THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2019, 123, 4497−4504

pubs.acs.org/JPCB

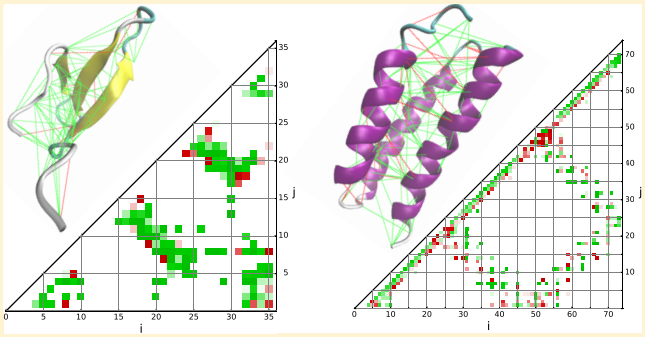
Localizing Frustration in Proteins Using All-Atom Energy Functions

Justin Chen,[†,‡] Nicholas P. Schafer,[†,§] Peter G. Wolynes,[†,‡,§,∥] and Cecilia Clementi*[,†,‡,§,⊥]

[†]Center for Theoretical Biological Physics, [‡]Department of Physics and Astronomy, [§]Department of Chemistry, [∥]Department of Biosciences, and [⊥]Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas 77005, United States

Supporting Information

ABSTRACT: The problems of protein folding and protein design are two sides of the same coin. Protein folding involves exploring a protein's configuration space given a fixed sequence, whereas protein design involves searching in sequence space given a particular target structure. For a protein to fold quickly and reliably, its energy landscape must be biased toward the folded ensemble throughout its configuration space and must lack deep kinetic traps that would otherwise frustrate folding. Evolution has "designed" the sequences of many naturally occurring proteins, through an eons-long process of random mutation and selection, to yield landscapes with a minimal degree of frustration. The task facing humans hoping to design protein sequences that fold into particular structures is to use the available approximate energy functions to sculpt funneled landscapes that work in the laboratory. In this work, we demonstrate how to calculate several localized frustration measures using an all-atom energy function. Specifically, we employ the Rosetta energy function, which has been used successfully to design proteins and which has a natural pairwise decomposition that is suitably solvent-averaged. We calculate these newly developed frustration measures for both a mutated WW domain, FiP35, and a three-helix bundle that was designed completely by humans, Alpha3D. The structure of FiP35 exhibits less localized frustration than that of Alpha3D. A mutation toward the consensus sequence for WW domains in FiP35, which has been shown unexpectedly in experiment to disrupt folding, induces localized frustration by disrupting the hydrophobic core. By performing a limited redesign on the sequence of Alpha3D, we show that some, but not all, mutations that lower the energy also result in decreased frustration. The results suggest that, in addition to being useful for detecting residual frustration in protein structures, optimizing the localized frustration measures presented here may be a useful and automatic means of balancing positive and negative design in protein design tasks.



INTRODUCTION

Successful protein design requires both a positive design to stabilize the target folded structure and a negative design to destabilize possible traps on the energy landscape that might result in misfolding.[1−3] Traps in the energy landscapes of proteins can be stabilized by relatively strong non-native interactions. Thus, simply achieving a stable design does not guarantee that clean folds will result. Not only must a design be stable but it must also be separated by an energetic gap from alternatives. This is one way of stating the minimal frustration principle of protein folding.[3−7] It was shown many years ago that a protein design based on energy gap could improve the foldability relative to designing based on stability alone.[3] Of course, designs can fail locally as well as globally. Thus, it is desirable to be able to localize where frustration exists in biomolecules.[8−14] An algorithm for computing localized frustration measures in protein structures has been proposed as a way of identifying frustrated contacts in folded structures.[8,11] The original "frustratometer" is based on the AMW/AWSEM energy function,[8] a coarse-grained solvent-averaged model, which has been successfully used to predict the folded structures of proteins and to study mechanisms of protein misfolding,[12] binding,[8,13] and aggregation.[13] What residual frustration remains in folded structures is found often to be related to the function. The AWSEM frustratometer has been used to show that residual frustration can be found in parts of protein structures that are involved in protein-binding interfaces,[8] conformational dynamics,[14] ligand binding,[13] enzyme active sites,[9] and other functions.[11]

Here, we introduce a way to compute localized frustration using an all-atom model, the Rosetta energy function, which has been used successfully to design proteins as well as to predict structures.[15] In analogy to what was done with the AWSEM frustratometer, two types of Rosetta-based frustration measures are developed here. These algorithms provide complementary information: one is an atomistic "mutational" frustration measure and the other is an atomistic "configurational" frustration measure. To estimate localized frustration, one must first compute the strength of interactions within the folded ensemble and then compare that strength with the

Received: February 17, 2019
Revised: May 7, 2019
Published: May 7, 2019

DOI: 10.1021/acs.jpcb.9b01545
J. Phys. Chem. B 2019, 123, 4497−4504

strengths of a set of decoy interactions. The two frustration measures differ primarily in the approach that is used to compute the decoy energy distribution. In the atomistic mutational frustration calculation, decoy interactions are computed by mutating pairs of residues that are in close contact in the folded structure and computing the pair energy between the mutated pairs. The atomistic mutational frustration measure therefore provides information about the specificity of evolution selecting particular interactions in the structural context in which they are found in the folded structure. In the atomistic configurational frustration calculation, the decoy energy distribution is obtained by computing the interaction energies of the same pairs of residues when they interact in an ensemble of non-native structures. Unlike the atomistic mutational frustration measure, which can be computed using only structural information about the native backbone configuration, computing the atomistic configurational frustration requires a means of generating non-native decoy configurations. The AWSEM frustratometer solves this problem by sampling randomly from the distribution of configurational parameters within the folded structure when computing decoy interactions. Here, we compute the atomistic configurational frustration measure by obtaining an ensemble of non-native structures from long molecular dynamics simulation trajectories.

We apply the new Rosetta-based frustration measures to the two proteins that are shown in Figure 1. The first protein we consider is FiP35 (Figure 1a), which is a fast-folding β-sheet protein in the WW domain family that was originally a designed variant of the hPin1 WW domain.[16,17] We pay special attention to two residues, W8 and F30, both of which show a high degree of conservation in the WW domain family.[18] The second protein is a human-designed α-helical protein, Alpha3D (Figure 1b), one of the first de novo designed proteins with a well-packed hydrophobic core.[19−21]

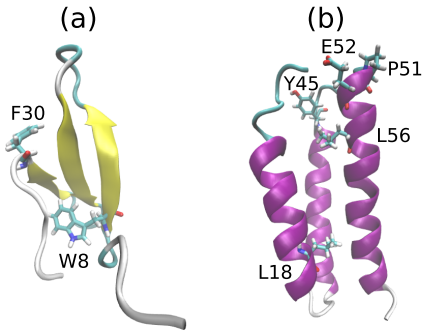


Figure 1. Folded structures of (a) FiP35 and (b) Alpha3D. β-Strands are shown in yellow, α-helices are shown in purple, and loops are shown in white and cyan. The side-chains of select residues are shown in an all-atom representation and are labeled with their corresponding one-letter amino acid-type code and sequence index. (a) Side-chains of residues that are highly conserved across WW domains (W8 and F30) are shown. (b) N-terminus is on the top left of the structure and residues L18, Y45, P51, E52, and L56 are shown.

METHOD: ALL-ATOM LOCALIZED FRUSTRATION CALCULATIONS

The pairwise frustration index is defined as follows[8]

$$F_{ij} = (\langle E_{ij}^{\text{decoy}} \rangle - E_{ij}^{\text{native}}) / \sqrt{1/N \sum_{k=1}^{N} (E_{ij}^{\text{decoy},k} - \langle E_{ij}^{\text{decoy}} \rangle)^2} \quad (1)$$

The pairwise frustration index, F_{ij}, is a Z-score of the native pair energy (E_{ij}^{native}) compared with a distribution of N decoy energies ($E_{ij}^{\text{decoy},k}$, for $k = 1, \ldots N$) for the pair of residues i and j. Frustration measures differ primarily in how the decoy energy distribution is generated, and all measures compute the pairwise frustration index, F_{ij}, as shown in eq 1. Large and positive frustration indices indicate that a pair of residues are minimally frustrated, whereas large and negative frustration indices indicate that the pair of residues is highly frustrated.

The pairwise energies E_{ij} entering eq 1 are computed using the Rosetta energy function as follows

$$E_{ij} = \sum_{k} w_k V_k(\mu_i, \mu_j) \quad (2)$$

We employed the REF2015 version of the Rosetta energy function, which has a set of well-tested default weights, w_k, for each energy term type, V_k, which depend on residue identities μ_i and μ_j.[15] The full Rosetta energy function includes many energy terms, but for this work we use only the energy terms that directly contribute to the pairwise residue interactions, which are the attractive and repulsive Lennard-Jones interactions, electrostatics, and an orientation-dependent solvation term. The individual energy terms entering eq 2 are themselves computed as a sum of interactions between atoms i' and j' in residues i and j, at positions $\mathbf{r}_{i'}$ and $\mathbf{r}_{j'}$, respectively, as follows

$$V_k(\mu_i, \mu_j) = \sum_{i'} \sum_{j'} v_k(\mathbf{r}_{i'}, \mathbf{r}_{j'}, \mu_i, \mu_j) \quad (3)$$

where v_k is the pairwise interaction energy for the atoms between a pair of residues. We compute the pairwise frustration indices for pairs of residues i and j that satisfy $j \geq i + 3$ and that interact in at least 5% of the native conformations. Decoy energy distributions are generated using two different methods, both of which are described below.

The thermally occupied folded ensemble is not necessarily well-represented by a single static structure. We therefore chose to compute native interaction energies using an average of over 100 randomly chosen structures from within an ensemble of structures that were generated during long molecular dynamics simulations. This approach provides a better estimate of the pairwise interaction energies within the native ensemble than would a single evaluation of the pairwise energy in, for example, the protein's X-ray crystal structure. Representatives of the native ensemble were taken from trajectories provided by the Shaw group.[22,23] The limits of the native ensemble were determined previously using a Markov state model.[24] Alternative means of generating native-like structures exist[2,25] and would likely be suitable for performing frustration calculations. The average native interaction energy for a pair of residues is estimated as follows

$$E_{ij}^{\text{native}} = \frac{1}{100} \sum_{k=1}^{100} E_{ij}^{k} \quad (4)$$

DOI: 10.1021/acs.jpcb.9b01545
J. Phys. Chem. B 2019, 123, 4497−4504

where E_{ij}^{k} is the energy between the pair of residues i and j in the representative structure with index k in the native ensemble.

As mentioned above, two methods were employed for generating the decoy energy distributions. To compute the atomistic mutational frustration measure, we generate decoy energies by starting from a set of native-like structures and mutating one or both of the residues in the pair of residues for which the frustration index is being calculated. Because mutating the side-chain without allowing the backbone to relax is likely to result in steric clashes, we take steps to avoid such clashes. The flexibility of the backbone is taken into account through the use of multiple native structures, and a local side-chain repacking is employed to try to avoid steric clashes. Rosetta has built-in methods for performing a mutation with local repacking. We used a "pack_radius" of 10 Å, meaning that Rosetta attempts to repack side-chains that are within 10 Å of the mutated side-chain to avoid steric clashes. Each decoy energy was obtained by randomly selecting a native-like structure from within a prespecified pool, mutating residues within the pair that is being examined, and then discarding the energy if the pairwise interactions were still unphysically large (≥10 Rosetta Energy Units) due to an inability to avoid clashes by repacking. Furthermore, in situations where the decoy energy standard deviation was unnaturally small due to the small number of acceptable mutants, the standard deviation was raised to 0.1 Rosetta Energy Units. The atomistic mutational decoy pairwise energy $E_{ij}^{\mathrm{mut},k}$ is obtained as follows

$$E_{ij}^{\mathrm{mut},k} = E_{ij}(\mu_s, \mu_l | \mu_s \neq \mu_i \text{ OR } \mu_l \neq \mu_j) \tag{5}$$

where the mutations of residues i and j to μ_s and μ_l must be done in such a way that at least one of the residues is different from the original identities μ_i and μ_j. Equation 5 can be substituted into eq 1 to compute the atomistic mutational pairwise frustration F_{ij}^{mut} for N_{mut} mutational decoys (N_{mut} = 1000 for this work).

$$F_{ij}^{\mathrm{mut}} = (\langle E_{ij}^{\mathrm{mut}} \rangle - E_{ij}^{\mathrm{native}}) / \sqrt{1/N_{\mathrm{mut}} \sum_{k=1}^{N_{\mathrm{mut}}} (E_{ij}^{\mathrm{mut},k} - \langle E_{ij}^{\mathrm{mut}} \rangle)^2} \tag{6}$$

To compute the atomistic configurational frustration, we used decoys that were directly taken from long molecular dynamics trajectories that include folding and unfolding events.[22,23] Interactions between any given pair of residues i and j can be observed throughout the protein trajectory. For the purpose of computing decoy energy distributions, we looked specifically at the conformations in the nonfolded states (fraction of native contacts <0.2). Furthermore, for a given residue pair, the interaction energies are included in the decoy energy distribution if and only if the their closest heavy-atom distance R satisfies $R \leq 3$ Å. Longer distance thresholds lead to strongly bimodal decoy energy distributions due to the inclusion of a large number of essentially noninteracting residue pairs with $E_{ij} \sim 0$. We therefore chose to use a relatively short cutoff distance and attempt to correct for any bias that is introduced by using this cutoff distance as discussed below. In effect, the atomistic configurational frustration measure then becomes a comparison between the residue pair's interaction energy in native-like states and their interaction energies in an ensemble of non-native states where they are nonetheless in close contact. Much as was done for the atomistic mutational frustratometer, configurations having pairwise energies above 10 Rosetta Energy Units were removed from the distribution. Also, when the resultant distribution was anomalously narrow, the standard deviation was set to 0.1 Rosetta Energy Units, which only affected a minority of interactions. The average value for the decoy energy, $\langle E_{ij}^{\mathrm{indiv}} \rangle$, between residue pair i and j is then obtained as follows

$$\langle E_{ij}^{\mathrm{indiv}} \rangle = \frac{1}{N_{ij}} \sum_{k}^{N_{ij}} E_{ij}^{k} \tag{7}$$

for the N_{ij} close interactions are found for the residue pair in the pool of all non-native decoy structures. The frustration from the individual pairs, F_{ij}^{indiv}, is then given as follows:

$$F_{ij}^{\mathrm{indiv}} = (\langle E_{ij}^{\mathrm{indiv}} \rangle - E_{ij}^{\mathrm{native}}) / \sqrt{1/N_{ij} \sum_{k=1}^{N_{ij}} (E_{ij}^{\mathrm{indiv},k} - \langle E_{ij}^{\mathrm{indiv}} \rangle)^2} \tag{8}$$

To account for the bias in the computed decoy energies that results from choosing a shorter cutoff distance, a group frustration is computed. The group energies E^{group}'s are computed by summing pairwise residue interaction energies over all residue pairs, i and j, and normalizing this sum by the total number of interactions that are included in the computation, $N_{\mathrm{total}} = \sum_i \sum_{j \geq i} N_{ij}$, to derive the following equation

$$\langle E^{\mathrm{group}} \rangle = \frac{1}{N_{\mathrm{total}}} \sum_{i} \sum_{j \geq i} \sum_{k}^{N_{ij}} E_{ij}^{k} \tag{9}$$

Then, the group frustration F^{group} is defined as follows

$$F^{\mathrm{group}} = (\langle E^{\mathrm{group}} \rangle) / \sqrt{1/N_{\mathrm{total}} \sum_{k=1}^{N_{\mathrm{total}}} (E^{\mathrm{group},k} - \langle E^{\mathrm{group}} \rangle)^2} \tag{10}$$

and the atomistic configurational frustration index is then given by uniformly shifting F_{ij}^{indiv} by F^{group}, giving the following equation

$$F_{ij}^{\mathrm{config}} = F_{ij}^{\mathrm{indiv}} - F^{\mathrm{group}} \tag{11}$$

Further discussion of the configurational frustration index computation is given in the Supporting Information.

RESULTS

Localized Frustration in FiP35: A Fast-Folding and Minimally Frustrated WW Domain. FiP35 is a fast-folding WW domain.[26] Its structure consists of a single β-sheet, which is shown in Figure 2. An analysis of the folding trajectory of FiP35 has indicated that the folding of FiP35 is unfrustrated.[27] Multiple sequence alignments of WW domains show that W8 is a conserved residue and position 8 shows strong mutational coupling to nearby residues according to direct-coupling analysis calculations.[28] Mutations of the conserved tryptophan at position 8 to other amino acids often result in variants that fail to fold.[18,29] The importance of W8 to the folding can be understood by examining the structure of FiP35; W8 appears to form the basis of the hydrophobic core. The first two β-strands, which includes W8, are structured in the transition state of FiP35.[24]

DOI: 10.1021/acs.jpcb.9b01545
J. Phys. Chem. B 2019, 123, 4497−4504

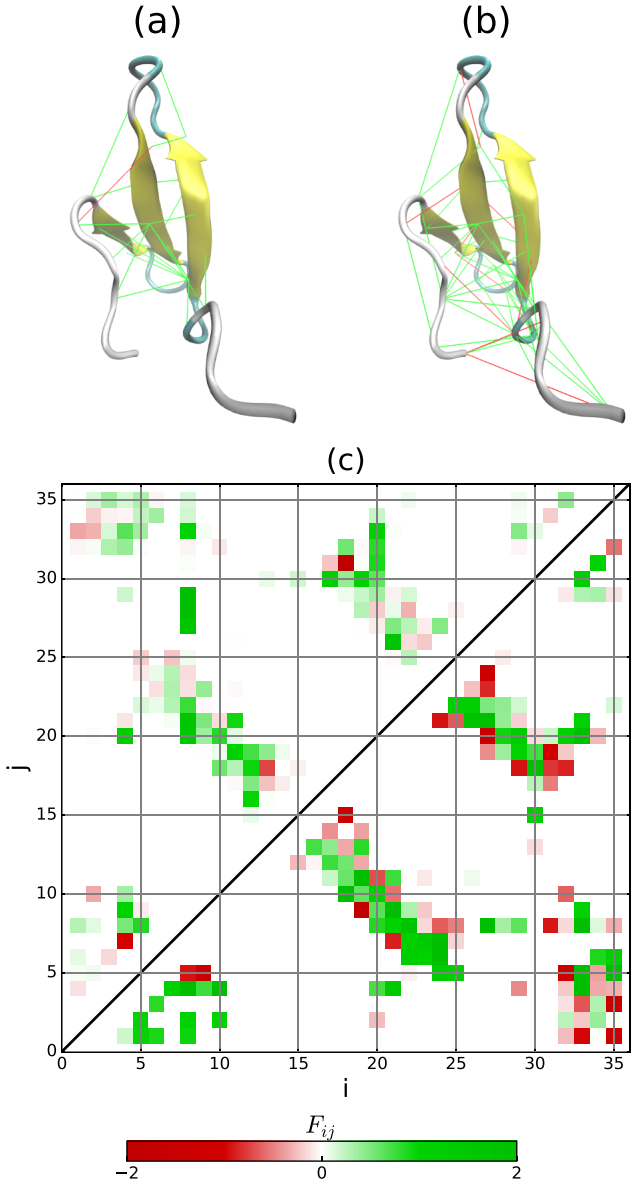


Figure 2. Frustratograms for FiP35 using the atomistic (a) mutational and (b) configurational frustratometer. A green line is drawn between residue pairs whose frustration index F_{ij} satisfies $F_{ij} > 0.78$. Such residue pairs are said to be "minimally frustrated". A red line is drawn between residue pairs with $F_{ij} < -1$, and these residue pairs are said to be "highly frustrated". Residue pairs that are in close contact in the native structure and have $-1 < F_{ij} < 0.78$ are said to be "neutrally frustrated" and are not indicated on the frustratogram for visual clarity. (c) FiP35 F_{ij} values for the atomistic mutational (top left) and configurational (bottom right) frustration measures. According to both frustration measures, FiP35 has very few highly frustrated interactions, which, when present, usually involve at least one residue in the flexible termini. The network of minimally frustrated interactions is more dense according to the atomistic configurational frustration measure than it is for the atomistic mutational frustration measure.

Panels a and b in Figure 2 show three-dimensional representations of the localized frustration patterns on the structure of FiP35 computed using the atomistic recipes. In these "frustratograms", green and red lines are used to indicate particularly unfrustrated and frustrated interactions, respectively. The F_{ij} values are also shown in a contact map style in

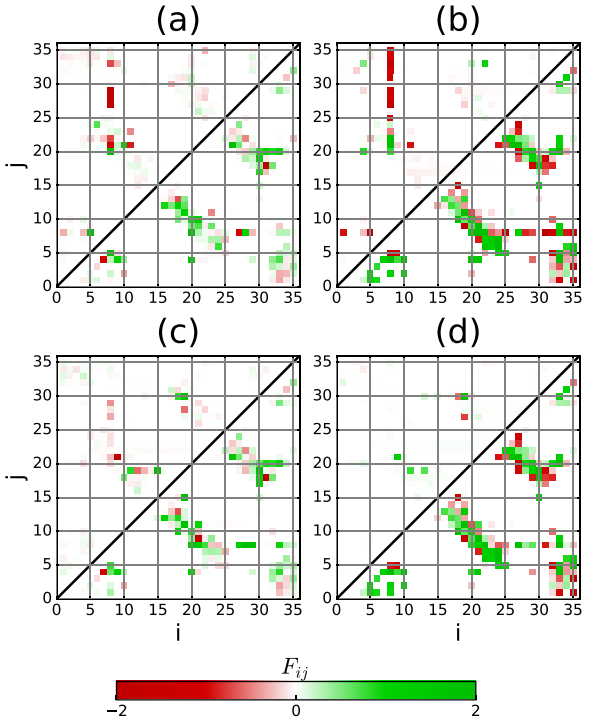


Figure 3. Top half shows the change in frustration upon mutation and the bottom half shows the resultant frustration. For the W8G mutation, both the atomistic (a) mutational and (b) configurational results show a mutation of the hydrophobic core surrounding W8G, and, as expected, it increases the frustration. The F30W mutation is shown for the atomistic (c) mutational and (d) configurational frustratometers.

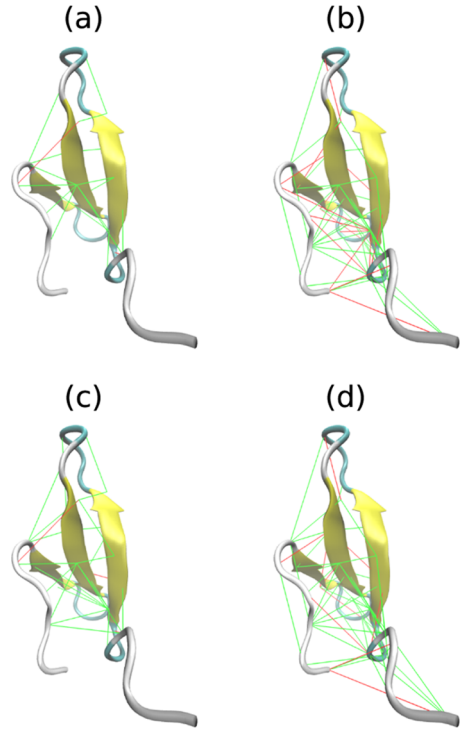


Figure 4. FiP35 frustratograms for the corresponding F_{ij} values in Figure 3 for the W8G mutation with the atomistic (a) mutational and (b) configurational frustratometers and the F30W mutation for the atomistic (c) mutational and (d) configurational frustratometers.

DOI: 10.1021/acs.jpcb.9b01545
J. Phys. Chem. B 2019, 123, 4497−4504

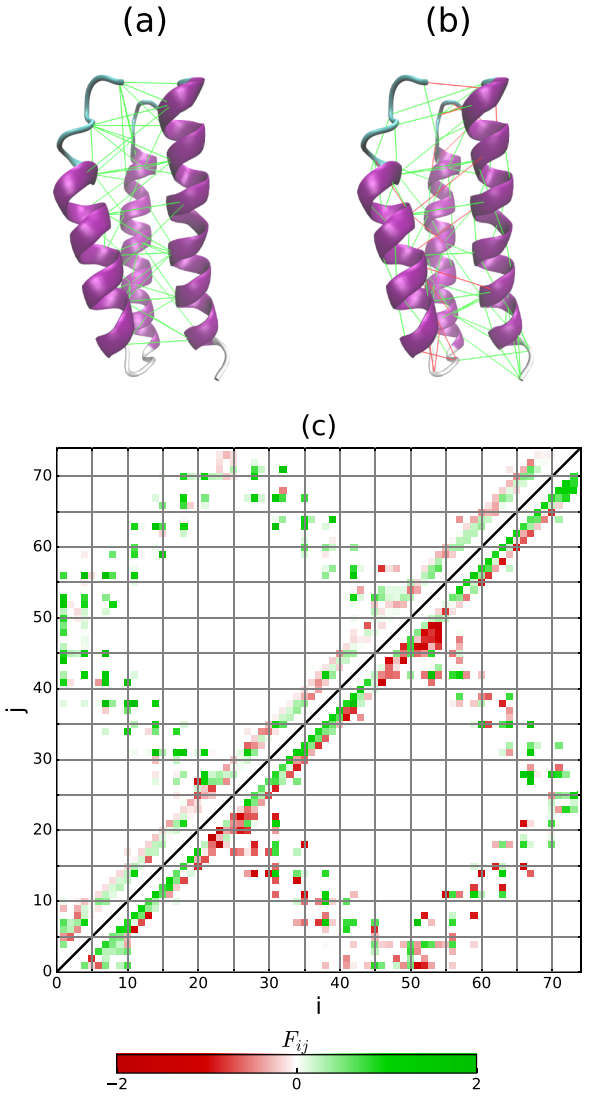


Figure 5. Frustratograms for Alpha3D using the atomistic (a) mutational and (b) configurational methods. (c) F_{ij} values for the atomistic mutational (top half) and configurational (bottom half) methods for Alpha3D. Alpha3D is one of the first de novo designed proteins. During the folding of Alpha3D, non-native contacts that are formed on the interface between helices cause some frustration. Notice the important interactions between the N-terminus and the turn regions of the α-helices, likely explaining why Alpha3D was more stable than its predecessor, Alpha3C.

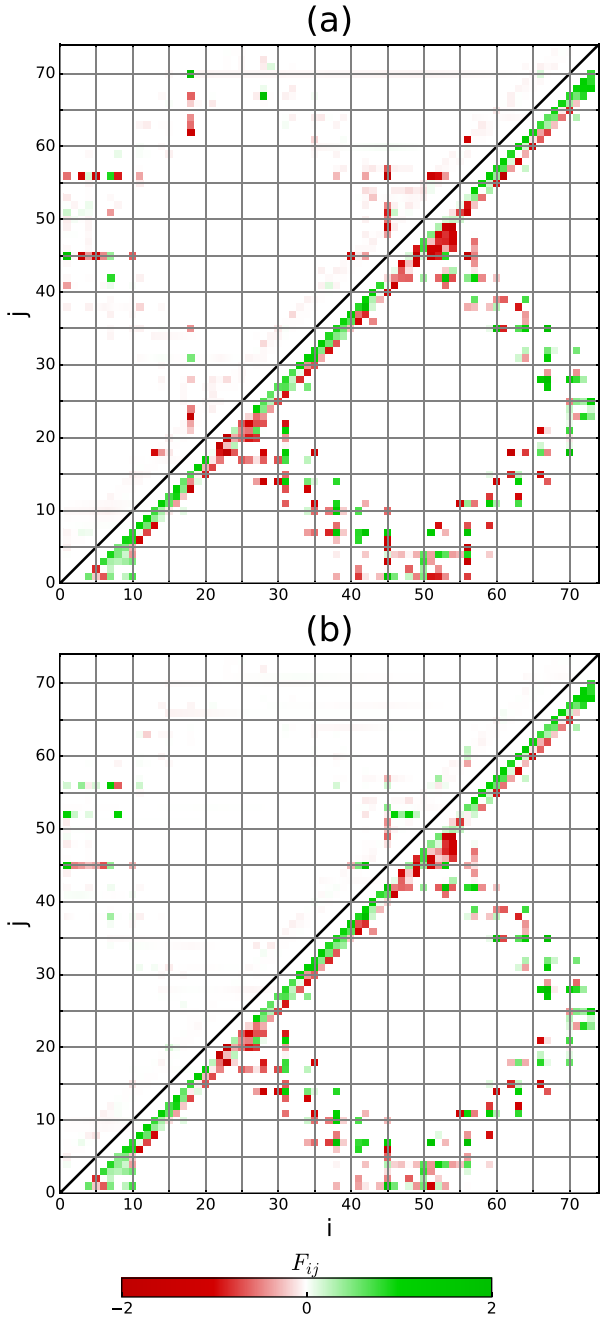


Figure 6. Atomistic configurational frustration for two proposed Alpha3D mutants with the difference (top half) and net result (bottom half). (a) Alpha3D-m1 and (b) Alpha3D-m2 are shown. Despite Alpha3D-m1 being the lowest-energy mutant found, it increases the frustration significantly, whereas Alpha3D-m2 is another low-energy mutant where despite its higher total energy compared with Alpha3D-m1, it has significantly less local frustration.

Figure 2c. Both frustration measures indicate that FiP35 is overall minimally frustrated, consistent with the experimental finding that it is a fast-folding protein. In particular, the conserved W8 residue forms a particularly large number of minimally frustrated interactions. According to the atomistic mutational frustration measure, most of the minimally frustrated interactions in the hydrophobic core involve W8 (Figure 2a,c), explaining the experimental observation of the immutability of W8. Most interactions throughout the structure are minimally frustrated according to the atomistic configurational frustratometer, which is in harmony with a previous simulational analysis[27] that suggested that FiP35 has a relatively unfrustrated landscape with very few non-native interactions that form during folding.

WW domains other than FiP35 often have a tryptophan both at position 8, as does FiP35, and also at position 30, where FiP35 instead has a phenylalanine. W30 is apparently more amenable to mutation than W8.[18] Below, we compare the localized frustration measures of the standard FiP35 sequence to those with either the W8G mutation or the F30W mutation.

Avoiding steric clashes that would otherwise be introduced when generating mutated structures at the all-atom level of detail requires, at a minimum, a repacking of side-chains around the mutated site. In this work, we opted for a relatively simple method for avoiding steric clashes, where we performed

DOI: 10.1021/acs.jpcb.9b01545
J. Phys. Chem. B 2019, 123, 4497−4504

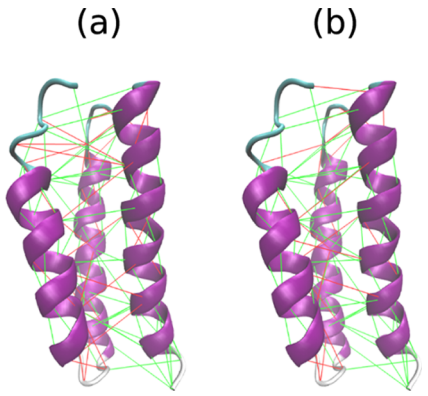


Figure 7. Alpha3D frustratograms corresponding to the F_{ij} values in Figure 6 for (a) Alpha3D-m1 and (b) Alpha3D-m2.

mutations on the native structures and then carried out a local repacking of the side-chains. When studying the mutated variants of FiP35 that are discussed below, we must also account for the other residue pair interactions that are nearby that might be needlessly repacked in this procedure. Therefore, we attempted to isolate the effects of repacking from the effects of performing the mutation itself when redesigning the sequences. To accomplish this, in addition to performing a frustration calculation on the mutated variants (F_{ij}^{mutated}), we also performed a frustration calculation on a repacked structure that did not have the mutation (F_{ij}^{repacked}). We then took the difference between these two results as a measure of frustration in the mutated variant (F_{ij}^{change}). The expression for F_{ij}^{change} is then represented as follows

$$\Delta F_{ij}^{\text{change}} = F_{ij}^{\text{mutated}} - F_{ij}^{\text{repacked}} \tag{12}$$

The resultant frustration value can then be recovered by adding the difference $\Delta F_{ij}^{\text{change}}$ to the original frustration result (either the atomistic mutational or configurational frustratometer)

$$F_{ij}^{\text{mutant}} = \Delta F_{ij}^{\text{change}} + F_{ij}^{\text{unmutated}} \tag{13}$$

Figure 3a,b demonstrate the effects of the W8G mutation for the atomistic mutational and configurational methods, respectively. The changes in frustration are plotted in the top halves of the figures, whereas the net frustration values are shown in the bottom halves. For the W8G mutation, there is a significant increase in the frustration specifically for residue 8. This increase is seen using either the atomistic mutational or the atomistic configurational recipes. The increase in frustration specifically affects the hydrophobic core of FiP35, where any local frustration can disrupt the folding behavior. The W8 residue is buried inside the hydrophobic core of FiP35. Conservative mutations to glycine or alanine are known to cause major disruptions to the stability of the hydrophobic core.[18] Our results for the W8G mutation clearly show that this mutation significantly increases the frustration specifically for residue 8 in the difference plots in the top half of Figure 3a,b. The net result of the mutation is an increase in frustration in the hydrophobic core of FiP35, as seen in the bottom half of Figure 3a. We also note that there is a significant loss of stabilizing interactions, as evident in the frustratograms of Figure 4a,b, which likely is the reason why W8G mutations often result in WW domains that do not fold.[29]

For the F30W mutation shown in Figure 3c,d, the changes are more subtle for the indices computed using either the atomistic mutational or configurational method. Overall, since this residue is far away from the main hydrophobic core, there is a very minimal, but slight, increase to the frustration of the W8 residue. On balance, it appears that the overall stability of the protein remains roughly the same according to the atomistic mutational frustratometer. There are some improvements for a few residue interactions, whereas the main hydrophobic core around W8 still remains overall unfrustrated for the frustratogram in Figure 4c. For the atomistic configurational frustratometer, the overall difference suggests that many interactions will become less frustrated when the protein folds. In general, W30F mutations for the WW domain typically result in stable proteins.[18] The results from the atomistic mutational and configurational frustratometers suggest that a mutation of FiP35's 30th residue back to a tryptophan should result in an overall stable folded state. There is the potential for further improvement to the folding rate as there are a few residue interactions that are made less frustrated with this mutation in the atomistic configurational frustratometer. Since the first two strands of FiP35 generally initiate folding, the changes in frustration here might have very little effect on the overall folding dynamics for FiP35. Clearly though, further mutational studies on FiP35 could potentially yield a faster-folding β-sheet.

Localized Frustration in Alpha3D: Configurational Frustration Leads to Transiently Populated Intermediate States. Alpha3D was one of the first de novo designed protein structures.[21] It has a three-helix bundle structure, with its tertiary structure held together by well-packed hydrophobic side-chains.[20,21] The predecessor to Alpha3D, Alpha3C, differs from Alpha3D by only a few changes. In going from Alpha3C to Alpha3D, two residues, M1 and G2, were added to the N-terminus. Three other mutations were also made: E9Q, S16T, and S65D. These three additional internal mutations, however, are not believed to significantly change the folding behavior of the three-helix bundle, as they are all along the solvent-exposed surface and were added to decrease the sequence similarity among the three helices.[21] The resultant protein, Alpha3D, is a fast folder, which primarily folds by nucleation around the hydrophobic core.[20] Interestingly, despite its high folding rate, Alpha3D does exhibit signs of frustration in its folding mechanism. Several studies show that Alpha3D visits many intermediate states during folding, wherein multiple non-native interactions are formed.[19,20]

Figure 5a,b show frustratograms for the frustration indices found using the atomistic mutational and configurational methods, respectively. The F_{ij} values are shown in Figure 5c. The atomistic configurational method indicates that there is some frustration along the interfaces between the α-helices, which is very evident in the frustratogram of Figure 5b. This indication of frustration is consistent with what was observed in the laboratory and in the simulation, where the nearby hydrophobic residues in the separate α-helices interact strongly with each other. These interactions lead to many frustrated intermediates being populated as the protein folds.[19,20,27] The atomistic mutational frustratometer instead does not indicate frustration along the α-helical interfaces or in the turn region. Instead, it suggests that most of those interactions are unfrustrated in Figure 5a. The α-helical interfaces, which form the hydrophobic core, are instead extremely unfrustrated, demonstrating that the packing of the core was well-

DOI: 10.1021/acs.jpcb.9b01545
J. Phys. Chem. B 2019, 123, 4497−4504

designed.[21] The lack of frustration along the α-helices in the atomistic mutational frustratometer could indicate that the α-helices themselves are actually less frustrated once formed and are relatively stable in the native state. In contrast, frustration around the turn regions is observed in the atomistic configurational frustratometer. This would seem to confirm the idea that Alpha3D primarily folds by nucleation of the hydrophobic core rather than by forming three separate α-helices that then latch together.[20] The high frustration in the turn region seen in the atomistic configurational frustratometer prevents the helices from coming together readily when folding, and instead Alpha3D must rely on the highly stabilizing interactions in the hydrophobic core to drive folding.

As mentioned above, one of the main changes incorporated in Alpha3D following the design of its predecessor, Alpha3C, was the addition of two residues to its N-terminus.[21] Particularly in the atomistic configurational frustratometer, many of the N-terminus interactions are highly unfrustrated, with the residues in the turn region having a high frustration. The additions to the N-terminus seem to have provided more unfrustrated interactions to help with the overall folding of Alpha3D.[19−21,27] To analyze this further, we performed random mutations to Alpha3D in its turn regions to see if a new sequence can be found that would reduce the overall frustration and possibly help speed up the folding process.

To find this redesign, we employed a simple procedure where 10 000 random mutations were first generated for residues in the turn regions of Alpha3D. These were then sorted by energy. Two of the generated mutants with a lower average native-energy than Alpha3D are shown in Figures 6 and 7: Alpha3D-m1 and Alpha3D-m2. Alpha3D-m1 was the lowest-energy mutant found and consists of the four mutations L18A, Y45A, P51S, and L56G from Alpha3D, whereas Alpha3D-m2 consists of three mutations Y45K, E52N, L56K. Despite Alpha3D-m1 being the lowest-energy mutant, it displays a significant increase in frustration for not only the N-terminus but also the loop region. Alpha3D-m2, however, has a similar but slightly higher energy than Alpha3D-m1. Nevertheless, this redesign lowers the frustration involving the key N-terminus as well as parts of the loop region. Given the importance of the N-terminus' interactions with the loop region, this redesign could potentially improve folding by providing less frustrated interactions around the loop region of Alpha3D. The procedure we have just outlined illustrates that minimizing the energy of a design does not necessarily result also in minimizing the frustration. We thus see how pairing energy minimization with a frustratometer can help select a sequence that both stabilizes the molecule and also decreases local frustration, leading to cleaner folding.

CONCLUDING REMARKS

Here, we have presented a way to localize frustration in proteins based on an all-atom model. The atomistic mutational and configurational frustration measures yield complementary results. Although the analyses that we have presented here make use of the Rosetta force field for computing localized frustration, the frustration measures described herein can be similarly calculated using other all-atom force fields as long as they are appropriately solvent-averaged and have a pairwise decomposition. Optimizing localized frustration is a means of automating design against potential traps on folding landscapes. Given the functional importance of locally frustrated sites in proteins, for the purposes of redesigning proteins with known and structurally annotated functions, it may be useful to target only a subset of the frustrated sites: this can be easily done within the framework of frustratometry.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.9b01545.

Frustration calculations, and a comparison of the atomistic frustratometer with the coarse-grained AWSEM frustratometer (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: cecilia@rice.edu.

ORCID

Peter G. Wolynes: 0000-0001-7975-9287

Cecilia Clementi: 0000-0001-9221-2358

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are grateful to D. E. Shaw Research for sharing the protein-folding simulation data. This work was supported by the National Science Foundation (CHE-1265929, CHE-1738990, CHE-1900374, and PHY-1427654), the Welch Foundation (C-1570), and the National Institute of General Medical Sciences (R01-GM44557). Additional support was also provided by the D. R. Bullard-Welch Chair at Rice University (C-0016). Simulations have been performed on the computer clusters of the Center for Research Computing at Rice University, supported in part by the Big-Data Private-Cloud Research Cyberinfrastructure MRI-award (NSF grant CNS-1338099).

REFERENCES

(1) Huang, P.-S.; Boyken, S. E.; Baker, D. The Coming of Age of De Novo Protein Design. *Nature* **2016**, *537*, 320−327.

(2) Davey, J.; Damry, A.; Euler, C.; Goto, N.; Chica, R. Prediction of Stable Globular Proteins Using Negative Design with Non-native Backbone Ensembles. *Structure* **2015**, *23*, 2011−2021.

(3) Jin, W.; Kambara, O.; Sasakawa, H.; Tamura, A.; Takada, S. De Novo Design of Foldable Proteins with Smooth Folding Funnel: Automated Negative Design and Experimental Verification. *Structure* **2003**, *11*, 581−590.

(4) 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.

(5) Wolynes, P. G. Energy Landscapes and Solved Protein-Folding Problems. *Philos. Trans. R. Soc., A* **2005**, *363*, 453−467.

(6) Fleishman, S. J.; Baker, D. Role of the Biomolecular Energy Gap in Protein Design, Structure, and Evolution. *Cell* **2012**, *149*, 262−273.

(7) Shakhnovich, E. I.; Gutin, A. M. Engineering of Stable and Fast-Folding Sequences of Model Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7195−7199.

(8) Ferreiro, D. U.; Hegler, J. A.; Komives, E. A.; Wolynes, P. G. Localizing Frustration in Native Proteins and Protein Assemblies. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 19819−19824.

(9) Ferreiro, D. U.; Hegler, J. A.; Komives, E. A.; Wolynes, P. G. On the Role of Frustration in the Energy Landscapes of Allosteric Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3499−3503.

(10) Ferreiro, D. U.; Komives, E. A.; Wolynes, P. G. Frustration in Biomolecules. *Q. Rev. Biophys.* **2014**, *47*, 285−363.

DOI: 10.1021/acs.jpcb.9b01545
J. Phys. Chem. B 2019, 123, 4497−4504

(11) Parra, R. G.; Schafer, N. P.; Radusky, L. G.; Tsai, M.-Y.; Guzovsky, A. B.; Wolynes, P. G.; Ferreiro, D. U. Protein Frustratometer 2: a Tool to Localize Energetic Frustration in Protein Molecules, Now with Electrostatics. *Nucleic Acids Res.* **2016**, *44*, W356−W360.
(12) Zheng, W.; Schafer, N. P.; Wolynes, P. G. Frustration in the Energy Landscapes of Multidomain Protein Misfolding. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 1680−1685.
(13) Gianni, S.; Camilloni, C.; Giri, R.; Toto, A.; Bonetti, D.; Morrone, A.; Sormanni, P.; Brunori, M.; Vendruscolo, M. Understanding the Frustration Arising from the Competition Between Function, Misfolding, and Aggregation in a Globular Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 14141−14146.
(14) Truong, H. H.; Kim, B. L.; Schafer, N. P.; Wolynes, P. G. Funneling and Frustration in the Energy Landscapes of Some Designed and Simplified Proteins. *J. Chem. Phys.* **2013**, *139*, No. 121908.
(15) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13*, 3031−3048.
(16) Liu, F.; Du, D.; Fuller, A. A.; Davoren, J. E.; Wipf, P.; Kelly, J. W.; Gruebele, M. An Experimental Survey of the Transition Between Two-State and Downhill Protein Folding Scenarios. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2369−2374.
(17) Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulten, K. Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain. *Biophys. J.* **2008**, *94*, L75−L77.
(18) Davis, C. M.; Dyer, R. B. WW Domain Folding Complexity Revealed by Infrared Spectroscopy. *Biochemistry* **2014**, *53*, 5476−5484.
(19) 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.
(20) Zhu, Y.; Alonso, D. O. V.; Maki, K.; Huang, C.-Y.; Lahr, S. J.; Daggett, V.; Roder, H.; DeGrado, W. F.; Gai, F. Ultrafast Folding of α3D: A De Novo Designed Three-Helix Bundle Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15486−15491.
(21) Walsh, S. T. R.; Cheng, H.; Bryson, J. W.; Roder, H.; DeGrado, W. F. Solution Structure and Dynamics of a De Novo Designed Three-Helix Bundle Protein. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 5486−5491.
(22) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517−520.
(23) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341−346.
(24) Boninsegna, L.; Gobbo, G.; Noé, F.; Clementi, C. Investigating Molecular Kinetics by Variationally Optimized Diffusion Maps. *J. Chem. Theory Comput.* **2015**, *11*, 5947−5960.
(25) Shehu, A.; Kavraki, L. E.; Clementi, C. On the Characterization of Protein Native State Ensembles. *Biophys. J.* **2007**, *92*, 1503−1511.
(26) Piana, S.; Sarkar, K.; Lindorff-Larsen, K.; Guo, M.; Gruebele, M.; Shaw, D. E. Computational Design and Experimental Testing of the Fastest-Folding β-Sheet Protein. *J. Mol. Biol.* **2011**, *405*, 43−48.
(27) Best, R. B.; Hummer, G.; Eaton, W. A. Native Contacts Determine Protein Folding Mechanisms in Atomistic Simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 17874−17879.
(28) Chen, J.; Chen, J.; Pinamonti, G.; Clementi, C. Learning Effective Molecular Models from Experimental Observables. *J. Chem. Theory Comput.* **2018**, *14*, 3849−3858.
(29) Petrovich, M.; Jonsson, A. L.; Ferguson, N.; Daggett, V.; Fersht, A. R. Φ-Analysis at the Experimental Limits: Mechanism of β-Hairpin Formation. *J. Mol. Biol.* **2006**, *360*, 865−881.

DOI: 10.1021/acs.jpcb.9b01545
J. Phys. Chem. B 2019, 123, 4497−4504