Article

Nascent β-Hairpin Formation of a Natively Unfolded Peptide Reveals the Role of Hydrophobic Contacts

Wei Chen,[1] Chuanyin Shi,[2] and Jana Shen[1,*]

[1]Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland; and [2]Eastern Hepatobiliary Surgery Hospital, Second Military Medical University, Shanghai, China

ABSTRACT   Despite the important role of the unfolded states in protein stability, folding, and aggregation, they remain poorly understood due to the lack of residue-specific experimental data. Here, we explore features of the unfolded state of the NTL9 protein by applying all-atom replica-exchange simulations to the two fragment peptides NTL9(1–22) and NTL9(6–17). We found that while NTL9(6–17) is unstructured, NTL9(1–22) transiently folds as various β-hairpins, a fraction of which contain a native β-sheet. Interestingly, despite a large number of charged residues, the formation of backbone hydrogen bonds is concomitant with hydrophobic but not electrostatic contacts. Although the fragment peptides lack a proposed specific contact between Asp^8 and Lys^{12}, the individually weak, nonspecific interactions with lysines together stabilize the charged Asp^8, leading to a pK_a shift of nearly 0.5 units, in agreement with the NMR data. Taken together, our data suggest that the unfolded state of NTL9 likely contains a β-hairpin in segment 1–22 with sequence-distant hydrophobic contacts, thus lending support to a long-standing hypothesis that the unfolded states of proteins exhibit native-like topology with hydrophobic clusters.

INTRODUCTION

In recent years, unfolded proteins have attracted much attention. They include the so-called intrinsically disordered proteins, which lack stable tertiary or secondary structures under physiological conditions but play critical roles in biological processes (1), including disease mechanisms (2,3). The unfolded states of natively folded proteins are also important to understand, as they play a key role in initiating folding and aggregation and serve as reference states for thermodynamic stability of folded proteins (4). We note that the unfolded states are also referred to by some workers as denatured states, since historically they were obtained using chemical denaturants or acids. The unfolded states have been traditionally thought of as random coils without preferential interactions (5,6); however, an increasing number of experimental studies point to the presence of residual structures such as the native-like topology, local and long-range order (7–9), as well as specific hydrophobic (10–12) and electrostatic interactions (13–15).

One of the best-studied proteins with regard to the unfolded state is the N-terminal domain of ribosomal protein L9 (NTL9), comprising 56 residues with a topology of $\beta1$-loop-$\beta2$-$\alpha1$-$\beta3$-$\alpha2$ (see Fig. 1). The minimum folding unit includes the first 39 residues and lacks the second helix (14). Thus, adding the third β-strand along with the intervening helix stabilizes the β-hairpin, $\beta1$-loop-$\beta2$, which is unstructured on its own (14). This β-hairpin, corresponding to residues 1–22, is the focus of this article.

Based on the mutagenesis data and measurements of pK_a, as well as pH-dependent stabilities of NTL9 and its fragment peptides, Raleigh and co-workers proposed the existence of a specific electrostatic contact between Asp^8 and Lys^{12} in the unfolded state of NTL9 (13,15,16). This is a nonnative contact, as the two side chains are >13.4 Å apart in the crystal structure of NTL9 (see Fig. 1, distance between Cγ of Asp^8 and Nζ of Lys^{12}). To our knowledge, the data from the Raleigh lab offered the first and only experimental evidence for specific electrostatic interactions in the unfolded state. A subsequent kinetic study also led Raleigh and colleagues to hypothesize that the nonnative interaction between Asp^8 and Lys^{12} persists in the folding-transition state (17). We note that residual electrostatic interactions and perturbed pK_as in the unfolded state had been (indirectly) observed by the Fersht lab two decades ago (18–20).

Considering the experimental challenges associated with the sparse population of the unfolded states under physiological conditions, Raleigh and co-workers used the constituent fragment peptides that are natively unfolded to model the unfolded state of the intact protein. Circular dichroism and NMR-based pK_a measurements were conducted on several fragment peptides of NTL9, including residues 1–11 and 1–22. It was found that even though both fragments are unstructured, the pK_a of Asp^8 is shifted lower than the model value (4.0) by 0.16 and 0.34 units, respectively (13,14). Interestingly, these shifts are slightly smaller than the pK_a of Asp^8 in a destabilizing mutant, V3AI4A, which was estimated to be 0.44 units (21). The observation that the pK_a of Asp^8 is slightly depressed in NTL9(1–11)

CrossMark

suggests the presence of other attractive interactions in addition to Lys[12], and raises the question about specificity of electrostatic interactions in the unfolded state. Most recently, Raleigh and co-workers conducted a study of the unfolded NTL9 using paramagnetic relaxation enhancement experiments in urea (22). The data indicate a lack of secondary structure but the presence of long-range contacts. To our knowledge, this is the only experimental study that has directly probed the structural features of the unfolded state of NTL9.

To directly probe the unfolded-state interactions in NTL9 and test the hypothesis regarding the nonnative interaction between Asp[8] and Lys[12], one of us performed generalized-Born (GB)-based constant pH molecular dynamics simulations (23) of the unfolded state of NTL9 and the mutant K12M (24). These simulations revealed a significantly depressed pK_a value of Asp[8] in the unfolded state, corroborating the experimental data (15) and a theoretical calculation derived from the experimental stability and pK_a data of the wild-type and mutant proteins (25). As to the origin of the depressed pK_a of Asp[8] in the unfolded state, the above simulation work showed that Asp[8] strongly interacts not only with Lys[12] but also with Lys[14] (24). Furthermore, Lys[12] also forms a nonnative interaction with Glu[17] (24). In another work using coarse-grained molecular dynamics simulations, Azia and Levy found that removal of the charge on Lys[12] or Lys[19] via a mutation to methionine eliminates the repulsive interaction between the two and reduces entropy of the unfolded state of NTL9 (26). Pande and co-workers successfully folded NTL9(1–39) using GB-based molecular dynamics simulations (27). They found that the unfolded state is compact and contains some residual native-like structures, with 20% of native contacts in the β_{12} hairpin, the formation of which is rate limiting (27). The latter fact was attributed to the large number of lysines and glycines in the loop region (27). Recently, Pappu and co-workers carried out a study using Monte Carlo simulations in implicit solvent at high temperature to examine the unfolded state of NTL9 (22). They observed low-likelihood long-range hydrophobic contacts, consistent with the NMR relaxation data from the Raleigh lab (22). These studies provided microscopic views of the unfolded state of NTL9; however, due to the known artifacts/limitations of the implicit-solvent and coarse-grained models, the insights are limited.

In this work, we seek to complement the experimental studies and further explore the residual structures in the unfolded state of NTL9. To make the study amenable to current computational resources, we performed 100-ns explicit-solvent molecular dynamics with temperature replica exchange on the two fragment peptides NTL9(1–22) and NTL9(6–17), which correspond to the β_{12} hairpin and its loop region, respectively, in the intact NTL9 (see Fig. 1). Simulations of these two fragments would allow us to test the aforementioned hypotheses regarding specific electro-static interactions in the unfolded state of NTL9, and address other questions, such as the role of hydrophobic contacts and the existence of native-like topology, which have been lingering in the community for more than a decade but remain somewhat unclear (7–10). Explicit-solvent molecular dynamics is the method of choice for studying conformational dynamics of biomolecules; it is not subject to the deficiencies of the implicit-solvent models such as the popular GB models, which generally overstabilize attractive electrostatic interactions, although recent developments are yielding encouraging results (28). Due to the short length of the two peptides, sampling limitation is unlikely to occur, which is another advantage compared with previous work (24). To avoid potential bias due to the use of a particular energy function, we employed three leading force fields, AMBER ff99SB-ILDN, CHARMM22 (C22)/CMAP, and CHARMM36 (C36). Simulation with the latter was repeated twice to verify convergence.

The article is organized as follows. After the Materials and Methods section, we analyze the residual structures for the two fragments, then investigate hydrophobic contacts and ionic interactions. Furthermore, we calculate the pK_a values of acidic residues based on trajectory snapshots and compare them with experiment. Our data reveal nascent formation of various β-hairpins and the correlation with sequence-distant hydrophobic contacts. Interestingly, in the absence of specific electrostatic interactions, nonspecific attractive interactions together stabilize the charged Asp[8], leading to a depressed pK_a, in agreement with experiment. Thus, our study complements the existing experimental data, greatly extending the microscopic insights into the structural features of the unfolded state of NTL9.

MATERIALS AND METHODS

Simulation protocol

All structure preparation and simulations were performed with the GROMACS package (version 4.5) (29). Three different force fields, AMBER ff99SB-ILDN (30), CHARMM C22/CMAP (31,32), and C36 (33), were used to represent the fragment peptides. The CHARMM modified TIP3P model (34) was used to represent water. NTL9(1–22) and NTL9(6–17), acetylated at the N-terminus (ACE) and N-methylamidated at the C-terminus (CT3), were built as extended chains using CHARMM (35). They were solvated in a truncated octahedral water box with a minimum distance of 10 Å between the peptide and the edge of the box. All acidic (Asp and Glu) and basic (Lys) side chains were in the charged form. Five and three chloride counterions were added to neutralize the systems of NTL9(1–22) and NTL9(6–17), respectively. Energy minimization with the steepest-descent algorithm was performed until the maximum force was <1000 kJ/mol/nm. Then, a 20-ps equilibration step was performed under constant NVT conditions at 300 K with 1000 kJ/mol/nm² restraint on the heavy atoms of the peptide, followed by a 2-ns unrestrained equilibration step conducted under constant NPT conditions at 400 K and 1 atm pressure. The temperature was controlled by the velocity rescaling algorithm (36) with a coupling constant of 0.1 ps, whereas the pressure was controlled by the Parrinello-Rahman method (37) with a coupling constant of 2 ps. To reduce the computational cost, a moderately collapsed structure from the 2-ns trajectory was selected to be resolvated and

equilibrated using 20 ps NVT and 100 ps NPT molecular dynamics at 300 K and 1 atm pressure.

The final snapshots from the above equilibration runs were used as the starting structures for the production runs. The particle-mesh Ewald method was used for treating long-range electrostatics. A spherical cutoff of 9 Å was applied to the calculation of van der Waals energies and forces. The bonds involving hydrogen atoms were constrained using the LINCS algorithm (38). A periodic boundary condition and 2-fs time step were used. The temperature replica-exchange (REX) protocol was applied with temperatures exponentially spaced between 280 K and 450 K (55 and 45 replicas for NTL9(1–22) and NTL9(6–17), respectively). The replicas adjacent in temperature were allowed to exchange every 2 ps based on the Metropolis criterion. The resulting exchange acceptance ratio was ~10–20% for adjacent replica pairs. The simulation lasted 100 ns for each replica, resulting in an aggregated simulation time of 5.5 and 4.5 μs for NTL9(1–22) and NTL9(6–17), respectively. Unless otherwise indicated, the replica with the temperature closest to 298 K was used for analysis. To test convergence, the REX simulation for NTL9(1–22) with the C36 force field was repeated twice, starting from independent random structures. Unless otherwise stated, the data with C36 are the average of the three independent runs.

Simulations of the folded, intact NTL9 protein were also performed using the same three force fields, based on the crystal structure (Protein Data Bank (PDB) ID 2HBB). The missing hydrogen atoms were added and all acidic and lysine residues were in the charged form. The energy minimization, equilibration steps, and simulation setting were identical to those used for the fragment peptides. The production run, however, was conducted under constant NPT conditions at 300 K and 1 atm pressure for 20 ns with each of the three force fields.

Data analysis

Unless otherwise stated, the analysis was performed using GROMACS (29) or VMD (39) based on the last 80 ns (per replica) of data. The DSSP algorithm was used to identify secondary structures (40). For clustering analysis, we first extracted all structures containing β-sheets from the simulations and then performed K-mean clustering based on the Cα RMSD with a cutoff of 3.5 Å using the MMTSB tool set (41). For defining backbone hydrogen bonds, the cutoff for the donor-receptor distance is 3.5 Å between heavy atoms and the cutoff for the donor-hydrogen-acceptor angle is 135°. For analysis of hydrophobic interactions, two side chains are defined as in contact if the minimal distance between any pair of heavy atoms is within 4.5 Å.

We employed the PROPKA program (version 3.0) (42) for the calculation of pK_a values for the fragment peptides and intact NTL9. PROPKA is an empirical method for rapid calculation of protein pK_a values based on a single structure. We chose the method because of its speed and high accuracy for predicting pK_a values of solvent-exposed side chains with small pK_a shifts. To account for conformational flexibility, we first calculated the microscopic pK_a values using the snapshots extracted from the simulations at intervals of 20 ps for the fragment peptides and 10 ps for the native state. These pK_a values were then combined to yield a macroscopic pK_a value using the equation

$$pK_a = log\left(\frac{1}{n}\sum_{i=1}^{n} 10^{pK_a^i}\right), \quad (1)$$

where n is the total number of conformers. Equation 1 can be derived as follows. For each conformer i, the Henderson-Hasselbalch equation relates the microscopic pK_a^i to the concentrations of the deprotonated and protonated species at certain pH,

$$pH - pK_a^i = log\frac{[A^-]_i}{[AH]_i}. \quad (2)$$

Using Eq. 2, the macroscopic pK_a can be related to the microscopic pK_a^i as

$$\begin{aligned}
pH - pK_a &= log\frac{\sum_{i=1}^{n} [A^-]_i}{\sum_{i=1}^{n} [AH]_i} \\
&= log\frac{\sum_{i=1}^{n} \frac{10^{pH-pK_a^i}}{1 + 10^{pH-pK_a^i}} \times F_i}{\sum_{i=1}^{n} \frac{1}{1 + 10^{pH-pK_a^i}} \times F_i},
\end{aligned} \quad (3)$$

where F_i is the population fraction of conformer i. Assuming a pH that is much greater than pK_a and replacing F_i with $1/n$, pH can be eliminated and the above equation becomes Eq. 1.

RESULTS AND DISCUSSION

Nascent formation of β-hairpin states

Explicit-solvent REX molecular dynamics simulations of 100 ns were performed for the two fragment peptides, NTL9(1–22) and NTL(6–17) (Fig. 1). The replicas at 298 K were used for analysis. We first examine the residual structures sampled by the two peptides. According to DSSP analysis (40), NTL9(6–17) remains unstructured, whereas the longer fragment, NTL9(1–22), transiently forms various β-sheets in all simulations (Fig. S1). The β-sheet states are present for 60%, 30%, and 36% of the time with the ff99SB-ILDN, C22, and C36 force fields, respectively, whereas for the rest of the time, the peptide samples unstructured states and little α-helix states. The higher β content in the ff99SB-ILDN simulation is consistent with the known β bias of the ff99SB force fields (43). A detailed comparison of the conformational states sampled by the three force fields is given in another study (50). Almost all (>97%) of the β-sheets are two-stranded, among which the majority (96%,
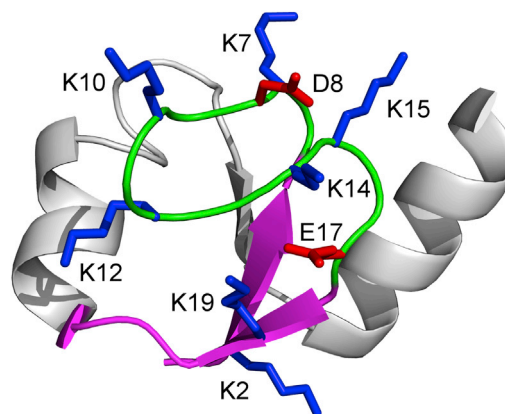

FIGURE 1 Crystal structure of NTL9 (PDB ID 2HBB). The fragment 1–22 (sequence MKVIFLKDVKGKGKKGEIKNVA) contains two β-strands (*magenta*) connected by a loop (residues 6–17, *green*). Two acidic and seven basic side chains are shown in red and blue, respectively. To see this figure in color, go online.

77%, and 96% for the ff99SB-ILDN, C22, and C36 force fields, respectively) are antiparallel β-hairpins. These β-hairpins are highly heterogeneous in terms of the lengths of the β-sheet and their registries. Most β-sheets are two to five residues long, with a very small fraction (<10%) having a length of six to eight residues (Fig. 2). With the AMBER force field, the sheets with two to five residues have nearly equal probabilities, whereas with the CHARMM force fields, especially C36, the shorter sheets dominate. To test convergence, the simulation with C36 was repeated twice, starting from independent random structures. Consistently, β-hairpin states were formed in all three runs. The standard deviation of the probabilities among the three runs is small (Fig. 2), demonstrating the statistical significance of the results. We note that despite the significant population of various β states, the macroscopic state of NTL9(1–22) is unstructured according to the NMR and circular dichroism data (14) as well as our NMR shift calculations based on the simulations (50).

To further examine the nascent folding of NTL9(1–22), K-mean clustering analysis with a Cα RMSD cutoff of 3.5 Å was carried out on the trajectory snapshots containing a β-sheet. A large number of clusters were obtained, with the largest cluster having a population of 12%, 18%, and 8% with the ff99SB-ILDN, C22, and C36 force fields, respectively (Fig. S2), indicating the diversity of the conformational ensemble, consistent with the overall picture of a largely random-coil structure with nascent formation of a β-hairpin. Fig. 3 displays the representative structures of the largest five clusters, which are various β-hairpins, except for one parallel β-sheet, the representative structure of the first cluster from the simulation with C22. This cluster also contains three-stranded mixed antiparallel/parallel β-sheets.
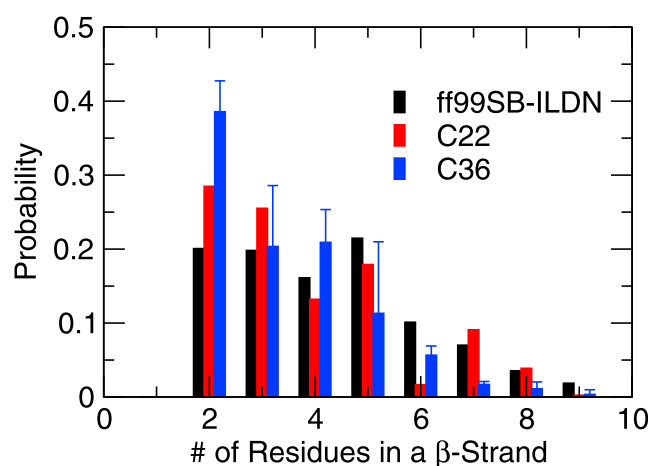

FIGURE 3 Representative folded states of NTL9(1–22) from the three simulations. The largest five clusters of the β-sheet structures are shown. The percentage of each cluster is indicated. The N- and C-termini are indicated by the blue and red spheres, respectively. To see this figure in color, go online.

The β-hairpin states formed by NTL9(1–22) are different from the corresponding native structure in the intact NTL9. The distribution of the Cα RMSD is centered in the range 6.5–7.5 Å (Fig. S3). However, if only segments 2–4 and 18–20 are considered, corresponding to the native β-sheet region in the intact protein, the RMSD distribution is almost flat, extending to the region of 1 Å (Fig. S3). This indicates the presence of a small fraction of structures that are very much like the native structure. An example is given in Fig. 4, which shows the representative structure of cluster 7 from the C36 simulation. There is a perfect alignment with the crystal structure in the β-sheet region (RMSD of 0.58 Å), whereas the deviation in the loop region is large. Remarkably, this structure contains several native hydrophobic contacts/clusters, Val[3]-Val[21]-Phe[5] and Ile[4]-Ile[18] (Fig. 4), which appear to stabilize the β-sheet.

Correlation between hydrophobic contacts and β-sheet formation

To understand the potential role of hydrophobic interactions in the nascent folding of NTL9(1–22), we contrast the hydrophobic contact maps in the presence and absence of the β-sheet formation. As seen from Fig. 5, many sequence-local contacts are formed in the N-terminal region

FIGURE 2 Probability distribution of the length of β-sheets in the simulations of NTL9(1–22). The longest β-strand was used in the calculation. The results with C36 are shown with error bars, which were calculated using the standard deviation among the three independent runs. To see this figure in color, go online.
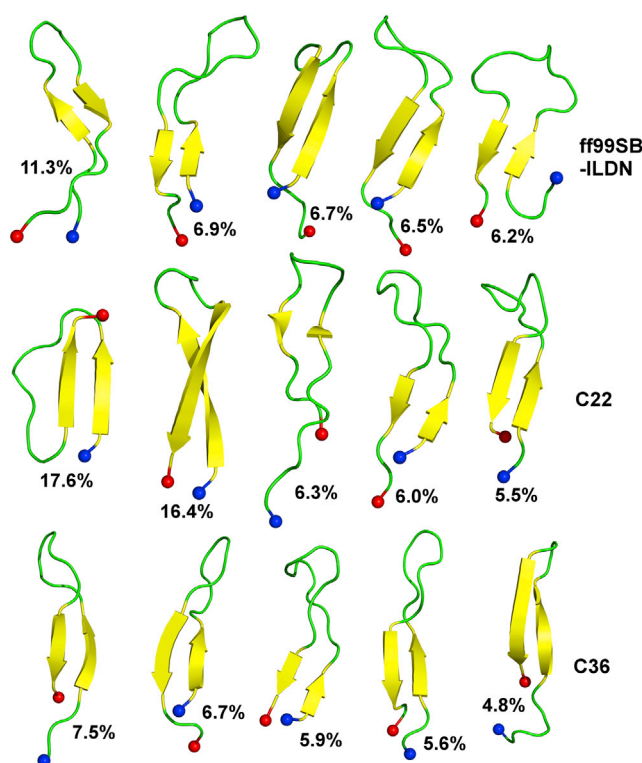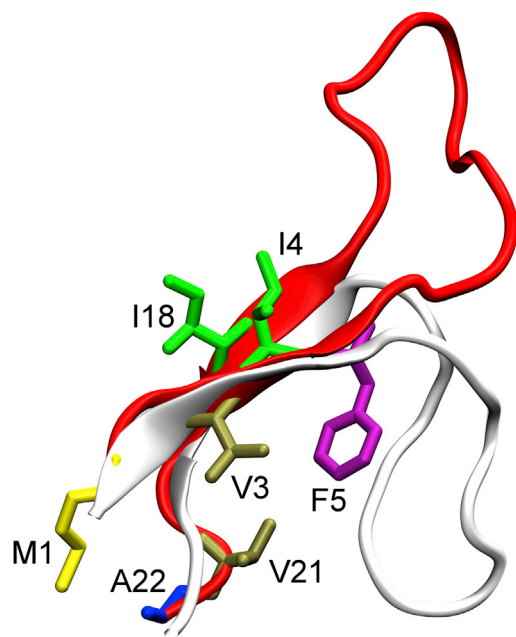
FIGURE 4   A sample snapshot with the native-like β-sheet, showing the representative structure of cluster 7 from the simulation of NTL9(1–22) with the C36 force field. Side chains making hydrophobic contacts are shown in stick representation. The crystal structure is displayed in silver. Alignment was made based on the Cα atoms of residues 2–4 and 18–20 (β-sheet region in the crystal structure), giving a Cα RMSD of 0.58 Å for this region. To see this figure in color, go online.

regardless of the β-sheet formation. However, sequence-distant hydrophobic contacts mainly occur in the β-hairpin states, which is particularly evident in the simulations with ff99SB-ILDN and C22. Among them, three are native and common for all force fields: Met[1]-Val[21], Val[3]-Val[21], and Phe[5]-Val[21], whereas three are nonnative and common for all force fields: Met[1]-Ile[18], Phe[5]-Ile[18], and Leu[6]-Ile[18]. The more frequent occurence of sequence-distant hydrophobic contacts in the β-hairpin states is in agreement with the representative structure shown in Fig. 4 and suggests a correlation between the two. We note that hydrophobic contacts

in the unfolded state of NTL9 were also seen in a recent simulation study based on an implicit-solvent model (22).

To further examine the relationship between hydrophobic contacts and β-sheet formation, we plot the probability distribution of the total number of backbone hydrogen bonds in the presence of 0, 1, 2, or 3 hydrophobic contacts, as the former is a proxy for β-sheet formation. As shown in Fig. 6, the distribution for zero hydrophobic contacts has a maximum at 2, 0, and 0 backbone hydrogen bonds in the ff99SB-ILDN, C22 and C36 simulations, respectively. Upon forming one hydrophobic contact, the peak of the distribution is shifted to the larger number of hydrogen bonds (5, 2, and 2, respectively). In the presence of two hydrophobic contacts, the peak of the distribution is further shifted toward a larger number of hydrogen bonds, with the degree of shift dependent on the particular force field. These data suggest that the formation of hydrophobic contacts is correlated with the formation of backbone hydrogen bonds. To further test this hypothesis, we obtained the free-energy surface as a function of the number of backbone hydrogen bonds and the solvent-accessible surface area (SASA) of Ile[18] and Val[21], which appear to form the most stable hydrophobic contacts (Figs. 4 and 5). Remarkably, the free-energy surface shows two local free-energy minima, one located at zero hydrogen bonds and large SASA, and one located at a larger number of hydrogen bonds but smaller SASA (Figs. 7 and S4). This anticorrelation between the solvent exposure of hydrophobic residues and formation of backbone hydrogen bonds provides additional evidence for the role of hydrophobic contacts in the folding of β-sheets.

Individual ion-pair interactions are weak and not correlated with the formation of β-sheets

An interesting question is whether electrostatic interactions also promote β-sheets. To address this question, we analyzed the probability distributions of the distances between the two acid side chains, Asp[8] and Glu[17], and any
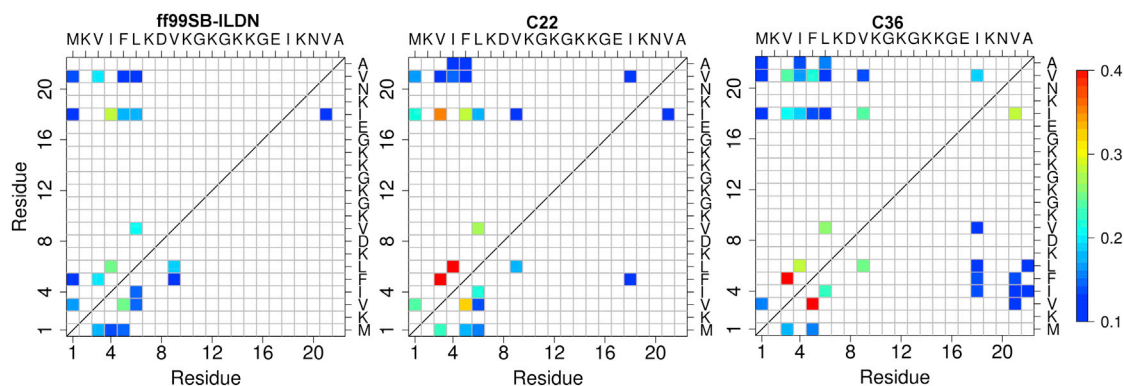

FIGURE 5   Correlation of side-chain hydrophobic contacts and formation of β-sheet in NTL9(1–22). The contacts in the upper and lower triangles are based on the β-sheet and unstructured states, respectively. The color scale indicates the fraction of time for a hydrophobic contact. To see this figure in color, go online.
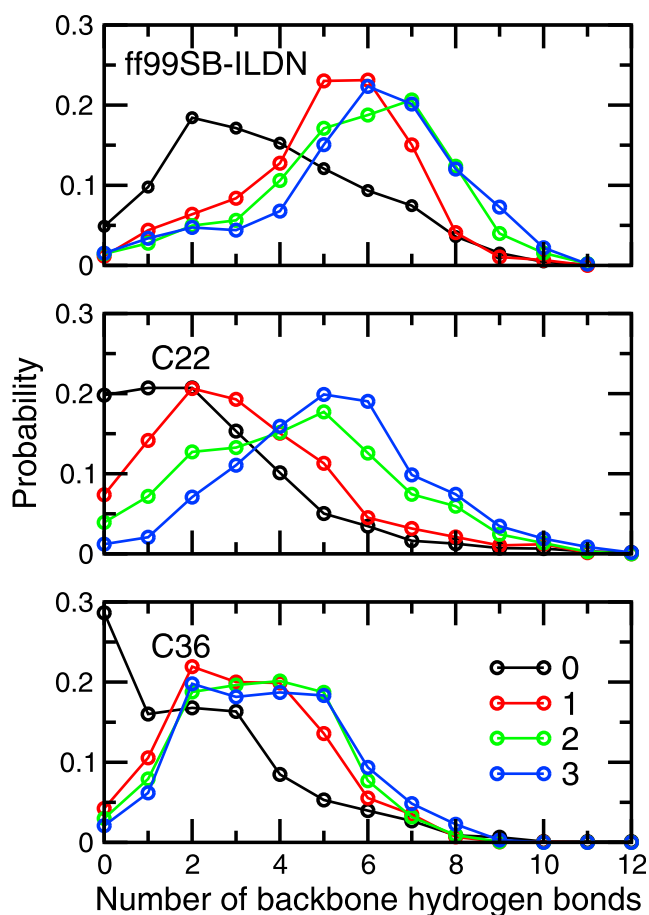
FIGURE 6 Correlation between the hydrophobic side-chain contacts and formation of backbone hydrogen bonds in the simulations of NTL9(1–22). The probability distribution for forming 0–12 backbone hydrogen bonds is plotted in the presence of 0, 1, 2, or 3 hydrophobic contacts. To see this figure in color, go online.

into β-sheet and unstructured states. Several observations can be made. First, all interactions are weak. The probability for the distance to be within 5 Å (defined here as an ion contact) is generally <10%, whereas the probability for the distance to be within 8 Å is no more than 30%. Second, no significant differences are found between the distributions for NTL9(1–22) and NTL9(6–17), suggesting that the presence of the β-sheet segments (residues 1–5 and 18–22) has no influence on the electrostatic interactions. Also, the distributions for the β-hairpin and unstructured states of NTL9(1–22) are similar, further suggesting the lack of correlation between electrostatic interactions and β-sheet formation.

One exception to the weak sequence-distant electrostatic interactions is the attraction between Asp[8] and Lys[19] in the simulations with C36. It is much stronger in the β-hairpin states (probability of 29% within 5 Å) than the unstructured states (probability of 9%). Interestingly, their sequence neighbors, Val[9] and Ile[18], form a persistent hydrophobic contact (occupancy of 24%), as well as a backbone hydrogen bond (occupancy of 27%). Thus, we speculated a possible correlation between the side-chain-to-side-chain distance Asp[8]-Lys[19], and the backbone amide-to-carbonyl distance Val[9]-Ile[18]. Somewhat surprisingly, the correlation is weak (Fig. S7). By contrast, the hydrophobic contact between Val[9] and Ile[18] coexists with their backbone hydrogen bonding (Fig. S7), confirming the conclusion made earlier that hydrophobic interactions facilitate β-sheet formation.

Previous experimental and computational studies suggest that Asp[8] is involved in significant electrostatic interactions in the unfolded state of NTL9 and may form a specific interaction with Lys[12] (15,16,24). In our simulations of the two fragments, the Asp[8]-Lys[12] contact is not stronger than the other ion pairs (Fig. S5), nor is the contact formation strengthened by the presence of the β-sheet (Fig. S5). Thus, the hypothesized contact may require stabilization by residues beyond sequence 1–22.

lysine separated by at least two amino acids (sequence distant) from simulations of the two fragment peptides (Figs. S5 and S6). The analysis of NTL9(1–22) is separated
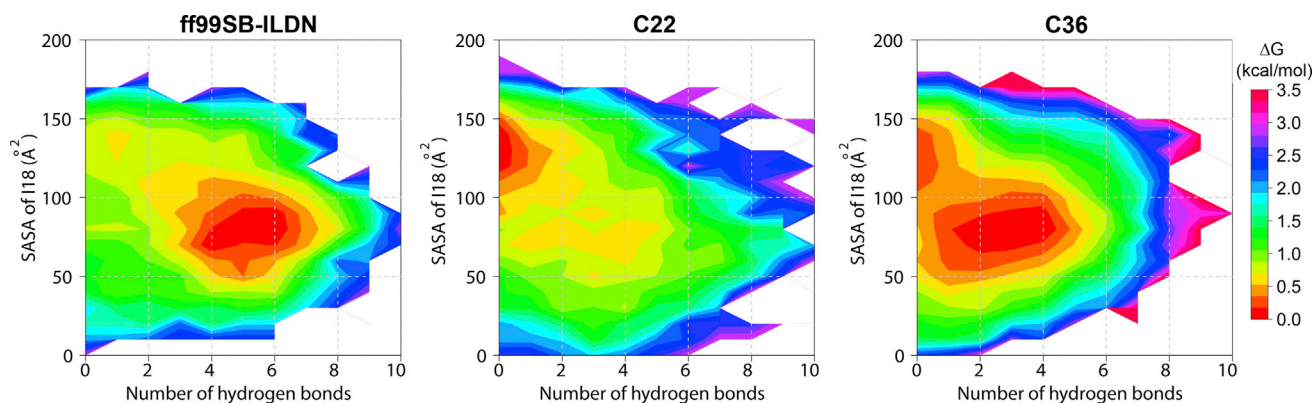

FIGURE 7 Free-energy surface as a function of the SASA of side chain Ile[18] and the number of backbone hydrogen bonds in the simulations of NTL9(1–22). A probe radius of 1.4 Å was used in the SASA calculation. A bin of 10 Å² was used in constructing the two-dimensional histogram. To see this figure in color, go online.

Depressed pK_a of Asp^8 is due to nonspecific electrostatic interactions

To further elucidate the role of Asp^8 and validate the electrostatic interactions observed in the simulations of NTL9(1–22), we calculated the pK_a values and compared them with the NMR titration data (14). The empirical program PROPKA (42) was used to afford speed. As a control, we first compare the calculated pK_a values of the intact NTL9 with experiment. This calculation was done in two ways: based on a single crystal structure or trajectory snapshots. As shown in Table 1, with the single crystal structure, the overall average absolute deviation from experiment is 0.53. We focus on the acidic residues involved in strong electrostatic attraction, i.e., Asp^8, Glu^{17}, and Asp^{23}. Their pK_a values are underestimated by ~0.7, 0.3, and 0.8 units, respectively. With the trajectory snapshots, these pK_a values are all increased, resulting in an average absolute deviation of ~0.3 for all three force fields. This improvement is due to the incorporation of conformational flexibility.

Differences among the force fields are worth noting. For Asp^8, although ff99SB-ILDN and C22 give the same pK_a value as experiment, C36 gives a lower pK_a by 0.3 units. For Glu^{17}, the pK_a values based on the three force fields are 0.1 units different from each other and are higher than the experimental value by 0.2–0.4 units. For Asp^{23}, the pK_a based on ff99SB-ILDN is too high by 0.4 units, whereas that based on C22 or C36 is too low by ~0.3 units. These data suggest that, given accurate conformational states, the

calculation error is about ± 0.3 units from the experimental value. For acidic residues involved in strong attractive interactions, CHARMM force fields tend to give lower pK_a values compared to the ff99SB-ILDN force field, which is not surprising, given the slightly stronger electrostatic interactions (Figs. S5 and S6).

The calculated pK_a values of the two fragments are examined next. Interestingly, the pK_a values of Asp^8 and Glu^{17} are nearly the same for both fragments, which supports our earlier conclusion that ionic interactions are not correlated with β-sheet formation. The pK_a of Asp^8 is almost identical to the experimental value, representing a pK_a downshift of ~0.4 unit relative to the model. The pK_a of Glu^{17} is similar to the model value, and ~0.4 higher than the experimental value. Similar overestimation is also seen for Glu^{17} in the intact protein and thus may be attributed to a systematic error. However, the pK_a downshift of 0.4 for Asp^8 is worth noting. As discussed earlier, the electrostatic attraction between Asp^8 and individual lysines (Lys^2/Lys^7/Lys^{10}/Lys^{12}/Lys^{14}/Lys^{15}/Lys^{19}) is weak, but the sum of these interactions is significant, which amounts to a total probability of 40% or higher for at least one lysine to be within a radius of 5 Å (Fig. S5). Thus, the depressed pK_a of Asp^8 in the fragment peptides is the result of nonspecific interactions with several lysines, i.e., seven for NTL9(1–22) and five for NTL9(6–17).

CONCLUSIONS

In this study, we probed the residual structures in the unfolded state of NTL9 using two fragment peptides, corresponding to the β_{12} hairpin and the loop, respectively. Whereas NTL9(6–17) is unstructured, NTL9(1–22) transiently samples diverse β-hairpins, a fraction of which bear strong resemblance in the β-sheet region to the β_{12} hairpin of the intact NTL9. This finding suggests that the β-hairpin is at least partially formed in the unfolded state of NTL9, in agreement with the GB-based folding simulation of NTL9 (27). The fact that the β-hairpin formation in the fragment peptide has a low probability and a higher degree of nonnativeness compared to the unfolded, intact NTL9 (27) may be attributed to the lack of stabilization by tertiary contacts with other parts of the protein. A similar observation has been made for partially folded helices in the fragment peptides of HP36 (44).

Our data showed that the formation of backbone hydrogen bonds is positively correlated with the formation of sequence-distant hydrophobic contacts and the resulting sequestration from solvent, which corroborates the results of several studies of the natively folded β-hairpins (45–47), strengthening the view of the cooperativity between backbone hydrogen bonding and hydrophobic contacts in β-sheet folding. Thus, our data support the notion that folding initiation involves the simultaneous minimization of solvent exposure of hydrophobic residues and formation of secondary structure (48).

TABLE 1 Calculated and experimental pK_a values of acidic residues in NTL9 and the two fragment peptides

Residue	Crystal[a]	ff99SB-ILDN	C22	C36	Experiment[b]
NTL9					
Asp^8	2.28	2.95 (0.22)	3.03 (0.18)	2.66 (0.03)	2.99 (0.05)
Glu^{17}	3.31	3.85 (0.10)	3.95 (0.26)	3.75 (0.12)	3.57 (0.05)
Asp^{23}	2.20	3.49 (0.01)	2.74 (0.02)	2.79 (0.02)	3.05 (0.04)
Glu^{38}	4.70	4.70 (0.01)	4.70 (0.01)	4.68 (0.02)	4.04 (0.05)
Glu^{48}	4.39	4.30 (0.08)	4.28 (0.08)	4.21 (0.07)	4.21 (0.08)
Avg. abs. dev.	0.53	0.30	0.29	0.28	
NTL9(1–22)					
Asp^8		3.57 (0.02)	3.56 (0.02)	3.55 (0.03)	3.66 (0.04)
Glu^{17}		4.38 (0.01)	4.24 (0.01)	4.34 (0.02)	4.05 (0.03)
NTL9(6–17)					
Asp^8		3.59 (0.02)	3.59 (0.01)	3.58 (0.03)	–
Glu^{17}		4.49 (0.01)	4.36 (0.01)	4.44 (0.01)	–

Average pK_a values were obtained from PROPKA calculations (42) based on the trajectory snapshots. The pK_a values of NTL9(1–22) with C36 are the averages over three simulations. Values in parentheses are the standard errors calculated using block analysis. For simulations of the intact NTL9 and peptides, the block lengths are 4.5 and 10 ns, respectively. Avg. abs. dev., average absolute deviation.
[a]PROPKA calculations based on the crystal structure (PDB ID 2HBB).
[b]NMR derived pKa values for NTL9 (13) and NTL9(1–22) (14).

Importantly, our data showed a lack of correlation between backbone hydrogen bonding and ionic interactions that, individually, are very weak. In particular, no specific interaction was found between Asp^8 and Lys^{12}, as suggested for the unfolded state of the intact NTL9 by both experiment ([15],[16]) and simulation ([24]). Thus, our data suggest that the formation of a nonnative ionic pair, Asp^8-Lys^{12}, may require interactions with residues beyond 1–22. A possible scenario is given by a most recent double-mutant study by Raleigh and co-workers, which revealed energetic coupling between Lys^{12} and hydrophobic core residues as far away in sequence as Leu^{35} and Leu^{47} in the unfolded state of NTL9 ([49]). Another interesting and perhaps surprising finding of this work is with regard to the pK_a of Asp^8 in the fragment peptides. The calculated pK_a downshift of nearly 0.5 units, in agreement with NMR measurement, is not related to a specific electrostatic interaction, but rather is due to the collective nonspecific attractive interactions with lysines, which are abundant in the fragment sequences. Together, our data support the view that the unfolded state contains native-like topology and hydrophobic contacts.

SUPPORTING MATERIAL

Seven figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00618-9.

AUTHOR CONTRIBUTIONS

J.S. designed the research; W.C. and C.S. performed the research; W.C. and J.S. analyzed the data; and W.C. and J.S. wrote the article.

ACKNOWLEDGMENTS

REFERENCES

1. Uversky, V. N., C. J. Oldfield, and A. K. Dunker. 2008. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* 37:215–246.

2. Karran, E., M. Mercken, and B. De Strooper. 2011. The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics. *Nat. Rev. Drug Discov.* 10:698–712.

3. Breydo, L., J. W. Wu, and V. N. Uversky. 2012. α-Synuclein misfolding and Parkinson's disease. *Biochim. Biophys. Acta.* 1822:261–285.

4. Bowler, B. E. 2012. Residual structure in unfolded proteins. *Curr. Opin. Struct. Biol.* 22:4–13.

5. Tanford, C., K. Kawahara, and S. Lapanje. 1966. Proteins in 6-M guanidine hydrochloride. Demonstration of random coil behavior. *J. Biol. Chem.* 241:1921–1923.

6. Jha, A. K., A. Colubri, …, T. R. Sosnick. 2005. Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc. Natl. Acad. Sci. USA.* 102:13099–13104.

7. Gillespie, J. R., and D. Shortle. 1997. Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.* 268:158–169.

8. Gillespie, J. R., and D. Shortle. 1997. Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol.* 268:170–184.

9. Shortle, D., and M. S. Ackerman. 2001. Persistence of native-like topology in a denatured protein in 8 M urea. *Science.* 293:487–489.

10. Mok, K. H., L. T. Kuhn, …, P. J. Hore. 2007. A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nature.* 447:106–109.

11. Felitsky, D. J., M. A. Lietzow, …, P. E. Wright. 2008. Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proc. Natl. Acad. Sci. USA.* 105:6278–6283.

12. Bruun, S. W., V. Iešmantavičius, …, F. M. Poulsen. 2010. Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein. *Proc. Natl. Acad. Sci. USA.* 107: 13306–13311.

13. Kuhlman, B., D. L. Luisi, …, D. P. Raleigh. 1999. pKa values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions. *Biochemistry.* 38:4896–4903.

14. Horng, J. C., V. Moroz, …, D. P. Raleigh. 2002. Characterization of large peptide fragments derived from the N-terminal domain of the ribosomal protein L9: definition of the minimum folding motif and characterization of local electrostatic interactions. *Biochemistry.* 41:13360–13369.

15. Cho, J.-H., and D. P. Raleigh. 2005. Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins. *J. Mol. Biol.* 353:174–185.

16. Cho, J.-H., S. Sato, and D. P. Raleigh. 2004. Thermodynamics and kinetics of non-native interactions in protein folding: a single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state. *J. Mol. Biol.* 338:827–837.

17. Cho, J.-H., and D. P. Raleigh. 2006. Electrostatic interactions in the denatured state and in the transition state for protein folding: effects of denatured state interactions on the analysis of transition state structure. *J. Mol. Biol.* 359:1437–1446.

18. Oliveberg, M., S. Vuilleumier, and A. R. Fersht. 1994. Thermodynamic study of the acid denaturation of barnase and its dependence on ionic strength: evidence for residual electrostatic interactions in the acid/thermally denatured state. *Biochemistry.* 33:8826–8832.

19. Oliveberg, M., V. L. Arcus, and A. R. Fersht. 1995. pKA values of carboxyl groups in the native and denatured states of barnase: the pKA values of the denatured state are on average 0.4 units lower than those of model compounds. *Biochemistry.* 34:9424–9433.

20. Tan, Y.-J., M. Oliveberg, …, A. R. Fersht. 1995. Perturbed pKA-values in the denatured states of proteins. *J. Mol. Biol.* 254:980–992.

21. Meng, W., and D. P. Raleigh. 2011. Analysis of electrostatic interactions in the denatured state ensemble of the N-terminal domain of L9 under native conditions. *Proteins.* 79:3500–3510.

22. Meng, W., N. Lyle, …, R. V. Pappu. 2013. Experiments and simulations show how long-range contacts can form in expanded unfolded proteins with negligible secondary structure. *Proc. Natl. Acad. Sci. USA.* 110:2123–2128.

23. Khandogin, J., and C. L. Brooks, 3rd. 2005. Constant pH molecular dynamics with proton tautomerism. *Biophys. J.* 89:141–157.

24. Shen, J. K. 2010. Uncovering specific electrostatic interactions in the denatured states of proteins. *Biophys. J.* 99:924–932.

25. Shen, J. K. 2010. A method to determine residue-specific unfolded-state pKa values from analysis of stability changes in single mutant cycles. *J. Am. Chem. Soc.* 132:7258–7259.

26. Azia, A., and Y. Levy. 2009. Nonnative electrostatic interactions can modulate protein folding: molecular dynamics with a grain of salt. *J. Mol. Biol.* 393:527–542.

27. Voelz, V. A., G. R. Bowman, …, V. S. Pande. 2010. Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1–39). *J. Am. Chem. Soc.* 132:1526–1528.

28. Nguyen, H., J. Maier, …, C. Simmerling. 2014. Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. *J. Am. Chem. Soc.* 136:13959–13962.

29. Pronk, S., S. Páll, …, E. Lindahl. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics.* 29:845–854.

30. Lindorff-Larsen, K., S. Piana, …, D. E. Shaw. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins.* 78:1950–1958.

31. MacKerell, Jr., A. D., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

32. Mackerell, Jr., A. D., M. Feig, and C. L. Brooks, 3rd. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400–1415.

33. Best, R. B., X. Zhu, …, A. D. Mackerell, Jr. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* 8:3257–3273.

34. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

35. Brooks, B. R., C. L. Brooks, 3rd, …, M. Karplus. 2009. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30:1545–1614.

36. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.

37. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.

38. Hess, B., H. Bekker, …, J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.

39. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.

40. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.

41. Feig, M., J. Karanicolas, and C. L. Brooks, 3rd. 2004. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.* 22:377–395.

42. Olsson, M. H. M., C. R. Søndergaard, …, J. H. Jensen. 2011. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* 7:525–537.

43. Best, R. B., N.-V. Buchete, and G. Hummer. 2008. Are current molecular dynamics force fields too helical? *Biophys. J.* 95:L07–L09.

44. Wickstrom, L., A. Okur, …, C. L. Simmerling. 2006. The unfolded state of the villin headpiece helical subdomain: computational studies of the role of locally stabilized structure. *J. Mol. Biol.* 360:1094–1107.

45. Zhou, R., B. J. Berne, and R. Germain. 2001. The free energy landscape for β hairpin folding in explicit water. *Proc. Natl. Acad. Sci. USA.* 98:14931–14936.

46. Zagrovic, B., E. J. Sorin, and V. Pande. 2001. β-hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* 313:151–169.

47. Best, R. B., and J. Mittal. 2011. Microscopic events in β-hairpin folding from alternative unfolded ensembles. *Proc. Natl. Acad. Sci. USA.* 108:11087–11092.

48. Dyson, H. J., P. E. Wright, and H. A. Scheraga. 2006. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc. Natl. Acad. Sci. USA.* 103:13057–13061.

49. Cho, J.-H., W. Meng, …, D. P. Raleigh. 2014. Energetically significant networks of coupled interactions within an unfolded protein. *Proc. Natl. Acad. Sci. USA.* 111:12079–12084.

50. Chen, W., C. Shi, A. D. Mackerell, Jr., and J. Shen. 2015. Conformational dynamics of two natively unfolded fragment peptides: comparison of the AMBER and CHARMM force fields. *J. Phys. Chem. B.* 119:7902–7910.