How Well Does a Funneled Energy Landscape Capture the Folding Mechanism of Spectrin Domains?

Robert B. Best*

Department of Chemistry, Cambridge University, Lensfield Road Cambridge CB2 1EW, U.K.

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, United States

Ⓢ Supporting Information

ABSTRACT: Three structurally similar domains from α-spectrin have been shown to fold very differently. First, there is a contrast in the folding mechanism, as probed by Φ-value analysis, between the R15 domain and the R16 and R17 domains. Second, there are very different contributions from internal friction to folding: the folding rate of the R15 domain was found to be inversely proportional to solvent viscosity, showing no apparent frictional contribution from the protein, but in the other two domains, a large internal friction component was evident. Non-native misdocking of helices has been suggested to be responsible for this phenomenon. Here, I study the folding of these three proteins with minimalist coarse-grained models based on a funneled energy landscape. Remarkably, I find that, despite the absence of non-native interactions, the differences in folding mechanism of the domains are well captured by the model, and the agreement of the Φ-values with experiment is fairly good. On the other hand, within the context of this model, there are no significant differences in diffusion coefficient along the chosen folding coordinate, and the model cannot explain the large differences in folding rates between the proteins found experimentally. These results are nonetheless consistent with the expectations from the energy landscape perspective of protein folding, namely, that the folding mechanism is primarily determined by the native-like interactions present in the Gō-like model, with missing non-native interactions being required to explain the differences in "internal friction" seen in experiment.
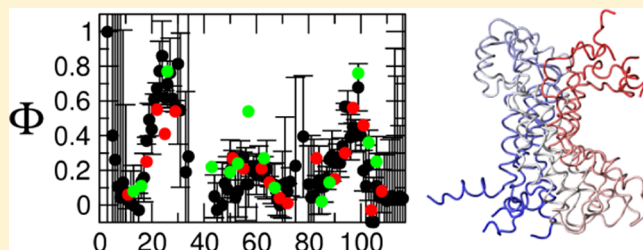
INTRODUCTION

One of the key features of the energy landscape theory of protein folding is that the folding mechanism is largely determined by the formation of native contacts, leading to a "funnel"-shaped energy landscape in which energy decreases with increasing formation of native structure.[1] An idealized funnel landscape may be realized for a given protein by constructing a Gō model, in which the only favorable pair interactions are those present in the native structure,[2] and such models have proved successful in predicting folding and binding mechanisms.[3,4] In many cases, a single order parameter, such as the fraction of native contacts, is a good reaction coordinate for capturing the dynamics of folding. Specifically, folding may be modeled as diffusion on the one-dimensional free energy surface for this coordinate, which is usually barrier-limited due to the imperfect cancellation of changes in energy and chain entropy as the native state is approached[5-7] (to be precise, the energy referred to here is the free energy as a function of the protein coordinates[8]). The diffusion coefficient should be determined by the local "roughness" of the energy landscape, that is, the heights of energy barriers between different protein configurations, for example, due to torsion angle rotations or formation of contacts, as well as by contributions from the solvent viscosity.[9-11] In this picture,

non-native contacts do not play any significant role in determining the mechanism of folding but may modify the effective diffusion coefficient along the folding coordinate.[5,12] Here, mechanism refers to the order in which native structure is assembled on folding transition paths; non-native contacts can of course affect the free energy surface and folding rates.[13,14]

Experimental measurements in which the solvent viscosity is varied therefore have the potential to tease apart the contributions to friction from solvent viscosity, and from internal friction due to the protein itself. Early efforts to identify a contribution from internal friction to the folding rate, by adding small molecule viscogenic agents, found it to be negligible for certain all-β and α/β proteins.[15-17] Internal friction was, however, detected for α-helical peptides[18] and small, α-helical fast-folding proteins[19-21] and has also been inferred for fast-folding β-proteins.[22] This discrepancy had been attributed to a greater sensitivity to internal friction contributions in the experiments on fast-folding proteins,[20]

due to the greater relative importance of the prefactor in determining the folding rate.

The folding of the three spectrin domains from α-spectrin (R15, R16, R17) has several interesting connections with energy landscape theory. Despite their similar structure, extensive experimental folding studies[23−29] have shown them to differ in two important respects. First, there are differences in the folding mechanism, which have been revealed by Φ-value analysis.[23−26] Second, despite folding on a millisecond time scale, evidence for internal friction has been obtained in two of the domains.[27] The viscosity of the solvent was varied by addition of a small-molecule viscogen, together with compensating amounts of chemical denaturant in order to offset the effect of the viscogen on protein stability, thereby (assuming a linear free energy relation to hold) minimizing the effect on the free energy barrier to protein folding. This revealed that, while the folding rate of one of the domains, R15, scaled inversely with the solvent viscosity (suggesting little contribution from internal friction to the folding rate), a strong indication of internal friction effects was obtained for the R16 and R17 domains. Given the close structural similarity of the domains (Figure 1), the exact origin of the differences in folding mechanism and internal friction effects remains unclear; molecular simulations may be able to shed some light on this question.
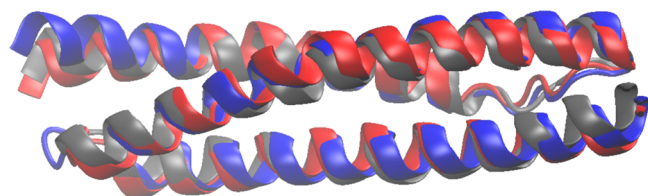

Figure 1. Native structures of the three spectrin domains overlaid: blue, R15; red, R16; gray, R17. Residues not present in the crystal structure 1U4Q[30] have been added in a helical conformation.

Here, I use structure-based (Gō) models in order to describe the folding of these three domains. The simulations predict the existence of sequential folding barriers in the energy landscape of R16 and a single barrier for R15, in qualitative agreement with experiment; a model with additional non-native interactions results in a broad barrier for both R16 and R17. Furthermore, Φ-values computed from the simulations are in fairly good agreement with experiment, considering the approximations involved. Thus, a smooth, "funneled" energy landscape appears to describe the qualitative features of the folding free energy landscape, and consequently of the folding mechanism. However, there are no evident differences in the diffusion coefficient along the reaction coordinate between the three domains, nor in the folding rates. Thus, the differences in internal friction between the domains are not captured by the model, suggesting that such effects are most likely due to non-native interactions, for example, the mis-docking of helices which has been previously suggested.[27] Such a mis-docking mechanism is very plausible, considering the structure of the early transition states of R16 and R17 in the simulation (those in which internal friction has been identified experimentally[29]).

■ METHODS

Model and Simulation Methods. The Gō model for each protein was constructed from its native structure, taken from protein data bank entry 1U4Q[30] (R15, 1662−1773; R16, 1764−1879; R17, 1870−1979). In order to have all domains of the same length (116), residues which were present in the folding experiments but not in the crystal structure were added to the termini in a helical conformation: for R15, the N-terminal sequence KLKE was added, and for R17, the C-terminal sequence NSAFLQ was added. For each domain, a coarse grained Gō model was constructed using 1 bead per residue, according to the procedure described by Karanicolas and Brooks.[31] The key features of the model are that the only favorable residue−residue contacts are those present in the native structure, pseudo-bond lengths and angles are taken from the experimental structure, and pseudo-dihedral angles are given by a statistical potential based on their distribution in a subset of the protein data bank.[32] The initial model was generated via the standard algorithm, following which all native contact energies were scaled by a factor of 1.25 in order to obtain folding temperatures of ∼350 K. The original version of the model includes only repulsive non-native interactions. A variant was also constructed in which non-native interactions were treated instead using a transferable protein−protein binding potential developed by Kim and Hummer;[33] the approach to combining this with the Gō model has been described previously.[11,34]

All simulations were run with the CHARMM simulation code,[35] using a Langevin dynamics integrator[36] with a friction coefficient of 0.1 ps^{-1}. Pseudo-bond lengths were fixed to their native values using the SHAKE algorithm[37] Free energy surfaces for each protein were constructed from a set of umbrella sampling simulations using the fraction of native contacts Q as a reaction coordinate. A restraint potential of the form $V(Q) = k_Q(Q - Q_0)^2/2$ was used, with a force constant of $k_Q = 600$ kcal/mol and 16 windows with different target values Q_0 spanning the range 0−1; replica exchange moves were allowed between umbrella windows adjacent on Q. The weighted histogram method was used to combine the results from different temperatures and umbrella windows to compute thermodynamic quantities.[38]

Diffusion coefficients on the reaction coordinate Q were calculated by discretizing Q and using a Bayesian procedure to determine local free energies $G(Q_i)$ and diffusion coefficients $D(Q_i)$ for each grid point Q_i from the propagators $p(Q_j, t|Q_i, 0)$ estimated from the simulations, as described previously.[11,39−41] One important difference from previous work on fast-folding proteins was that, due to the slow folding rates of these domains, it was not possible to estimate diffusion coefficients from long equilibrium simulations. Instead, diffusion coefficients were estimated from simulations in which an inverse umbrella potential $-F(Q)$ was added to the total system energy. The potential was represented as a cubic spline function. This approach has been shown to yield comparable results to equilibrium simulations, where Q is a good reaction coordinate.[41]

■ RESULTS AND DISCUSSION

The approach taken here is to investigate to what extent "structure-based" or "Gō-like" models can capture the folding mechanism and unusual kinetics of the three spectrin domains considered. In these models, only native contacts are considered to be favorable energetically, and consequently, the energy landscape is as close to a perfect "funnel" as possible, where increasing formation of native structure leads to lowering of the protein energy. The initial model considered is the

generic Gō potential described by Karanicolas and Brooks,[31] where the only element of the energy function not based on the native structure is the torsional potential, which is a statistical potential.

Free Energy Surfaces. The free energy surfaces $F(Q)$ for the fraction of native contacts Q are shown in Figure 2A,C,E for
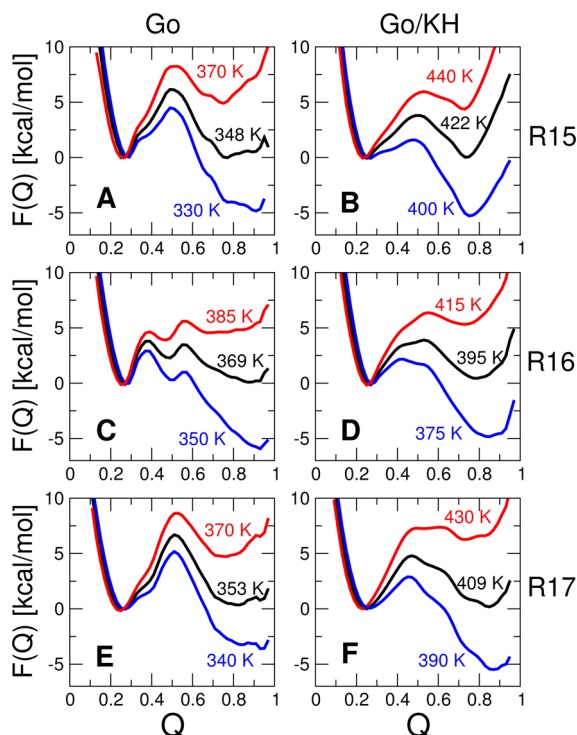

Figure 2. Free energy surfaces $F(Q)$ for the fraction of native contacts Q. A, C, and E are surfaces for the pure Gō model for R15, R16, and R17, respectively, and B, D, and F are the corresponding surfaces for a Gō model with a transferable non-native potential. Black, blue, and red curves are potentials of mean force at the midpoint (folding) temperature T_f and at selected temperatures (displayed) below and above it, respectively.

each protein. As is evident, all the proteins are essentially "two-state"; i.e., there are only two states which are predominantly populated under any given equilibrium conditions.

A few features are immediately worth noting. At the midpoint, R15 and R17 each display a single energy barrier in $F(Q)$, while two "sequential" barriers are evident for R16. This is consistent with the experimental evidence for either sequential barriers[42] or a broad barrier[43] for R16, manifested in a downward curvature of the unfolding rate at high denaturant concentrations,[23,24] while there is no experimental evidence for sequential barriers in R15 and only certain mutants of R17 alter the slope of the unfolding arm of the chevron plot.[25] As in the experiments, the rate-limiting barrier is the early barrier under stabilizing conditions (low temperature in simulations, low denaturant in experiment), with the late barrier becoming prominent under destabilizing conditions (high temperature, high denaturant), consistent with the Hammond postulate, and earlier findings using lattice models.[44] There is also some evidence for heterogeneity in the native states of R15 and R17 (broad native minimum).

Gō-like models, as described above, can in fact exhibit "internal friction" effects, for example, due to topological frustration.[45] However, since Gō models clearly exclude the

possibility of internal friction effects due to attractive non-native interactions, I have also considered a model in which a "transferable" potential is used to describe non-native interactions. That is, instead of having only repulsive non-native interactions, as in the Gō model, the non-native interactions are described by a pair potential which depends only on the chemical identity of the participating residues. This is achieved by combining the standard Gō model as previously described[11] with a potential derived for protein–protein interactions due to Kim and Hummer;[33] this combination will be referred to as the Gō/KH potential. The idea of combining a non-native potential with a Gō model is not novel and has been pioneered by others.[8,46−50] It is important to note that residue pairs may be either mutually attractive or repulsive in this potential, depending on their types. The free energy surfaces for this modified model are shown in Figure 2B,D,F. In general, the form of the free energy surfaces is still similar to that for the pure Gō model. R15 is still clearly two-state, with the barrier in much the same position (on Q) as before; the fully native state at high Q has become less stable, probably due to frustration arising from non-native interactions. The barrier is slightly lowered (seen also for the other two proteins) by non-native interactions, a finding which is consistent with predictions from theory and simulation,[51,52] although some simulation studies have found that non-native interactions may raise the folding barrier.[46] However, the sequential transition states of R16 are less distinct, and appear as part of a single broad barrier. This may be because the non-native interactions destabilize the high energy intermediate, but it may also be due partly to the Q coordinate being slightly less suitable for describing free energy surfaces including non-native inter-actions. Another interesting difference from the pure Gō models is that the surface for R17 also starts to show more of a "broad" barrier, which may be consistent with the ability of certain mutants to shift the slope of the unfolding arm of the chevron.[25] Notably, the addition of the non-native interactions shifts the major barrier at the midpoint temperature and below toward an earlier position on the reaction coordinate, similar to R16.

In order to confirm the relation between the observed sequential transition states or broad free energy barriers and the experimental observation of a nonlinear dependence of unfolding rate on denaturant, I have determined the analogue of an experimental denaturant-dependent rate by varying the temperature in the simulations. Since carrying out explicit simulations at every temperature would be quite demanding given the height of the barrier for some of the proteins, we have instead used Kramers theory (eq 3, described below) to approximate the rate at each temperature from the temper-ature-dependent free energy surface. Eyring plots of the unfolding rate in Figure 3 show that I indeed obtain a slight inflection in the rate at the temperature at which the position of the transition state changes, for both R16 and R17 (the inflection is also slight in experiment). Such an inflection is missing for the Gō model of R15 and is much weaker for the Gō/KH model of that protein.

It should also be noted that sequential transition states or broad barriers, while sufficient to explain nonlinear chevron plots, may not necessarily be the cause. For example, under conditions strongly favoring folding, strong rollover effects may have other causes, such as trapping in compact non-native configurations, or internal friction effects.[45,53] This type of effect may be the origin of the strong rollover seen in the
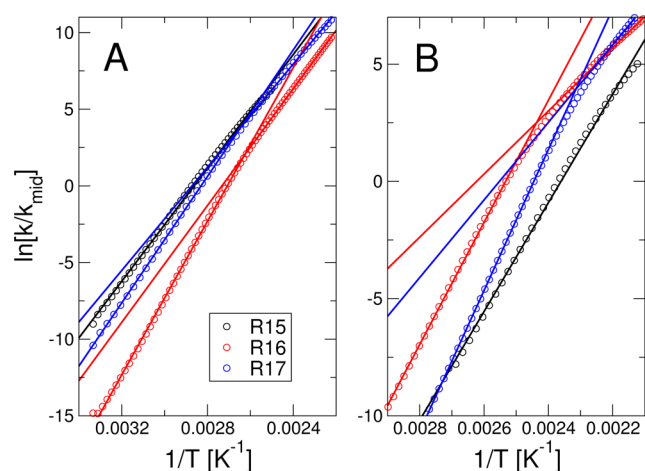
Figure 3. Unfolding "limb" of chevron plots for (A) the pure Gō model and (B) the Gō model with non-native contacts added. Symbols represent the log-ratio of the unfolding rate relative to the midpoint unfolding rate, as estimated from the free energy surfaces using eq 3. Lines are linear fits to the full data range for R15 and separate fits to high and low temperature data for R16 and R17.

folding of certain designed repeat proteins.[54] In the case of the spectrin domains, such an explanation is less likely, because there is no evidence in experiment for "rollover" in the folding rates, with the folding arm of the chevron being linear; only at relatively high denaturant concentration is there an inflection of the unfolding arm, which can be interpreted as a switch in barrier position.[25]

Folding Mechanism. In order to compare the Gō model and Gō/KH models more quantitatively with the mechanism deduced from experiment, I have turned to folding Φ-values. The folding of each of the three spectrin domains has been extensively characterized via Φ-value analysis, including separate sets of Φ-values for the early and late transition states of R16. Experimentally, Φ-values are determined by considering the relative effect of a single-point mutation on the folding rate k_{fold} and the equilibrium constant K_{eq} of a protein. That is, the Φ-value for a particular mutation is given by

$$\Phi = \frac{\ln[k_{fold}(wt)/k_{fold}(mut)]}{\ln[K_{eq}(wt)/K_{eq}(mut)]} \quad (1)$$

The effect of any single point mutation on the energy function has been defined as a weakening of all native contact pairs involving the mutated residue; I have chosen to scale the pair contact energies by a factor of 1/2. I obtain the equilibrium constant directly from the umbrella sampling simulations used to generate the free energy surfaces, by integrating over $F(Q)$ as

$$K_{eq} = \int_{1/2}^{1} \exp[-\beta F(Q)] \, dQ \Big/ \int_{0}^{1/2} \exp[-\beta F(Q)] \, dQ \quad (2)$$

where $\beta = 1/k_B T$. A dividing surface of $Q = 1/2$ has been used in all cases, although the equilibrium constant is expected to be insensitive to this choice given the relatively high free energy barriers for these proteins. The free energy surfaces for the mutants were generated by perturbation using multidimensional WHAM[38] and equilibrium constants computed accordingly.

In principle, the rate coefficients could also be computed directly from simulations. Because of the relatively high barriers

involved, it would be inefficient to do this via long equilibrium runs or first-passage time simulations, but it could be achieved via transition-path sampling.[55,56] However, because similar Φ-values have been obtained by considering just the rates or only the contribution of the change in barrier height,[57] a thermodynamic approximation has been adopted here, where the diffusion coefficient along Q, $D(Q)$, is assumed to be unaffected by the mutation; a similar approximation has been used successfully in previous simulation studies,[57,58] and I examine the validity of this approximation further below. To estimate the effect of barrier height on the rate, I use the overdamped Kramers result[59] that

$$k_{fold}^{-1} = D^{-1} \int_{1/2-\Delta}^{1/2+\Delta} \exp[\beta F(Q)] \, dQ \int_{0}^{1/2} \exp[-\beta F(Q)] \, dQ \quad (3)$$

and again compute the wild-type and mutant rates directly from their respective free energy surfaces, using a value of $\Delta = 0.15$. In this expression, I am also making the assumption that D, the diffusion coefficient at the barrier top, is not affected by mutations (the actual value of D is not important for computing Φ-values). The results should be relatively insensitive to the choice of 1/2 as an approximate dividing surface, as long as $[1/2 - \Delta, 1/2 + \Delta]$ includes the barrier region. Where possible, Φ-values were computed at the midpoint temperature (Figure 2). To compute separate Φ-values for the early and late transition states of R16, temperatures below and above the folding midpoint, respectively, were used. For the pure Gō model, these temperatures were 350 and 385 K, while, for the Gō/KH model, temperatures of 375 and 415 K were used (see Figure 2 for corresponding free-energy surfaces). Φ-values determined at the folding midpoint for R17 were very scattered—this likely reflects the very similar free energy of the early and late transition states for this protein, such that some mutations cause a shift in barrier near the midpoint. Therefore, Φ-values were computed at a temperature below the midpoint, 390 K, as the experimental data refer to the early transition state.

The resulting Φ-values are shown in Figure 4, together with their experimental counterparts. For completeness, I have shown the calculated Φ-values for all residues, with the result that some of the errors are very large. This occurs for a similar reason to that causing large errors in experiment, namely, that the change in stability for these residues upon mutation is comparable to the statistical error in the free energy estimate. This is because these residues tend to be forming few or only weak contacts and consequently would most likely be excluded from the experimental data set; the large errors should therefore not affect the comparison with experiment.

The agreement with experiment is generally good, particularly for R15 and R16. The addition of non-native contacts has a relatively small effect, slightly improving agreement with experiment for the late R16 transition state. This result is similar to what was obtained by Gin et al., who found that inclusion of non-native interactions affected the folding rates but not the mechanism.[12] There are several obvious outliers toward the center of the sequence and toward the end: these are the mutations H48A, A46G, A50G, A63G, A101G, A103G, A106G, and L108A. All of these residues have very high Φ-values, with $\Phi \geq 0.8$. Nonetheless, the overall pattern of Φ-values through the sequence is reasonably well reproduced, including the increase of structure toward the N-terminus and center of the sequence in the late R16 transition
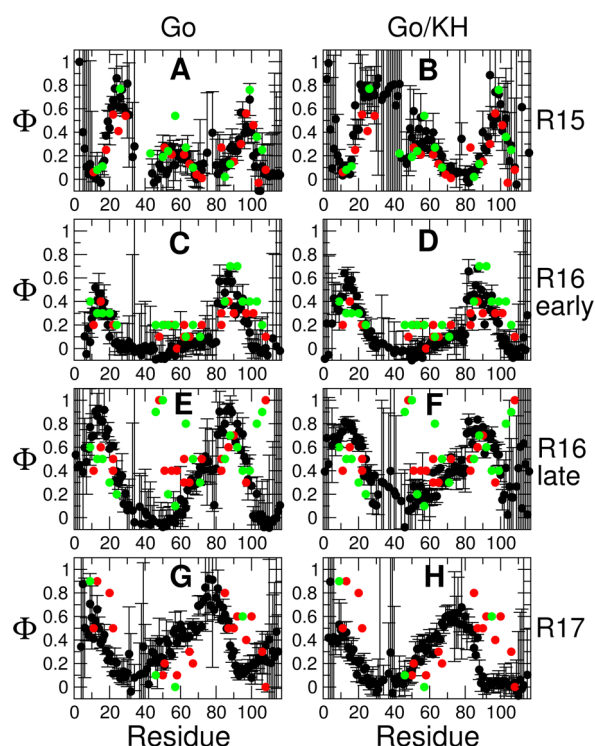
Figure 4. Φ-values estimated from simulation are shown by black symbols, with experimental Φ-values from "core" and Ala → Gly mutations indicated by red and green, respectively. A, C, E, and G are Φ-values estimated from a pure Gō model for R15, the early R16 transition state, late R16 transition state, and R17; B, D, F, and H are the corresponding results for a Gō model in which non-native interactions are treated by a transferable potential.

state, relative to the early one. The match of the calculated Φ-values for R17 is less satisfactory. It appears that the experimental pattern of Φ-values is quite similar for R16 and R17, while the high Φ-value region of the R17 simulations has shifted toward the center of the sequence. Thus, it is most likely that the relative interaction strengths stabilizing different regions of the structure are less well captured for R17 than for R16. Nonetheless, in comparing the results with R15, it is clear that the Gō and Gō/KH models do reflect the essential differences in mechanism between R15 and R16/R17. In the Supporting Information, we provide corresponding scatter plots for the data, and a table of parameters describing the agreement with experiment. For the pure Gō model, the overall RMSD to experiment is 0.16, 0.19, 0.26, and 0.25 for R15, R16 early transition state, R16 late transition state (after excluding outliers), and R17, respectively, with the corresponding linear correlation coefficients being 0.76, 0.67, 0.68, and 0.25. A calculation of the mutual correlation between the different sets of experimental and simulated Φ-values (Table S2, Supporting Information) indicates that each experimental data set matches most closely the simulation of the corresponding protein, with the exception of the R17 data which matches the R16 simulatons more closely; however, the experimental Φ-values for R16 and R17 are quite similar. Because we model mutations using only contacts, one might expect the agreement to be worse for the helix-scanning Ala → Gly mutations on the surface, relative to the hydrophobic core mutations. Indeed, a number of the very high "outlier" Φ-values for the late R16 transition state are for Ala → Gly. However, the statistics in Table S2 (Supporting Information) do not indicate a

systematic difference in accuracy between surface and core mutations, with the surface residues having lower RMSD than core residues in some cases and higher RMSD than core residues in others. Thus, it seems that the contact scaling model is no worse at capturing the effect of the Ala → Gly mutations than that of hydrophobic core mutations.

Having established consistency between the Φ-values from simulation and experiment, the simulations can be used to provide some structural insights into the folding mechanism. I do this by drawing structures from an equilibrium distribution at key values along the reaction coordinate Q, on the assumption that Q is a sufficiently good reaction coordinate—as it has been found to be for other proteins.[11,56,60] I have chosen to focus on the putative transition states for folding as being the key to establishing the folding mechanism. In order to simplify the analysis, we show only the results for the Gō/KH model in Figure 5; results for the pure Gō model are very similar.

The transition state for R15 folding in Figure 5A shows the formation of ordered helical structure at the center of each of the three helices (diagonal), which are docked against each other, forming native contacts; the overall picture is very consistent with what would be deduced from a conventional interpretation of the Φ-values.[61] The first transition states for R16 (Figure 5B) and R17 (Figure 5D) have some features in common, with native helix formation toward one end of the protein (N-terminus of helices 1 and 3 and C-terminus of helix 2) but relatively few contacts formed between the helices. In the corresponding late transition states, the native helix structure is even more extensive, with the three helices already docked in a native conformation at one end of the molecule; nascent contacts are being formed in the regions which are not yet helical.

What are the key differences between the different mechanisms, as deduced here? As has been concluded from a direct analysis of the experimental Φ-values, all of the transition states are somewhat similar, as reflected in the experimental Tanford β-values, and the transition-state Q-values which are all near 0.5 in the simulations. There is some overall similarity between all of the transition state contact maps. I focus on the comparison of the R15 and R16 transition states, as these are the ones best validated against the experimental Φ-values. In terms of tertiary structure, the R15 transition state is most similar to the R16 early transition state, with an RMSD for long-range contacts (difference of residue indices greater than 5) between contact maps of 0.039, calculated over contacts that are formed in at least one of the transition states. This compares with an RMSD for long-range contacts between the early and late transition states in R16 of 0.038. However, the differences in short-range contacts (defined as not long-range contacts) are larger, an RMSD of 0.23 between R15 and the early R16 transition state but a difference of 0.11 between the early and late R16 transition states. This highlights the key difference between R15 and R16, namely, the greater degree of secondary structure formation already in the early transition state of R16. The reason for the multistate folding of R16/R17 appears to be a slight decoupling between the interactions in the two ends of the (rather elongated) spectrin domain, such that the end with the more stable interactions forms first. In R15, there does not appear to be such a decoupling and therefore native structure forms in a single step. So far, our results do not shed any light directly on the origin of the differences in internal friction observed for these domains.
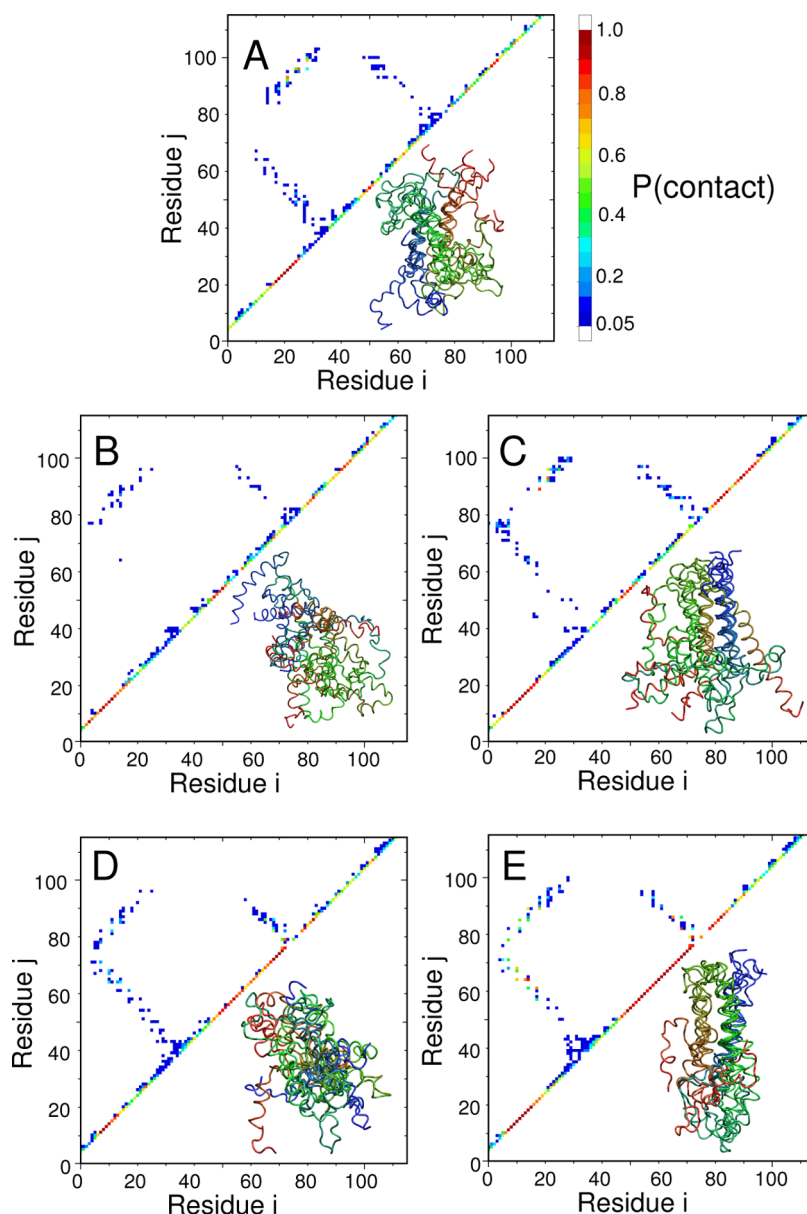
E

dx.doi.org/10.1021/jp403305a | J. Phys. Chem. B XXXX, XXX, XXX–XXX

Figure 5. Structures of putative transitions states: (A) R15 ($Q = 0.5$), (B) R16 early ($Q = 0.39$), (C) R16 late ($Q = 0.55$), (D) R17 early ($Q = 0.46$), (E) R17 late ($Q = 0.63$). All structures were drawn randomly from an equilbrium distribution at the specified Q-values for the Gō/KH model. For each transition state ensemble, an average contact map (upper left) and a representative set of structures (lower right) are shown. Results for the pure Gō model are similar and are not shown for clarity (the R17 transition state in the Gō model is similar to the late R17 transition state in the Gō/KH model).

However, they are consistent: the large internal friction contribution is only observed for the early transition state for R16 and R17, which is absent from R15.[29] Second, the lack of ordered tertiary structure in the first transition state of R16/R17, in contrast to the relatively ordered secondary structure, would support a mechanism for internal friction involving misdocking of helices.

Although folding free energy barrier heights cannot be measured experimentally, it is nonetheless interesting to compare the relative midpoint barrier heights obtained in our model with the order of folding rates measured experimentally at the folding midpoint, i.e., R15 > R16 > R17. I find no correlation with barrier height for the pure Gō model, but the folding barriers of ~3.5, ~3.6, and ~4.5 kcal/mol for the Gō/KH model are consistent with the rank order of experimental

folding rates. The improvement in agreement upon adding non-native interactions is most likely coincidental; however, it is noteworthy that a similar result was obtained by introducing electrostatic interactions into an Ising-like statistical mechanical model for these three proteins (i.e., the correct ranking was only obtained upon addition of electrostatics).[62] However, in neither that work nor the present work was the difference in barrier height sufficient to explain the difference of over 2 orders of magnitude in folding rate between R15 and the other two domains.

Folding Dynamics. The unusual feature of the folding of this group of spectrin domains is the difference in internal friction, which may help to explain partially the large differences in folding rate. I therefore consider next the folding dynamics as represented by a one-dimensional diffusion model. In principle,
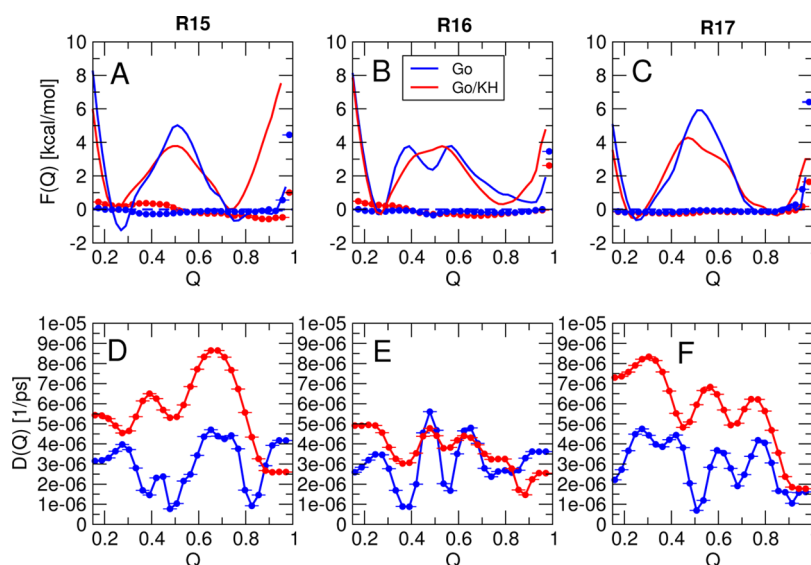
Figure 6. Free energies and position-dependent diffusion coefficients for Q. Top: free energy surfaces for (A) R15, (B) R16, and (C) R17. Bottom: diffusion coefficients for (D) R15, (E) R16, and (F) R17. Blue lines and symbols correspond to the original Gō model, while red lines and symbols correspond to the Gō/KH model. In parts A−C, the lines represent the original potential of mean force $F(Q)$ (see Figure 2), while the symbols represent the free energy estimate from simulations in which an estimate of $-F(Q)$ was added to the simulation energy as a cubic spline.

differences in internal friction should manifest as differences in the diffusion coefficient along the folding coordinate, Q. I use a well established Bayesian method for determining free energies and diffusion coefficients for representing dynamics on a 1D reaction coordinate;[11,39,40] however, due to the high folding energy barriers, an additional trick was needed to sample diffusion coefficients in the high energy regions. Specifically, the free energy surface was "flattened" by adding the inverse potential $-F(Q)$ to the total system energy. If Q is a good reaction coordinate and the dynamics can be represented by 1D diffusion along it, then diffusion coefficients obtained by this procedure should be identical to those for the original energy landscape, as we have previously shown.[41]

In Figure 6A−C, I show the original free energy surfaces (lines), as well as the free energy surfaces resulting from sampling with the added inverse potential (symbols). As intended, the resulting free energy surfaces are flat, allowing sampling of diffusion at all values of Q. Diffusion coefficients for the Gō model and Gō/KH model are presented in Figure 6D−F.

As has previously been observed for other proteins, the variation in diffusion coefficients over the full range of Q is relatively small.[11] This occurs because, even though diffusion is faster in Cartesian space for more unfolded states, the distance which must be traveled to form a contact is larger than it would be for a more native-like state, and the two effects approximately compensate. In addition, there is very little difference between the diffusion coefficients for the different proteins when the Gō model only is considered. While that may not be unexpected, the diffusion coefficients obtained with the additional "non-native" interactions are also rather similar. Across all proteins and the two energy functions, the diffusion coefficients vary by only a factor of 2−3. Therefore, this model of non-native interactions does not explain the differences in folding rate between the different spectrin domains.

It must be noted, however, that the internal friction in the early transition states of R16 and R17 is highly localized, being undetectable in both the unfolded state and second transition

state;[29] in fact, the folding models I use here do correctly capture the similarity of the diffusion coefficients in the unfolded state, for example. Most likely, the internal friction in the first transition state is quite structurally specific in origin, and the specific model of non-native interactions employed here is simply not accurate enough to capture it.

To quantify how much of the variation in rates between the three proteins is captured by the diffusion model, I have determined the midpoint folding rates by constructing a 1D diffusion model in which I combine the unbiased free energy surfaces with the diffusion coefficients obtained with the flat free energy surface (Table 1). The computed folding rates in

Table 1. Folding Rates from the 1D Diffusion Model and from Experiment[26] at the Folding Midpoint (Units of s^{-1})

	R15	R16	R17
Gō	3.2×10^4	7.0×10^5	1.3×10^4
Gō/KH	4.7×10^6	1.1×10^6	2.0×10^6
experiment	9.0×10^1	2.2×10^{-1}	8.2×10^{-2}

Table 1 indicate that neither the Gō nor the Gō/KH model captures the variation in folding rates seen in experiment. The Gō model suggests that R17 is the fastest folder, and while the Gō/KH model correctly indicates that R15 is fastest, the variation in folding rate between the three domains is much smaller than in experiment.

Combining unbiased free energy surfaces with diffusion coefficients obtained from the flat free energy surface assumes that 1D diffusion along Q is a good model for the folding dynamics. That assumption was previously shown to give reliable results for a model of protein G.[41] However, I also find here that the rates from the diffusion model are in reasonable agreement with independent estimates for the same molecular simulation model. For example, computing the folding rate for R15 from the 1D diffusion model gives a rate of 3.2×10^4 s^{-1}, whereas a separate estimate of the folding rate from transition-path sampling[55] resulted in a rate of $\sim 6.4 \times 10^4$ s^{-1}. Note that because of the low solvent friction and smooth energy

landscapes used in the simulations, these rates are much faster than the experimental midpoint folding rate of ~90 s^{-1}; this is a well-known deficiency of coarse-grained Gō models with two-body contact potentials.[63]

The most likely reason for the failure of the Gō/KH model to capture the difference in folding rates between domains is that the model used for non-native interactions underestimates their strength relative to the native contacts. Although one cannot pinpoint the exact cause for this based on the results presented, one weakness of the current non-native model is the treatment of electrostatic interactions such that they may be too weak at short-range. Thus, if the misdocking of helices were driven by misformation of non-native salt-bridges, their strength may be underestimated by the model. Given the involvement of charged residues in slowing the folding of R16 and R17,[28] this may be an significant deficiency in the model. Another respect in which the gap with experiment may be narrowed would be to adopt the same operational definition of internal friction in terms of the dependence of folding rate on solvent viscosity, as done in a recent simulation study.[64]

CONCLUSIONS

The results presented here indicate that a purely "structure-based" model for folding based on only native-like interactions describes fairly well the folding mechanism of the set of three spectrin domains I have studied, as measured by folding Φ-values. The addition of a "transferable" non-native contact potential does improve the free energy surface in some cases (R17) to be more similar to that inferred experimentally. The accuracy of the prediction from a structure-based model is in general support of the energy landscape theory, in which native contacts play the dominant role in determining the folding mechanism,[3] although there are exceptions in which non-native interactions do appear to play an important role in stabilizing folding intermediates.[65,66]

The fact that a purely structure-based model does not distinguish the roughness of the different spectrin domains would suggest that such roughness must arise from non-native interactions—precisely the role these interactions play according to the energy landscape perspective. For example, the structure of the first transition state for the Gō-based models would allow for non-native helix docking as suggested previously. A clear remaining challenge for future simulations and experiments is to resolve the remarkable differences in internal friction between these domains at a molecular level.

ASSOCIATED CONTENT

Supporting Information

Two tables of parameters characterizing agreement between simulated and experimental Φ-values and one figure with scatter plots of simulated and experimental Φ-values. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: robertbe@helix.nih.gov. Phone: (301)496-5414.
Notes
The authors declare no competing financial interest.

REFERENCES

(1) Wolynes, P. G. Recent Successes of the Energy Landscape Theory of Protein Folding and Function. Q. Rev. Biophys. 2005, 38, 405−410.

(2) Ueda, Y.; Taketomi, H.; Gō, N. Studies on Protein Folding, Unfolding and Fluctuations by Computer Simulation. I. The Effects of Specific Amino Acid Sequence Represented by Specific Inter-Unit Interactions. Int. J. Pept. Protein Res. 1975, 7, 445−459.

(3) 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.

(4) Levy, Y.; Wolynes, P. G.; Onuchic, J. N. Protein Topology Determines Binding Mechanism. Proc. Natl. Acad. Sci. U.S.A. 2004, 101, 511−516.

(5) Bryngelson, J. D.; Wolynes, P. G. Intermediates and Barrier Crossing in a Random Energy Model (with Applications to Protein Folding). J. Phys. Chem. 1989, 93, 6902−6915.

(6) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. Navigating the Folding Routes. Science 1995, 267, 1619−1620.

(7) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. Diffusive Dynamics of the Reaction Coordinate for Protein Folding Funnels. J. Chem. Phys. 1996, 104, 5860−5868.

(8) Chan, H. S.; Zhang, Z.; Wallin, S.; Liu, Z. Cooperativity, Local-Nonlocal Coupling, and Nonnative Interactions: Principles of Protein Folding from Coarse-Grained Models. Annu. Rev. Phys. Chem. 2011, 62, 301−326.

(9) Bryngelson, J. D.; Onuchic, J.; Socci, N. D.; Wolynes, P. G. Funnels, Pathways, and the Energy Landscape of Protein Folding: a Synthesis. Proteins 1995, 21, 167−195.

(10) Portman, J. J.; Takada, S.; Wolynes, P. G. Microscopic Theory of Protein Folding Rates. II. Local Reaction Coordinates and Chain Dynamics. J. Chem. Phys. 2001, 114, 5082−5096.

(11) Best, R. B.; Hummer, G. Coordinate-Dependent Diffusion in Protein Folding. Proc. Natl. Acad. Sci. U.S.A. 2010, 107, 1088−1093.

(12) Gin, B. C.; Garrahan, J. P.; Geissler, P. L. The Limited Role of Nonnative Contacts in the Folding Pathways of a Lattice Protein. J. Mol. Biol. 2009, 392, 1303−1313.

(13) Chan, H. S.; Dill, K. A. Transition States and Folding Dynamics of Proteins and Heteropolymers. J. Chem. Phys. 1994, 100, 9238−9257.

(14) Camacho, C. J.; Thirumalai, D. Modeling the Role of Disulfide Bonds in Protein Folding: Entropic Barriers and Pathways. Proteins 1995, 22, 27−40.

(15) Plaxco, K. W.; Baker, D. Limited Internal Friction in the Rate-Limiting Step of a Two-State Protein Folding Reaction. Proc. Natl. Acad. Sci. U.S.A. 1998, 95, 13591−13596.

(16) Jacob, M.; Geeves, M.; Holtermann, G.; Schmid, F. X. Diffusional Barrier Crossing in a Two-State Protein Folding Reaction. Nat. Struct. Biol. 1999, 6, 923−926.

(17) Frauenfelder, H.; Fenimore, P. W.; Chen, G.; McMahon, B. H. Protein Folding Is Slaved to Solvent Motions. Proc. Natl. Acad. Sci. U.S.A. 2006, 103, 15469−15472.

(18) Jas, G. S.; Eaton, W. A.; Hofrichter, J. Effect of Viscosity on the Kinetics of α-Helix and β-Hairpin Formation. J. Phys. Chem. B 2001, 105, 261−272.

(19) Qiu, L.; Hagen, S. J. A Limiting Speed for Protein Folding at Low Solvent Viscosity. J. Am. Chem. Soc. 2004, 126, 3398−3399.

(20) Pabit, S.; Roder, H.; Hagen, S. J. Internal Friction Controls the Speed of Protein Folding from a Compact Configuration. *Biochemistry* **2004**, *126*, 0−6.

(21) Cellmer, T.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. Measuring Internal Friction in Ultra-Fast Protein Folding Kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18320−18325.

(22) Liu, F.; Nakaema, M.; Gruebele, M. The Transition State Transit Time of WW Domain Folding Is Controlled by Energy Landscape Roughness. *J. Chem. Phys.* **2009**, *131*, 195101.

(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.; Steward, A.; Fowler, S. B.; Clarke, J. The Folding of Spectrin Domains II: Phi-Value Analysis of R16. *J. Mol. Biol.* **2004**, *344*, 207−221.

(25) Scott, K. A.; Randles, L. G.; Moran, S. J.; Daggett, V.; Clarke, J. The Folding Pathway of Spectrin R16 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.; Gärtner, M.; Choo, W. X.; Batey, S.; Clarke, J. Different Members of a Simple Three-Helix Bundle Protein Family Have Very Different Folding Rate Constants and Fold by Different Mechanisms. *J. Mol. Biol.* **2009**, *390*, 1074−1085.

(27) 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.

(28) 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 Frustrated Folding of Spectrin Domains. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 17795−17799.

(29) 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.

(30) 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.

(31) Karanicolas, J.; Brooks, C. L., III The Origins of Asymmetry in the Folding Transition States of Protein L and Protein G. *Protein Sci.* **2002**, *11*, 2351−2361.

(32) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Authors, M. The Protein Data Bank. *Acta Crystallogr.* **2002**, *D58*, 899−907.

(33) Kim, Y. C.; Hummer, G. Coarse-Grained Models for Simulation of Multiprotein Complexes: Application to Ubiquitin Binding. *J. Mol. Biol.* **2008**, *375*, 1416−1433.

(34) Sirur, A.; Best, R. B. Effects of Interactions with the GroEL Cavity on Protein Folding Rates. *Biophys. J.* **2013**, *104*, 1098−1106.

(35) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(36) Brooks, C. L., III; Brünger, A.; Karplus, M. Stochastic Boundary-Conditions for Molecular Dynamics Simulations of ST2 Water. *Chem. Phys. Lett.* **1984**, *105*, 495−500.

(37) Ryckaert, J. P.; Cicotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(38) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(39) Hummer, G. Position-Dependent Diffusion Coefficients and Free Energies from Bayesian Analysis of Equilibrium and Replica Molecular Dynamics Simulations. *New J. Phys.* **2005**, *7*, 34.

(40) Best, R. B.; Hummer, G. Diffusive Model of Protein Folding Dynamics with Kramers Turnover in Rate. *Phys. Rev. Lett.* **2006**, *96*, 228104.

(41) Best, R. B.; Hummer, G. Diffusion Models of Protein Folding. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16902−16911.

(42) Sánchez, I. E.; Kiefhaber, T. Hammond Behaviour versus Ground State Effects in Protein Folding: Evidence for Narrow Free Energy Barriers and Residual Structure in Unfolded States. *J. Mol. Biol.* **2003**, *327*, 867−884.

(43) Ternström, T.; Mayor, U.; Akke, M.; Oliveberg, M. From Snapshot to Movie: ϕ Analysis of Protein Folding Transition States Taken One Step Further. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14854−14859.

(44) Gutin, A. M.; Abkevich, V. I.; Shakhnovich, E. I. A Protein Engineering Analysis of the Transition State for Protein Folding: Simulation in the Lattice Model. *Folding Des.* **1997**, *3*, 183−194.

(45) Kaya, H.; Chan, H. S. Origins of Chevron Rollovers in Non-Two-State Protein Folding Kinetics. *Phys. Rev. Lett.* **2003**, *90*, 258104.

(46) 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.

(47) 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.

(48) 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.

(49) 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* **2012**, *80*, 858−870.

(50) Shental-Bechor, D.; Smith, M. T. J.; MacKenzie, D.; Broom, A.; Marcovitz, A.; Ghashut, F.; Go, C.; Bralha, F.; Meiering, E. M.; Levy, Y. Nonnative Interactions Regulate Folding and Switching of Myristoylated Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 17839−17844.

(51) Plotkin, S. S. Speeding Protein Folding beyond the Go Model: How a Little Frustration Sometimes Helps. *Proteins* **2001**, *45*, 337−345.

(52) Clementi, C.; Plotkin, S. S. The Effect of Nonnative Interactions on Protein Folding Rates: Theory and Simulation. *Protein Sci.* **2004**, *13*, 1750−1766.

(53) Kaya, H.; Chan, H. S. Solvation Effects and Driving Forces for Protein Thermodynamic and Kinetic Cooperativity: How Adequate Is Native-Centric Topological Modeling. *J. Mol. Biol.* **2003**, *326*, 911−931.

(54) Javadi, Y.; Main, E. R. Exploring the Folding Energy Landscape of a Series of Designed Consensus Tetratricopeptide Repeat Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 17383−17388.

(55) Hummer, G. From Transition Paths to Transition States and Rate Coefficients. *J. Chem. Phys.* **2004**, *120*, 516−523.

(56) Best, R. B.; Hummer, G. Reaction Coordinates and Rates from Transition Paths. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6732−6737.

(57) Shea, J.-E.; Onuchic, J. N.; Brooks, C. L., III. Exploring the Origins of Topological Frustration: Design of a Minimally Frustrated Model of Fragment B of Protein A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12512−12517.

(58) Naganathan, A. N.; Orozco, M. The Protein Folding Transition-State Ensemble from a Gō-Like Model. *Phys. Chem. Chem. Phys.* **2011**, *13*, 15166−15174.

(59) Kramers, H. A. Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions. *Physica* **1940**, *7*, 284−303.

(60) Nymeyer, H.; García, A. E.; Onuchic, J. N. Folding Funnels and Frustration in Off-Lattice Minimalist Protein Landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5921−5928.

(61) Fersht, A. R.; Matouschek, A.; Serrano, L. The Folding of an Enzyme. 1. Theory of Protein Engineering Analysis of Stability and Pathway of Protein Folding. *J. Mol. Biol.* **1992**, *224*, 771−782.

(62) Naganathan, A. N. Predictions from an Ising-Like Statistical Mechanical Model on the Dynamics and Thermodynamic Effects of Protein Surface Electrostatics. *J. Chem. Theory Comput.* **2012**, *8*, 4646−4656.

(63) 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.

(64) Shulz, J. C. F.; Schmidt, L.; Best, R. B.; Dzubiella, J.; Netz, R. R. Peptide Chain Dynamics in Light and Heavy Water: Zooming in on Internal Friction. *J. Am. Chem. Soc.* **2012**, *134*, 6273−6279.

(65) Capaldi, A. P.; Kleanthous, C.; Radford, S. E. Im7 Folding Mechanism: Misfolding on a Path to the Native State. *Nat. Struct. Biol.* **2002**, *9*, 209−216.

(66) Forman, J. R.; Yew, Z. T.; Qamar, S.; Sandford, R. N.; Paci, E.; Clarke, J. Non-Native Interactions Are Critical for Mechanical Strength in PKD Domains. *Structure* **2009**, *17*, 1582−1590.