanism of the R16 and R17 domains became like R15, from a framework-like to a nucleation condensation mechanism.[39] In this study, to investigate R15, R16, and $R16M_5$ folding routes, we use the energy landscape visualization method (ELViM) technique[40,41] to identify the metastable states and the collective motions responsible for mediating the transition between unfolded and transition states that is responsible for changing the folding mechanism.

Here, we use a coarse-grained model with the structure-based Cα potential (SBM-Cα)[42,43] with the addition of non-native interactions[8,12,36,38] to study the folding process of R15, R16, and $R16M_5$ domains. Molecular dynamics simulations uncovered the different folding pathways among wild-type proteins and proteins with specific mutations. While the folding pathway connects the unfolded to folded ensembles, the representative pathways which go through metastable conformations shed light on the correlations among folding rate, mutation effects, and intermediate states. The energy landscape computed by the ELViM technique identified the metastable conformation responsible for speeding down the folding, which corresponds to the similar configuration reported by the Clarke and Schuler group.[28] Meanwhile, the frustrating search for the correct docking of partly preformed helical was identified as the primary source of this undefined behavior into folding rates. Our findings with R15, R16, and $R16M_5$ α-spectrin domains are consistent with experimental data. Our observations provide new insights into the folding mechanism changes by some amino acid substitutions and the role of these orchestrated sets of non-native interactions that may also be associated with secondary structure misdocking. Thus, this paper presents a multidimensional projection into lower dimensions based on the technique developed by de Oliveira et al.,[40,41] in which the R15, R16, and $R16M_5$ folding pathways are studied in detail. As observed in the previous study, the R16 domain shows substantial differences in the folding route compared to the R15 and R17 ones. So, we focus the present work on the analysis of the R15WT, R16WT, and $R16M_5$ domains.

METHODS

Structure-Based Cα Model with Non-native Potentials. Molecular dynamics simulations were performed using the structure-based Cα model (SBM-Cα).[15,42] In the current SBM-Cα, the amino acids are represented by a single bead of unit mass in the Cα position.[8,44] Harmonic potentials maintain bond lengths, bond angles, and planar dihedrals, which are represented, respectively, by the first, second, and third terms of eq 1. The nonbonded interactions between residues i and j are given by the fourth term, which is represented by Lennard-Jones 10−12 potential. They describe contacts that are present in the native structure and are identified by the Shadow Map Algorithm.[45] All the pairwise residues, which are not in contact in the native structure interact via nonspecific repulsion, as can be seen by the fifth term in eq 1:

$$
\begin{aligned}
V(\Gamma, \Gamma_o) = & \sum_{bonds} \epsilon_r (r - r_o)^2 + \sum_{angles} \epsilon_\theta (\theta - \theta_o)^2 \\
& + \sum_{dihedrals} \epsilon_\phi \Big\{ [1 - \cos(\phi - \phi_o)] \\
& + \frac{1}{2}[1 - \cos(3(\phi - \phi_o))] \Big\} + \sum_{contacts} \epsilon_C \left[5 \left(\frac{d_{ij}}{r_{ij}} \right)^{12} \right. \\
& \left. - 6 \left(\frac{d_{ij}}{r_{ij}} \right)^{10} \right] + \sum_{non\text{-}contacts} \epsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12} \\
& + \sum_{i}^{M} \sum_{j=i+4}^{M} K_{HP} \kappa_{ij} \exp[-(r_{ij}^{HP} - \sigma)^2] \\
& + \sum_{Elec} K_{Elec} \frac{q_i q_j \exp(-\kappa r_{ij}^{Elec})}{\epsilon_K r_{ij}^{Elec}}
\end{aligned}
\tag{1}
$$

in which the distance between two subsequent residues and the angles formed by three and four subsequent residues are represented by r, θ, and ϕ. The parameters ϵ_r, ϵ_θ, and ϵ_ϕ correspond to the strength of bonds, angles, and dihedrals, with $\epsilon_r = 100\epsilon_C$, $\epsilon_\theta = 20\epsilon_C$, and $\epsilon_\phi = \epsilon_{NC} = \epsilon_C = 1$. The constants r_o, θ_o, ϕ_o, and d_{ij} are taken from the native structure; r_{ij} corresponds to the distance between two native contacts from each simulated step, and $\sigma_{NC} = 4$ Å represents the excluded volume term.

The last two terms represent the non-native potential that take into account the hydrophobic (HP^{MJ}) and the electrostatic (Elec) interactions, respectively. The hydrophobic non-native interaction is given by a Gaussian potential,[7,35,46] where the contact energy, κ_{ij}, is taken from the upper triangle in Table 5 of Miyazawa and Jernigan.[47] M is the number of hydrophobic amino acids; $K_{HP} = 0.1$ is the overall strength of the hydrophobic forces, and r_{ij}^{HP} represents the distance between two non-native hydrophobic amino acids (Ala, Val, Leu, Ile, Met, Trp, and Phe), with $\sigma = 5$ Å, which is the minimum interaction distance. The last term, the Debey-Hückel model[48−51] represents the electrostatic interactions by adding charged points at beads in all acidic/basic residues (i.e., Arg, His, Lys, Glu, and Asp). The charged residues, i and j, are represented by q_i and q_j; $K_{Elec} = 332$ kcalÅ/(mol e^2). The dielectric constant (ϵ_K) is 80; r_{ij}^{Elec} is the distance between two charged residues, and κ is the inverse of Debye length.[52]

https://doi.org/10.1021/acs.jpcb.2c06178
J. Phys. Chem. B 2023, 127, 1291−1300

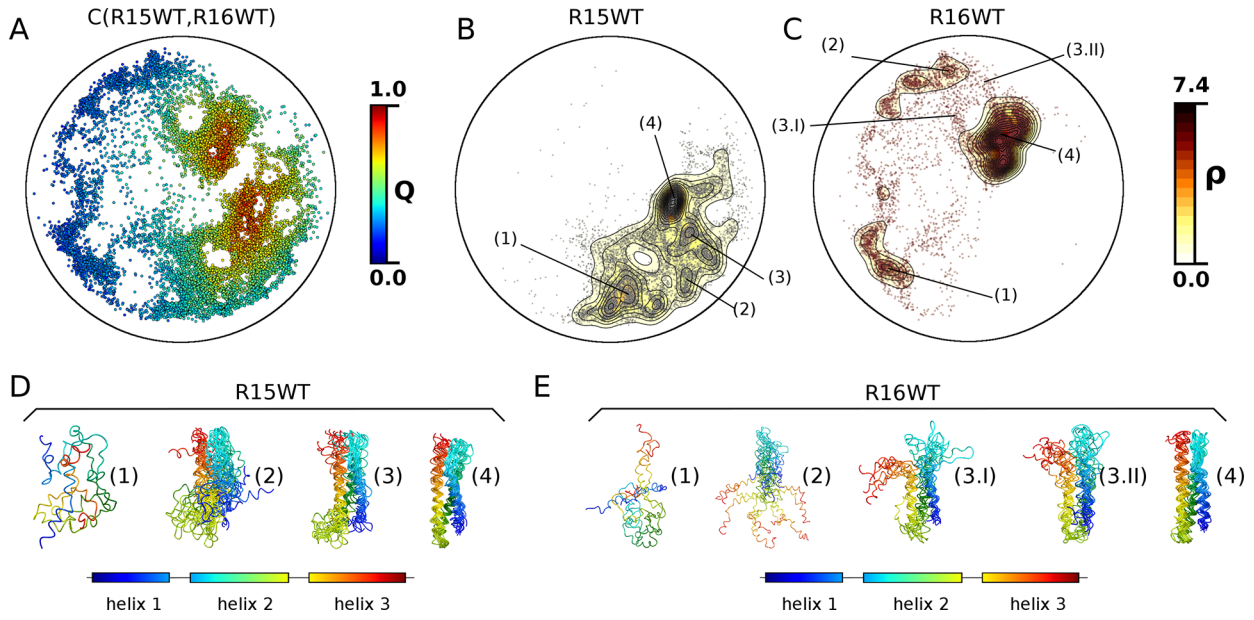


Figure 1. Folding trajectory analysis of R15WT and R16WT. (A) Folding trajectory visualization for C(R15WT,R16WT) mapped into 2D phase space using ELViM. The color bars are a function of the fraction of native contacts, Q. (B, C) 2D density map to R15WT and R16WT, respectively. The color bars correspond to the density of states, ρ, calculated by the package seaborn in python version 3. In (A–C), the euclidean distance between two conformations (dots) corresponds to the structural difference. (D, E) Ensemble of structures taken from (B) and (C), respectively, represent each stage of the folding process.

Non-native Contact Map. In our model, all the hydrophobic and charged residues, i and j, are considered to be in contact if $|i - j| > 3$ with $r^{HP} < 8.0$ Å and $r^{Elec} < 16.0$ Å, respectively. According to eq 2, the non-native contacts group, C, is obtained by the difference between two lists, hydrophobic/electrostatic (A) and native (B) contacts, respectively, with the x_{ij} contact present only in the A group.

$$C = A - B = \{x_{ij} | x_{ij} \in A \text{ and } x_{ij} \notin B\} \tag{2}$$

Simulation Details. The SBM-Cα+HP^{MJ}+Elec simulations were performed using the molecular dynamics package Gromacs version 4.5.5.[53] To maintain coupling to an external bath, the Langevin thermostat algorithm was employed with a constant equal to 1 ps and using a leapfrog integration. All the input files were obtained with the SMOG-Server webtool.[45] Proteins were initialized in an open random configuration and simulated over 5×10^9 steps with timesteps equal to 0.5 fs, and the configurations were saved every 5×10^3 steps. The reaction coordinate used to follow the folding events is defined as the fraction of native contacts (Q). The native contact between two residues, i and j, is considered to have been formed when the distance between them is shorter than $1.2d_{ij}$, where d_{ij} is the distance between two Cα residues in the native structure, with $j > i + 3$. The contact map was determined by the software Shadow Contact Map.[45] R15 and R16 correspond to the positions from 1665 to 1771 and from 1772 to 1878, respectively, of the full-length PDB ID: 1Q4U,[54] and the mutations were generated using Modeller software version 9.17.[55]

Energy Landscape Visualization Method. The visual energy landscape representation is obtained by projecting the structural data set into a 2D effective phase space.[40,41] This method is called ELViM (Energy Landscape Visualization Method) and is based primarily on computing a matrix of conformational distances between all pairs of structures. For two conformations k and l, their similarity measurement is given by

$$q^{k,l} = \frac{1}{N} \sum_{N} \exp \frac{-(r_{i,j}^{k} - r_{i,j}^{l})^2}{\sigma_{i,j}^{\ 2}} \tag{3}$$

where $r_{i,j}^{k}(l)$ is the distance between the atom i and j of the conformation $k(l)$. N is the normalization over all non-neighboring pairs of residues and $\sigma_{i,j} = \sigma_0 |i - j|^{\epsilon}$, with $\sigma_0 = 1$ Å and $\epsilon = 0.15$.[56] The dissimilarity between any two structures is described as

$$\delta^{k,l} = 1 - q^{k,l} \tag{4}$$

where $\delta^{k,l}$ ranges from 0 (identical) to 1 (different).

The landscape description is obtained through a 2D representation of the dissimilarity matrix ($\delta^{k,l}$) which aims to describe the computed effective distances optimally. The method goes beyond the usual one-dimensional representation, and it does not require a reference conformation or a reaction coordinate. This approach relies only on structural information, but the energies of states are implicitly accounted for since they are strongly correlated with their molecular conformations.

The general workflow for this method is based on four steps: (i) an ensemble of structures that, in general, are obtained through simulation is generated; (ii) the dissimilarity matrix is calculated by applying a metric throughout the simulated trajectory; (iii) a data processing procedure may be carried out to cluster very similar structures into a single conformation, lowering the effective total number of conformations; (iv) a multidimensional projection is performed and, then, the dissimilarity matrix is transformed to a two-dimensional (2D) projection.

There are many methods for the analysis of complex multidimensional data which rely on pairwise distances between elements.[57] In this method, we address this question

https://doi.org/10.1021/acs.jpcb.2c06178
J. Phys. Chem. B 2023, 127, 1291−1300

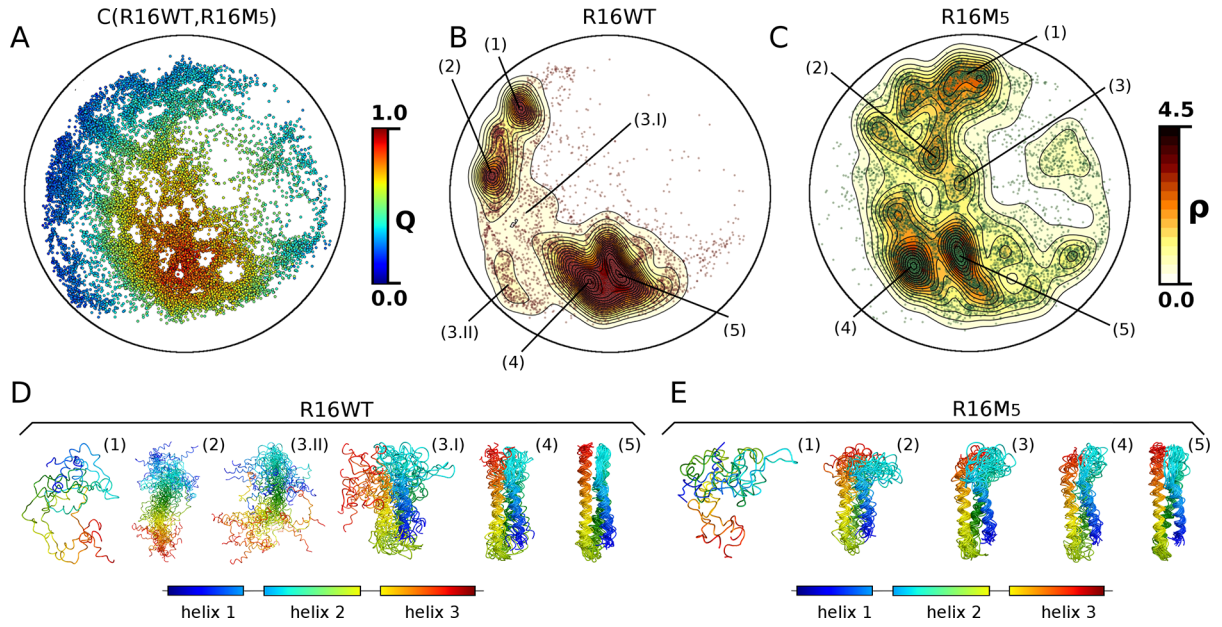


Figure 2. Folding trajectory analysis of R16WT and R16M$_5$. (A) Folding trajectory visualization for C(R16WT,R16M$_5$) mapped into 2D phase space using ELViM. The color bar is a function of fraction of native contacts, Q. (B, C) 2D density map for R16WT and R16M$_5$, respectively. The color bar corresponds to the density of states, ρ, calculated by the package seaborn in python version 3. In (A−C), the euclidean distance between two conformations (dots) corresponds to the structural difference. (D,E) Ensemble of structures taken from (B) and (C), respectively, that represent each stage of the folding process.

by mapping the data into 2-dimensional spaces, preserving as much information about dissimilarity between the data as possible.[58] Transition states are usually identified through peaks in the free energy profile, which are obtained as a function of reaction coordinates, which are defined *a priori* (e.g., fraction of native contacts) or identified using computational methods.[59] In 1D cases, this procedure has the potential to shroud the richness of the energy landscape and its dynamics (e.g., in case there are multiple paths). In a 2D representation, instead of using the peak in the free energy profile, transition states are identified as the saddle points in the surface. ELViM does not depend on reference structures or any other *a priori* variable to be used as reaction coordinates, and the energy landscape structure is an emerging property of the conformation data set. However, the determination of the transition state ensembles is analogous to the 2D analysis using reaction coordinates. Using the density of states obtained from 2D ELViM projection, the transition state ensembles are identified as the conformations at the saddle points.

In brief, ELViM uses only structural analysis over time-uncorrelated structures. The proximity of the structures in the projection conveys the similarity between them. The assumption is that it is possible to have a good representation of the configurational phase space with a reasonable number of conformations (usually around 10k to 50k states). These sampled states convey the underlying structure of the energy landscape, in which the kinetic information is implicitly included since the data set is obtained with realistic molecular dynamics simulations. From the ELViM representation, one can infer the stable ensembles and transition state ensembles associated with interconversion between them.

RESULTS AND DISCUSSION

All the simulations of the wild-type proteins, R15WT and R16WT, and R16M$_5$ were herein carried out with the SBM-Cα model with the non-native potentials, electrostatic (Elec) and hydrophobic (HP^{MJ}). The details about the simulation results to SBM-Cα+HP^{MJ} and SBM-Cα+Elec are in our previous study.[12] Such work evidenced that the folding rate differences between R15WT and R16WT from α-spectrin emerge from the combination of both hydrophobic and electrostatic contributions. Moreover, the set of mutations inserting R15WT residues to create R16M$_5$ reduces the frustration associated with the search for the correct docking of the

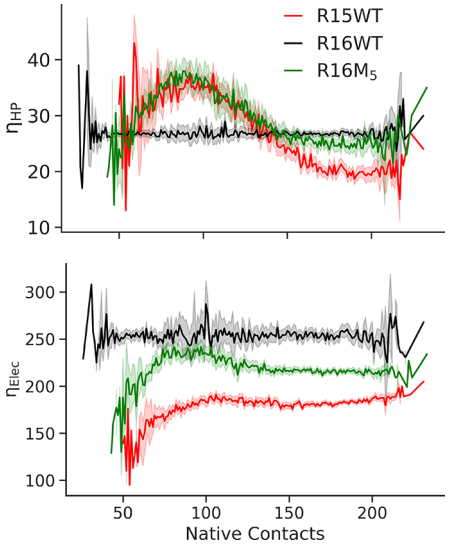


Figure 3. R15WT, R16WT, and R16M$_5$ non-native contacts. Average number of non-native hydrophobic, η_{HP}, and charged, η_{Elec}, contacts as a function of native contacts. The domains R15WT, R16WT, and the mutant R16M$_5$ are in red, black, and green solid lines, respectively. The shadow in each line corresponds to the fluctuation. The cutoff distances to charged and hydrophobic non-native contact are r^{elec} < 16.0 Å and r^{HP} < 8.0 Å, respectively.

https://doi.org/10.1021/acs.jpcb.2c06178
J. Phys. Chem. B 2023, 127, 1291−1300

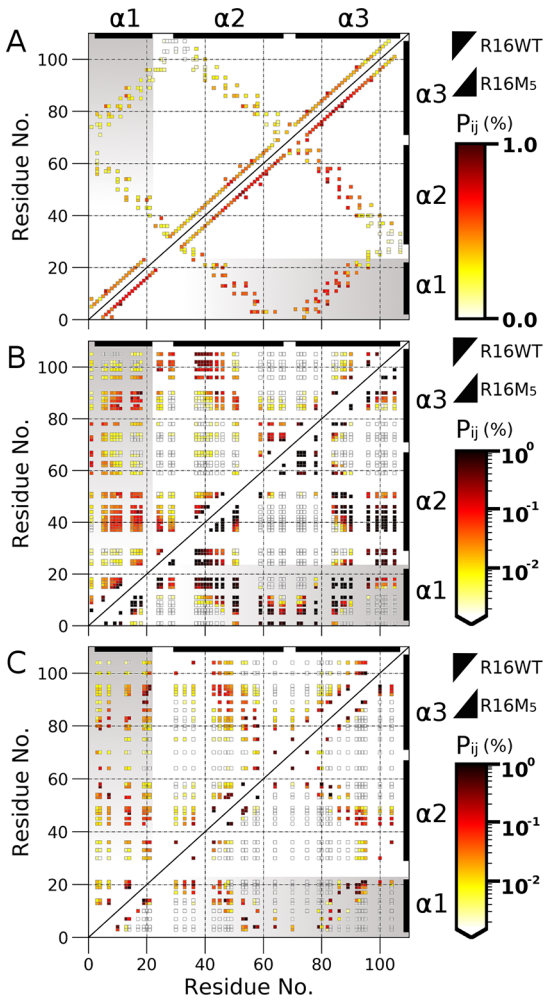


Figure 4. Native and non-native contact probability formation for R16WT and R16M$_5$. (A−C) Native and the non-native charged and hydrophobic contact maps, respectively, for R16WT (upper triangle) and R16M$_5$ (lower triangle). The contacts i and j are colored by their contact formation probability, P_{ij} (%), of the ensemble of structures selected in Figure S2. The native contacts i and j are determined by the Shadow Map Algorithm, and the non-native charged and hydrophobic contacts are defined as those within a cutoff distance r^{Elec} < 16.0 Å and r^{HP} < 8.0 Å, respectively. Black rectangles at the top and right side of each graph correspond to the R16WT and R16M$_5$ α-helix 1, 2, and 3 (α_1, α_2, and α_3). Contacts of α-helix 1 with α-helices 2 and 3 are shown on the gray background.

helices, subsequently decreasing the landscape roughness and changing the folding route, which was also reported by experiments.[27]

Although previous works were able to shed light on the effect of the non-native interactions in the α-spectrin folding speed, the details about the molecular mechanism of its folding process remains unclear. Thus, to investigate the folding route difference between R15WT, R16WT, and R16M$_5$, the energy landscape of these proteins was calculated by combining pairs of trajectories: R15WT and R16WT called C(R15WT,R16WT) and R16WT and R16M$_5$ called C(R16WT,R16M$_5$). ELViM allows multiple data sets to be analyzed using the same configuration phase space. Of course, these proteins have to be of the same sizes. As in the present study, the proteins have essentially the same native structures, and this procedure allows a better comparison between the two groups, as we have shown in other studies.[13,60] The simulation trajectories of all domains were taken at their folding temperature. For ELViM analysis, each trajectory consisted of 10^4 configurations.

R15WT Presents a High Folding Route Specificity Compared to R16WT. The 2D phase space visualizations given by ELViM using two trajectories combinations, C(R15WT,R16WT) and C(R16WT,R16M$_5$), are shown in Figures 1A and 2A, respectively. In the scatter plot design visualization, each dot represents a structure conformation. The fraction of native contacts, Q, of each protein was used to color each conformation. In Figure 1A, two different folding routes were observed using the R15WT and R16WT folding trajectories. With 2D effective phase space, one can start analyzing each ensemble data separately, as shown in Figure 1B,C. The protein's density map displayed different folding pathways that each domain takes on. In the R15WT domain, with the 2D phase space and the density map, four main clusters were identified, and the ensemble of representative structures is shown in Figure 1D (see Figure S1 in the Supporting Information).

According to the experimental results,[39] R15WT folds via a nucleation−condensation mechanism with only the centers of the helices 1 and 2 structured and docked. The ensemble of structures taken from our computational results revealed that the folding mechanism of the R15WT domain follows the same line reported in the literature.[23] In Figure 1D, the unfolded ensemble (1) and the representative structures in the transition state (2) are in the concurrent formation of secondary and tertiary structure interactions. In the early native state (3), the correct arrangement of helices is formed, and a remarkable coil region fluctuation with the N and C terminals of the protein is observed. Finally, the R15WT domain reaches its native state (4).

On the other hand, R16WT, which folds three orders magnitude slower than the R15WT domain, folds via a framework-like mechanism that involves the docking of partly preformed helices 1 and 3 along the entire length.[22] Also, the framework-like mechanism is regularly associated with a faster fold, but this is not the case for the R16WT domain. In Figure 1C, which shows the R16WT density map, it is observed that the trajectory spread over a 2D phase space completely different from R15WT (see Figure 1B). More than that, the routes that lead the R16WT unfolded state to its native structure emerge from different folding trajectories. These molecular paths present different unfolded ensembles, for example, ensembles (1) and (2), and the transition state can be represented by two routes, as indicated in (3.I) and (3.II). Unlike route (3.II), in (3.I), helix 3 is not located in its native position and the coil region between helix 1 and 2 is highly frustrated. At last, the ensemble (4) represents the native state of the R16WT domain.

The folding mechanism and the route of these two domains are significantly different, resulting in the faster folding of R15WT than R16WT. In general, the folding rate difference between R15WT and R16WT emerges from cooperative effects of non-native interactions, as shown in Bruno da Silva et al.[12] In Figure 1B,C, as a result of these cooperative effects, the unlikeness in their energy landscape showed intrinsic aspects of their folding. R16WT folds from the unfolded to the folded state through several metastable ensembles and may assume two different routes in the transition state, one more structured than the other. In this case, one of these routes may be

https://doi.org/10.1021/acs.jpcb.2c06178
J. Phys. Chem. B 2023, 127, 1291−1300

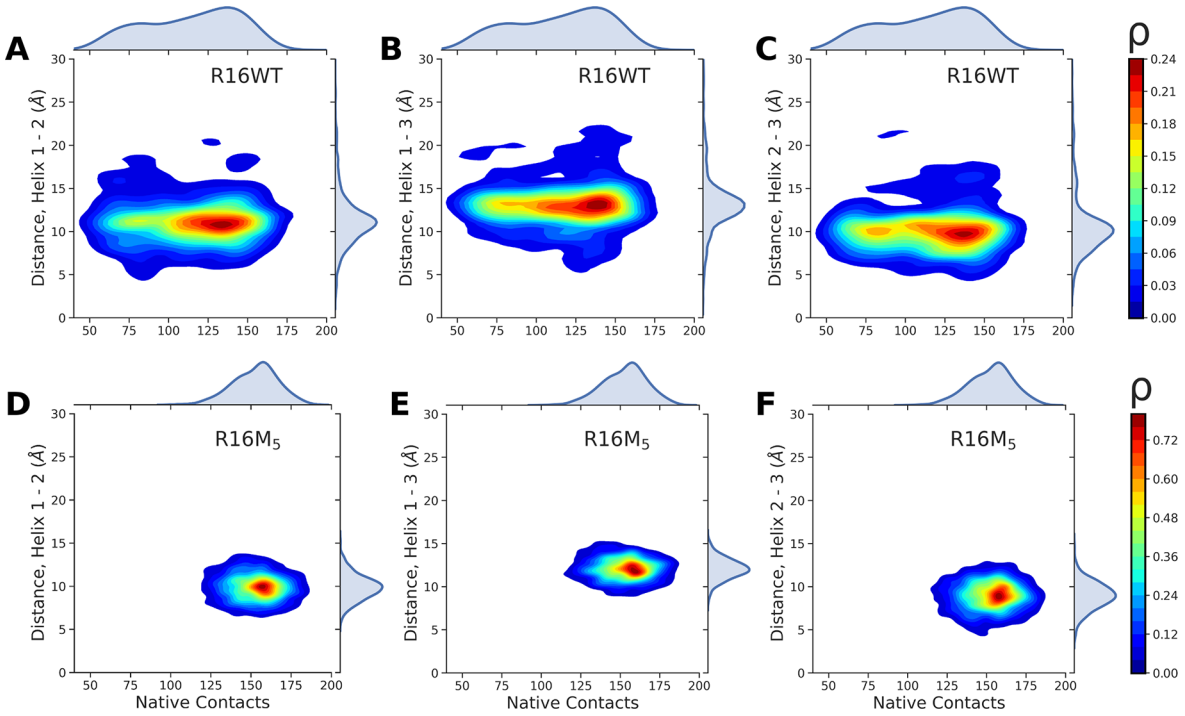


Figure 5. The center of mass distance from the helices. R16WT (A−C) and R16M$_5$ (D−F) display the center of mass distance (Å) between α-helix 1 and 2, 1 and 3, and 2 and 3, respectively, as a function of native contacts with their corresponding distribution (vertical and horizontal). The color bars represent the density of states, ρ, calculated by the package seaborn in python version 3.

interpreted as a trap, making the protein spend more time searching for the correct docking of the helices, as shown in Figure 1C,E. In contrast, R15WT folds from the unfolded to the folded state through a directed route, as seen in Figure 1B,D. The dominant transition state ensembles are exemplified by the structures 1D.2 and 1D.3.

Mutations Optimize the Non-native Interactions and Speed up the Folding of R16WT. As previously mentioned, theoretical and experimental data have suggested that slower folding is a result of a rugged landscape.[8,12,28,36,61] In this section, we have presented evidence that one possible reason for the slow folding of R16WT compared to R15WT is that kinetic traps in the energy landscape prevent its diffusion. For this reason, the protein takes longer to escape from local minima. Some ensembles of these kinetic traps are presented in Figure 1A−C. Experimentally, Clarke's group showed that R16M$_5$ folds and unfolds significantly faster than its wild-type version.[27] Furthermore, they have indicated that the R16M$_5$ folding mechanism is altered compared to R16WT, which includes the correct dock of helix 3 promoted by non-native interaction events.

In Figure 2, the combination of folding trajectories of R16WT and R16M$_5$, C(R16WT, R16M$_5$), in 2D phase space and the density map for individual proteins show crucial differences in their energy landscape by changing some residues. Even though R16WT and R16M$_5$ display a superposition in the 2D phase space (see Figure 2A), the density map of each protein reveals a significant density variation between unfolded and folded states, as seen in Figure 2B,C. The ensembles of structures visited by R16WT are compounded by two unfolded clusters, Figure 2B,D in (1) and (2), where it is possible to observe the secondary structure formation before the interchain interactions. The early native ensemble (3.II) is established in the same region observed in Figure 1C, close to the transition state and follows the same structure formation, ensemble (2) and (3.II) in Figures 1E and 2D, respectively. The transition state structure representation in (3.I) in Figure 2B is approximately the same set of structures in (3.I) in Figure 1C. Such a result demonstrates a representative ensemble of structures that may be part of a trap during protein folding. In both cases, helix 3 and the coil region between helix 1 and 2 are highly frustrated. The folded ensemble is characterized by two clusters, one with few frustrated structures (5) and the other with more fluctuations (4).

Experimental and computational results are in qualitative agreement in the folding time calculation. Moreover, ELViM analysis appropriately accounts for the changes in the R16M$_5$ folding mechanism. In Figure 2C, the R16M$_5$ density map displays a significant increase in the transition state region, indicated by (2) and (3). Following the folding path of R16M$_5$, the folding goes from the unfolded (1) to the folded ensemble, (4) and (5), through highly populated intermediate states, (2) and (3). The changes in the folding mechanism occur by the substitution of five residues, promoting a highly collapsed structure, as seen in Figure 2E (see Figure S2 in the Supporting Information). In summary, the wild-type proteins, R15WT and R16WT, assume different folding mechanisms, nucleation-condensation and framework-like mechanisms, respectively. In the first case, the helices 1 and 3 of R15WT established the correct docking. In contrast, in the second case, helix 3 of R16WT must find the proper dock of its helix in a more roughed energy landscape. In this sense, the helices misfolding is the primary source of frustration in the energy landscape, and the set of mutations to speed up is crucial to reduce the roughness and promote non-native cooperative interactions that speed up the folding process. The reason to combine two folding trajectories is to understand better, visualize, and

https://doi.org/10.1021/acs.jpcb.2c06178
J. Phys. Chem. B 2023, 127, 1291−1300

distinguish the differences between folding routes. In Figures 1 and 2, the 2D phase space map is colored according to the fraction of native contacts and rescaled from 0 to 1. While Figure 1A shows that R15WT and R16WT have two different folding routes, Figure 2A shows R16WT and $R16M_5$ have similar transition state ensembles.

Non-native Contact Density of the R16WT Mutant Mimics That of R15WT. The number of non-native contacts was measured to analyze in which state of the folding process the non-native interactions promote changes in the $R16M_5$.[8,21,36] In Figure 3, the average number of hydrophobic and electrostatic non-native contacts, η_{HP} and η_{Elec}, are shown for R15WT, R16WT, and $R16M_5$ in red, black, and green solid lines, respectively. For the R15WT domain, the same behavior is observed, for hydrophobic and electrostatic non-native contacts, until the protein reaches $Q \sim 100$. The number of non-native contacts increases equally, but for $Q > 100$, η_{HP} decreases until reaching a plateau. Unlike hydrophobic non-native contacts, for $Q > 100$, η_{Elec} remains almost the same with slight fluctuations. The R16WT domain displays no variation in the number of non-native contacts during folding, remaining with the same behavior for both non-native contacts. Such results demonstrate that hydrophobic and electrostatic non-native interactions act with the same weight in the entire folding, with a slight fluctuation around $Q = 100$ and also in its unfolded and folded state.

As discussed in the previous section, the changes by the set of mutations that speed up the folding emerge from the cooperative effects of non-native interactions. $R16M_5$ displays an interesting behavior; the protein presents almost the same pattern as protein R15WT. Moreover, in both cases, hydrophobic and electrostatic, the number of non-native contacts increases up to $Q = 100$; however, for $Q > 150$, η_{HP} presents the same values of R16WT, while R15WT has lower values. On the other hand, η_{Elec} has a considerable increase up to $Q = 100$, and a slight decrease is observed, reaching values between those of R15WT and R16WT.

The folding mechanisms of R15WT and $R16M_5$ are similar, as shown by experimental data[27] and by Figures 1 and 2. With the results in Figure 3, the contribution of non-native interactions displays different roles during their folding, giving some shreds of evidence of how the non-native interactions affect and change the energy landscape. It is worth mentioning that the non-native energy follows the same interpretation and all nonbonded energies in our model are temperature independent and pairwise additive (see Figure S3 in the Supporting Information). To sum up, the modification of the folding mechanism of $R16M_5$ is due to the role that non-native interactions play by the insertion of a set of mutations.

Formations of Specific Contacts Promoted by Non-native Interactions Speed up the Folding. Structural properties of our simulated R15WT, R16WT, and $R16M_5$ protein folding are in agreement with the experimental results, indicating that non-native interactions play a crucial role in their folding mechanism. Conformational changes that speed up the folding were investigated by selecting the structures of R16WT and $R16M_5$ from the transition state regions (see Figure S4 in the Supporting Information). In Figure 4A, the native contact map presents a slight difference in the contact probability formation (P_{ij}) in helices 1, 2, and 3 ($\alpha1$, $\alpha2$, and $\alpha3$) for proteins R16WT and $R16M_5$. The R16WT secondary and tertiary native structure interactions (upper triangle) display different propensities. In this case, the helix is more structured, while the tertiary structure interactions have a low probability. Meanwhile, the $R16M_5$ native contact map (lower triangle) displayed an ensemble of conformations more structured than the R16WT ensemble of structures. Thus, the results in Figure 4A follow the same arguments discussed in Figures 1 and 2, in which the folding energy landscape of $R16M_5$ becomes less frustrated and has no significant traps.

The electrostatic non-native map for R16WT and $R16M_5$, Figure 4B, displays a significant difference in their probabilities. There is a higher density of contact formation of $\alpha1$ with $\alpha2$ and $\alpha3$ for R16WT than for $R16M_5$. However, the few contacts of $R16M_5$ are more specific and with higher probability. The number of non-native contacts helps us to understand the correct docking of the helices. In the case of $R16M_5$, the contacts with greater specificity indicate a correct helix docking, whereas lower specificity and a higher amount of contacts imply the search for the correct dock. In the literature,[26,39] the folding rate difference, the roughness in the energy landscape, and the misdocking were attributed to the frustrating search of helix 3 for its native position. Our results have shown an example of how the noncooperative effects of helix 1 may perturb the docking of helix 3. Based on this assumption, many of the non-native contacts increase the helices fluctuation, leading the protein to a frustrated search of its native state (see the ensembles of structures in Figures 1 and 2). A similar non-native contact formation pattern was noticed in the hydrophobic non-native contacts for R16WT and $R16M_5$, as shown in Figure 4C. In both cases, in the hydrophobic and electrostatic non-native contact map (Figure 4B,C), the helix 1 plays an important role in the kinetic trap for the R16WT folding.

Misdocking between helices yields in distance fluctuations between them. Therefore, the distances of the center of mass between helix 1 and 2, helix 1 and 3, and helix 2 and 3 were calculated to investigate their behavior as these helices approach each other. Figure 5 presents the distance of the center of mass for the ensemble of structures corresponding to the same region discussed above for R16WT and $R16M_5$. The number of native contacts for R16WT ensemble structures ranges from 50 to 175. The distance between helix 1 and 2, with Q around 75, displays a significant fluctuation, ranging from 5 to 20 Å, as shown in Figure 5A. However, while the number of native contacts increases, the fluctuation of the distances between helices 1 and 2 decreases, reaching, at $125 < Q < 150$, approximately its native distance. Nonetheless, we observe an opposite behavior for the distance between helices 1 and 3 and helices 2 and 3, showing a small fluctuation close to $Q = 75$ and significant variation for $Q > 125$ (Figure 5B,C). As seen here, we suggest that the correct docking of helices 1 and 2 promotes the misdocking of helix 3 with helices 1 and 2. The R16WT folds through a framework-like mechanism, in which the formation of the helices precedes their packing. Our results corroborated with the structuring process of helices formation, in which helix 1 packs with helix 2, then helix 3 must find the correct docking, and the kinetic trap emerges from a non-native docking event.

In contrast to the R16WT results, $R16M_5$ displays an ensemble with more structured conformations, $120 < Q < 180$, and according to the contact formation probability (Figure 4), the helices distance display a highly concentrated density to values close to the native distance (Figure 5D–F). The ensemble of structures selected for $R16M_5$ ranges from the early transition state to the late transition state. This region

https://doi.org/10.1021/acs.jpcb.2c06178
J. Phys. Chem. B 2023, 127, 1291−1300

corresponds approximately to the same stage of R16WT. Still, as a result of five mutations, E18F+E19D+I22L+K25V+V29L, the folding mechanism, the roughness in the energy landscape, and the misdocking events have been affected in a manner that speeds up the $R16M_S$ folding process.

CONCLUDING REMARKS

In the present work, we have determined an ensemble of representative conformations for each folding stage of the R15WT, R16WT, and $R16M_S$ by using the energy landscape visualization method (ELViM). The folding trajectories were obtained with the SBM-Cα+HP^{MJ}+Elec potential. The detailed analysis of the features of these structural ensembles has revealed a crucial aspect of the kinetic trapping of R16WT. The helix misfolding is associated as the primary source of frustration in the folding landscape of R16WT. Moreover, non-native docking events are related to these misfolding incidents. The addition of non-native potentials, hydrophobic (HP^{MJ}) and electrostatic (Elec), into the SBM-Cα has shown that the folding time difference is due to the energy landscape roughness, which emerges from the cooperative effects of these non-native interactions.

We have also investigated the set of mutations that speeds up the R16WT folding. The substitution of five residues into R16WT and $R16M_S$ significantly increased the folding speed due to a decrease in the energy landscape roughness. The folding times can not be compared directly with the experimental times, but only the correlation between them, as it was discussed in previous work in which the results for several mutations calculated both by experiments and simulations were shown. The smoother energy landscape of this mutant is a result of the reduced frustrations of helix 3, which causes a number of non-native docking events, and changes the folding mechanism from framework-like to nucleation-like. Also, the non-native contact map has shown the cooperative effects of the tertiary structure interactions between the helices through the increase in the likelihood of non-native contacts. In conclusion, depending on the amount of frustration, the non-native interactions may increase or decrease the folding speed by reducing the energy landscape roughness.[36,43,62,63] Finally, the ELViM provided a uniquely detailed description of the different ensembles during the protein folding, showing the main stage that represents the kinetic trap, which in turn leads the R16WT protein to seek the correct docking of its helix.

The difference in folding rates of α-spectrin domains has been a long-standing puzzling problem. The analysis of non-native interactions, along with the ELViM approach, has provided valuable insights into their folding mechanism, and due to its generality, this methodology is feasible to apply to other complex folding systems.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.2c06178.

Figure S1: R15WT and R16WT ensemble of structures taken from Figure 1B,C; Figure S2: R16WT and $R16M_S$ ensemble of structures taken from Figure 2B,C; Figure S3: 2D density map to R16WT and $R16M_S$; Figure S4: average non-native hydrophobic, E_{HP}, and electrostatic, E_{Elec}, energies as a function of native contacts (PDF)

AUTHOR INFORMATION

Corresponding Author

Vitor B. P. Leite − *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;* orcid.org/0000-0003-0008-9079; Email: vitor.leite@unesp.br

Authors

Fernando Bruno da Silva − *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;* orcid.org/0000-0002-0285-8700

Vinícius Martins de Oliveira − *Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, Maryland 21201, United States;* orcid.org/0000-0003-0927-3825

Antonio Bento de Oliveira Junior − *Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005-1892, United States*

Vinícius de Godoi Contessoto − *Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005-1892, United States;* orcid.org/0000-0002-1891-9563

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.2c06178

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We want to thank José Nelson Onuchic for sharing his wisdom, for his enthusiastic and insightful discussions, for being an inspiration to do science, and for his friendship. F.B.S. was supported by the National Council for Scientific and Technological Development (CNPq - Grant 141715/2017-0). V.G.C. was funded by FAPESP (São Paulo State Research Foundation) and CAPES (Higher Education Personnel Improvement Coordination) Grants 2016/13998-8 and 2017/09662-7. V.M.O. was supported by FAPESP (Grant 2018/11614-3). V.B.P.L. was supported by CNPq and FAPESP (Grant 2019/22540-3).

REFERENCES

(1) Bryngelson, J. D.; Wolynes, P. G. Spin Glasses and the Statistical Mechanics of Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 7524−7528.

(2) Leopold, P. E.; Montal, M.; Onuchic, J. N. Protein Folding Funnels: A Kinetic Approach to the Sequence-Structure Relationship. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 8721−8725.

(3) Kubelka, J.; Hofrichter, J.; Eaton, W. A. The Protein Folding 'Speed Limit'. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76−88.

(4) 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., Bioinf.* **1995**, *21*, 167−195.

(5) Fersht, A. R.; Matouschek, A.; Serrano, L. The Folding of an Enzyme. *J. Mol. Biol.* **1992**, *224*, 771−782.

(6) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116*, 7898−7936.

(7) Chen, T.; Chan, H. S. Native Contact Density and Nonnative Hydrophobic Effects in the Folding of Bacterial Immunity Proteins. *PLoS Comput. Biol.* **2015**, *11*, e1004260.

(8) 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.
(9) Martins de Oliveira, V.; Godoi Contessoto, V. d.; Bruno da Silva, F.; Zago Caetano, D. L.; Jurado de Carvalho, S.; Pereira Leite, V. B. Effects of pH and Salt Concentration on Stability of a Protein G Variant Using Coarse-Grained Models. *Biophys. J.* **2018**, *114*, 65−75.
(10) Basdevant, N.; Borgis, D.; Ha-Duong, T. A Coarse-Grained Protein-Protein Potential Derived from an All-Atom Force Field. *J. Phys. Chem. B* **2007**, *111*, 9390−9399.
(11) de Oliveira, V. M.; Caetano, D. L. Z.; da Silva, F. B.; Mouro, P. R.; de Oliveira, A. B.; de Carvalho, S. J.; Leite, V. B. P. pH and Charged Mutations Modulate Cold Shock Protein Folding and Stability: A Constant pH Monte Carlo Study. *J. Chem. Theory Comput.* **2020**, *16*, 765−772.
(12) 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.
(13) Oliveira Junior, A. B.; Lin, X.; Kulkarni, P.; Onuchic, J. N.; Roy, S.; Leite, V. B. P. Exploring Energy Landscapes of Intrinsically Disordered Proteins: Insights into Functional Mechanisms. *J. Chem. Theory Comput.* **2021**, *17*, 3178−3187.
(14) 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.
(15) de Oliveira, A. B., Jr.; Contessoto, V. G.; Hassan, A.; Byju, S.; Wang, A.; Wang, Y.; Dodero-Rojas, E.; Mohanty, U.; Noel, J. K.; Onuchic, J. N.; Whitford, P. C. SMOG 2 and OpenSMOG: Extending the limits of structure-based models. *Protein Sci.* **2022**, *31*, 158−172.
(16) Freitas, F. C.; Lima, A. N.; Contessoto, V. d. G.; Whitford, P. C.; de Oliveira, R. J. Drift-Diffusion (DrDiff) Framework Determines Kinetics and Thermodynamics of Two-State Folding Trajectory and Tunes Diffusion Models. *J. Chem. Phys.* **2019**, *151*, 114106.
(17) 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*, e1004794.
(18) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. Y.; Onuchic, J. N. An All-Atom Structure-Based Potential for Proteins: Bridging Minimal Models with All-Atom Empirical Forcefields. *Proteins: Struct., Funct., Bioinf.* **2009**, *75*, 430−441.
(19) 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.
(20) Martins de Oliveira, V.; de Godoi Contessoto, V.; Bruno da Silva, F.; Zago Caetano, D. L.; Jurado de Carvalho, S.; Pereira Leite, V. B. Effects of pH and Salt Concentration on Stability of a Protein G Variant Using Coarse-Grained Models. *Biophys. J.* **2018**, *114*, 65−75.
(21) Ngo, K.; Bruno da Silva, F.; Leite, V. B. P.; Contessoto, V. G.; Onuchic, J. N. Improving the Thermostability of Xylanase A from Bacillus subtilis by Combining Bioinformatics and Electrostatic Interactions Optimization. *J. Phys. Chem. B* **2021**, *125*, 4359−4367.
(22) Scott, K. A.; Randles, L. G.; Clarke, J. The Folding of Spectrin Domains II: Phi-value Analysis of R16. *J. Mol. Biol.* **2004**, *344*, 207−221.
(23) Scott, K. A.; Batey, S.; Hooton, K. A.; Clarke, J. The Folding of Spectrin Domains I: Wild-type Domains Have the Same Stability but very Different Kinetic Properties. *J. Mol. Biol.* **2004**, *344*, 195−205.
(24) Scott, K. A.; Clarke, J. Spectrin R16: Broad Energy Barrier or Sequential Transition States? *Protein Sci.* **2005**, *14*, 1617−1629.
(25) Scott, K. A.; Randles, L. G.; Moran, S. J.; Daggett, V.; Clarke, J. The Folding Pathway of Spectrin R17 from Experiment and Simulation: Using Experimentally Validated MD Simulations to Characterize States Hinted at by Experiment. *J. Mol. Biol.* **2006**, *359*, 159−173.
(26) Wensley, B. G.; Batey, S.; Bone, F. A. C.; Chan, Z. M.; Tumelty, N. R.; Steward, A.; Kwa, L. G.; Borgia, A.; Clarke, J. Experimental Evidence for a Frustrated Energy Landscape in a Three-Helix-Bundle Protein Family. *Nature* **2010**, *463*, 685−688.
(27) Wensley, B. G.; Kwa, L. G.; Shammas, S. L.; Rogers, J. M.; Browning, S.; Yang, Z.; Clarke, J. Separating the Effects of Internal Friction and Transition State Energy to Explain the Slow, Frustrated Folding of Spectrin Domains. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17795−17799.
(28) Borgia, A.; Wensley, B. G.; Soranno, A.; Nettels, D.; Borgia, M. B.; Hoffmann, A.; Pfeil, S. H.; Lipman, E. A.; Clarke, J.; Schuler, B. Localizing Internal Friction Along the Reaction Coordinate of Protein Folding by Combining Ensemble and Single-Molecule Fluorescence Spectroscopy. *Nat. Commun.* **2012**, *3*, 1195.
(29) Best, R. B. How Well Does a Funneled Energy Landscape Capture the Folding Mechanism of Spectrin Domains? *J. Phys. Chem. B* **2013**, *117*, 13235−13244.
(30) Hagen, S. J.; Qiu, L.; Pabit, S. A. Diffusional Limits to the Speed of Protein Folding: Fact or Friction? *J. Phys.: Condens. Matter* **2005**, *17*, S1503.
(31) Chung, H. S.; Piana-Agostinetti, S.; Shaw, D. E.; Eaton, W. A. Structural Origin of Slow Diffusion in Protein Folding. *Science* **2015**, *349*, 1504−1510.
(32) Schulz, J. C. F.; Miettinen, M. S.; Netz, R. R. Unfolding and Folding Internal Friction of β-Hairpins Is Smaller than That of α-Helices. *J. Phys. Chem. B* **2015**, *119*, 4565−4574.
(33) Sashi, P.; Ramakrishna, D.; Bhuyan, A. K. Dispersion Forces and the Molecular Origin of Internal Friction in Protein. *Biochemistry* **2016**, *55*, 4595−4602.
(34) Echeverria, I.; Makarov, D. E.; Papoian, G. A. Concerted Dihedral Rotations Give Rise to Internal Friction in Unfolded Proteins. *J. Am. Chem. Soc.* **2014**, *136*, 8708−8713.
(35) Zarrine-Afsar, A.; Wallin, S.; Neculai, A. M.; Neudecker, P.; Howell, P. L.; Davidson, A. R.; Chan, H. S. Theoretical and Experimental Demonstration of the Importance of Specific Nonnative Interactions in Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 9999−10004.
(36) 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., Bioinf.* **2013**, *81*, 1727−1737.
(37) Clementi, C.; Plotkin, S. S. The Effects of Nonnative Interactions on Protein Folding Rates: Theory and Simulation. *Protein Sci.* **2004**, *13*, 1750−1766.
(38) Chen, T.; Song, J.; Chan, H. S. Theoretical Perspectives on Nonnative Interactions and Intrinsic Disorder in Protein Folding and Binding. *Curr. Opin. Struct. Biol.* **2015**, *30*, 32−42.
(39) Kwa, L. G.; Wensley, B. G.; Alexander, C. G.; Browning, S. J.; Lichman, B. R.; Clarke, J. The Folding of a Family of Three-Helix Bundle Proteins: Spectrin R15 Has a Robust Folding Nucleus, Unlike Its Homologous Neighbours. *J. Mol. Biol.* **2014**, *426*, 1600−1610.
(40) de Oliveira, A. B.; Yang, H.; Whitford, P. C.; Leite, V. B. P. Distinguishing Biomolecular Pathways and Metastable States. *J. Chem. Theory Comput.* **2019**, *15*, 6482−6490.
(41) de Oliveira, A. B., Jr; Fatore, F. M.; Paulovich, F. V.; Oliveira, O. N., Jr; Leite, V. B. P. Visualization of Protein Folding Funnels in Lattice Models. *PLoS One* **2014**, *9*, e100861.
(42) 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.
(43) de Mendonça, M. R.; Rizzi, L. G.; Contessoto, V.; Leite, V. B. P.; Alves, N. A. Inferring a Weighted Elastic Network from Partial Unfolding with Coarse-Grained Simulations. *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 119−129.
(44) 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., Bioinf.* **2013**, *81*, 1727−1737.
(45) 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.
(46) Zhang, Z.; Chan, H. S. Competition Between Native Topology and Nonnative Interactions in Simple and Complex Folding Kinetics of Natural and Designed Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 2920−2925.
(47) Miyazawa, S.; Jernigan, R. L. Self-Consistent Estimation of Inter-Residue Protein Contact Energies Based on an Equilibrium Mixture Approximation of Residues. *Proteins: Struct., Funct., Bioinf.* **1999**, *34*, 49−68.
(48) 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.
(49) 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.
(50) 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.
(51) de Oliveira, V. M.; Dias, M. M. G.; Avelino, T. M.; Videira, N. B.; da Silva, F. B.; Doratioto, T. R.; Whitford, P. C.; Leite, V. B. P.; Figueira, A. C. M. pH and the Breast Cancer Recurrent Mutation D538G Affect the Process of Activation of Estrogen Receptor α. *Biochemistry* **2022**, *61*, 455−463.
(52) Ullner, M.; Woodward, C. E.; Jonsson, B. A Debye−Hückel Theory for Electrostatic Interactions in Proteins. *J. Chem. Phys.* **1996**, *105*, 2056−2065.
(53) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(54) Kusunoki, H.; Minasov, G.; MacDonald, R. I.; Mondragón, A. Independent Movement, Dimerization and Stability of Tandem Repeats of Chicken Brain α-Spectrin. *J. Mol. Biol.* **2004**, *344*, 495−511.
(55) Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinf.* **2016**, *54*, 5.6.1−5.6.37.
(56) Hardin, C.; Eastwood, M. P.; Prentiss, M. C.; Luthey-Schulten, Z.; Wolynes, P. G. Associative memory Hamiltonians for structure prediction without homology: alpha/beta proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1679−1684.
(57) Ragonnet-Cronin, M.; Hodcroft, E.; Hué, S.; Fearnhill, E.; Delpech, V.; Leigh Brown, A.; Lycett, S. Automated analysis of phylogenetic clusters. *BMC Bioinf.* **2013**, *14*, 317.
(58) Tejada, E.; Minghim, R.; Nonato, L. G. On Improved Projection Techniques to Support Visual Exploration of Multi-Dimensional Data Sets. *Inf. Vis.* **2003**, *2*, 218−231.
(59) Best, R. B.; Hummer, G. Microscopic interpretation of folding ϕ-values using the transition path ensemble. *Proc. Natl. Acad. Sci. U.S.A* **2016**, *113*, 3263−3268.
(60) Sanches, M. N.; Knapp, K.; Oliveira, A. B.; Wolynes, P. G.; Onuchic, J. N.; Leite, V. B. P. Examining the Ensembles of Amyloid-β Monomer Variants and Their Propensities to Form Fibers Using an Energy Landscape Visualization Method. *J. Phys. Chem. B* **2022**, *126*, 93−99.
(61) 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., Bioinf.* **2018**, *86*, 1184−1188.
(62) de Godoi Contessoto, V.; Ramos, F. C.; de Melo, R. R.; de Oliveira, V. M.; Scarpassa, J. A.; de Sousa, A. S.; Zanphorlin, L.ıc. M.; Slade, G. G.; Leite, V. B. P.; Ruller, R. Electrostatic Interaction Optimization Improves Catalytic Rates and Thermotolerance on Xylanases. *Biophys. J.* **2021**, *120*, 2172−2180.
(63) 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.

https://doi.org/10.1021/acs.jpcb.2c06178
J. Phys. Chem. B 2023, 127, 1291−1300