Balance between α and β Structures in Ab Initio Protein Folding

Robert B. Best*,† and **Jeetain Mittal*,‡**

Department of Chemistry, Cambridge University, Lensfield Road, Cambridge CB2 1EW, U.K., and Department of Chemical Engineering, Lehigh University, Bethlehem, Pennsylvania

Despite initial successes in folding of proteins by molecular simulation, it is becoming increasingly evident that current energy functions (force fields) tend to favor either α or β secondary structure, such that the choice of force field is governed by the structural class of the protein. Here, we study the folding of peptides with either predominantly α (Trp cage) or β (GB1 hairpin) structure with a modified version of the Amber ff03 force field, optimized to reproduce structural propensity in a helix-forming peptide. Using extensive replica exchange molecular dynamics simulations starting from completely unfolded configurations, we obtain the correct folded structure for each peptide, in close agreement with the experimental native structure (<1.5 Å all-atom root-mean-square deviation). We obtain converged equilibrium distributions, with folded populations at standard conditions (≈300 K), in remarkable accord with experiment. Further comparison to experimental data from NMR spectroscopy and FRET suggests that although the folded structures are accurately reproduced, the unfolded state remains too structured and compact. Our results suggest that the backbone correction results in a force field that is transferable to the folding of proteins from different structural classes.

Introduction

All-atom molecular dynamics simulations have been of great value in understanding protein folding and function,[1] including the folding of several small proteins.[2–4] Protein folding simulations provide molecular details of the folding reaction that would be hard to obtain directly from experiment, although single-molecule studies are rapidly bridging the gap.[5] Since advances in computation now allow microsecond[6,7] and soon millisecond[8] time scales to be accessed, the principal limitation of this approach is the accuracy of the models themselves, largely determined by the energy function or force field employed.[8] A central problem in force field parametrization is the balance between α and β structures: many commonly used force fields are known to have a bias toward helical structure,[9–11] leading to a number of efforts to improve secondary structure balance in all-atom force fields.[9,12] To address this deficiency, we have recently optimized commonly used force fields to reproduce experimental data for the helix−coil transition, a prototype for protein folding.[13] Notably, although the corrections were only a fraction of thermal energy $k_B T$ per residue, they resulted in a significant population shift when summed over the whole molecule.

Since it is far from clear whether such optimizations will be transferable to other peptides and proteins, as parametrization was based primarily on helix−coil data for a single peptide, we have chosen to investigate this issue further. As an example of a peptide forming β structure, we use the GB1 hairpin (Figure 1), residues 41−56 of the immunoglobulin binding domain of streptococcal protein G. This molecule was observed to fold independently[14] into a hairpin similar to that in the native protein structure,[15] and the coincidence of melting curves derived

from different probes indicates that folding is cooperative.[16,17] It is ∼50% folded at room temperature and folds in ≈6 μs at 300 K.[16]

As an example of a helical peptide (including both α and 3_{10} helix), we use Trp cage, a designed miniprotein derived from a fragment of a larger protein[18] (Figure 1). At 300 K, this peptide is ≈90% folded and folds in ≈4 μs.[19] Although most experimental studies found that Trp cage folded in a two-state fashion,[18–21] some studies suggested a less cooperative transition[22] or the population of an intermediate.[23] Due to their size, both GB1[24−47] and Trp cage[48−59] have been the subject of numerous simulation studies. Nonetheless, due to the previously mentioned force field bias, it is usually necessary to choose force fields favoring α or β structure for Trp cage or GB1, respectively.

We present results of replica exchange molecular dynamics (REMD) simulations of Trp Cage and the GB1 hairpin using the modified ff03* variant[13] of the Amber ff03 force field.[60] The original ff03 has too high a helical propensity:[12,13] although the correctly folded Trp cage structure has been obtained with ff03 in explicit solvent[57] starting from unfolded configurations, the lowest free energy states of the GB1 hairpin in this force field are helical. We obtain a near-equilibrium distribution of folded and unfolded configurations for both peptides in explicit solvent, on the basis of the convergence of the folded population starting from either the folded or unfolded state. At 300 K, GB1 and Trp cage are, respectively, ∼65% and ∼90% folded, in good agreement with experiment. To our knowledge, this is the first time both of these peptides have been folded with the same all-atom force field and explicit solvent, with the correct (unblocked) termini. The accuracy of ff03* is further corroborated by comparison of chemical shift, Förster resonance energy transfer (FRET) efficiency data, and NOE distances from our simulation with experiment. These results clearly suggest that this force field will be well-suited to studies of protein folding and related processes. The mechanism for the GB1 hairpin folding inferred

* Address correspondence to either author. E-mails: (R.B.B.) rbb24@cam.ac.uk; (J.M.) jeetain@lehigh.edu.
† Cambridge University.
‡ Lehigh University.

α and β Structures in Ab Initio Protein Folding

J. Phys. Chem. B, Vol. 114, No. 26, 2010 **8791**
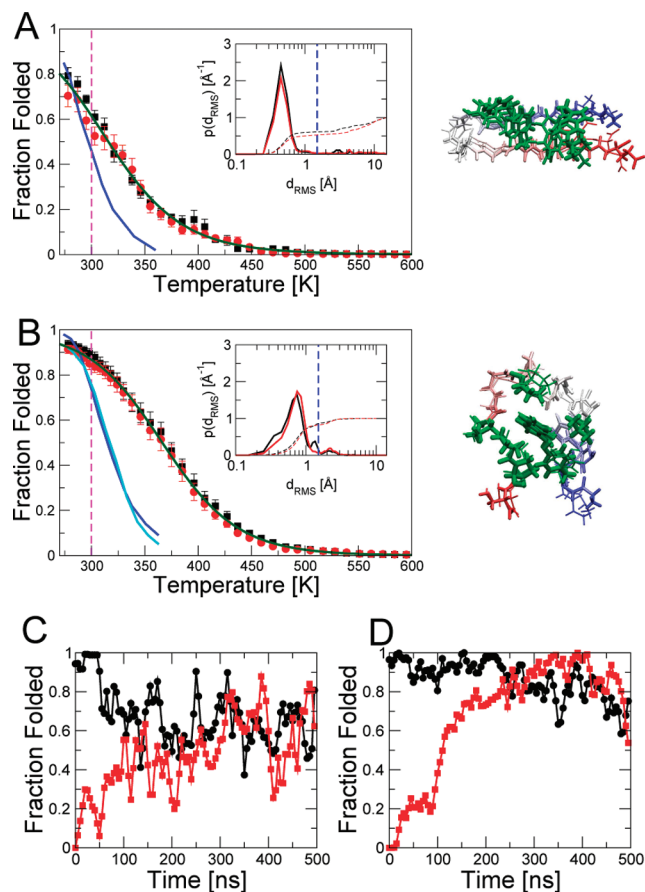
Figure 1. Convergence of folded populations. Folded populations from simulations starting from folded (black squares) and unfolded (red circles) configurations for (A) GB1 and (B) Trp cage along with fit to a two-state thermodynamic model (see text) shown by green solid lines. The experimental data are shown by solid blue lines for GB1[16] and blue[18] and cyan[21] for Trp cage. Insets to A and B: distribution of d_{RMS} for the 303 K replica for runs started from folded (black) and unfolded (red). The broken blue line indicates the d_{RMS} threshold for folded configurations (1.5 Å), and broken black and red lines indicate the cumulative distributions. To the right of A and B, the folded structures from the simulations (thin lines) are superimposed on the experimental structures (thick lines) for the GB1 hairpin (1GB1)[14,15] and Trp cage (1L2Y).[18] Averages of the folded population over a 10 ns window are shown as a function of the window origin in C and D for GB1 and Trp cage, respectively: black and red curves give the averages from simulations in which all replicas were started from folded, or unfolded, respectively.

from our simulations is similar to the "zipper" mechanism proposed from experiment,[16,61–63] in contrast to the results of many previous simulation studies. Our simulations also indicate areas in which the current energy function can be improved in the future; namely, (i) the temperature dependence of the folded population is too weak, analogous to previous findings for the helix−coil transition with this and other force-fields,[13,64] and (ii) the unfolded state is too collapsed relative to experiment.[65]

Results and Discussion

Equilibrium Folding of GB1 Hairpin and Trp Cage. In long REMD simulations starting from only unfolded structures, we observe a population of correctly folded molecules for both the peptides, defined as having a backbone $d_{RMS} < 1.5$ Å (see Models and Methods for definition). Although the REMD simulations result in discontinuous trajectories at a single temperature, we obtain continuous trajectories by following a

given replica through the exchanges in temperature space. In Supporting Information (SI) Figure 1, we present several such trajectories, monitored by the fraction of all-atom (excluding hydrogen atoms) native contacts, Q_{aa} (see Models and Methods for definition), which is a convenient coordinate for monitoring folding because the fluctuations in Q_{aa} in the folded and unfolded states are similar.[66] The trajectories reveal cooperative folding and unfolding events for each peptide (in total: 114 for GB1 and 97 for Trp cage, summed over all replicas). Although there are clearly intermediates present in the simulations, there are two dominant states, at low Q_{aa} (unfolded) and high Q_{aa} (folded), suggesting that these peptides are two-state, at least from a thermodynamic perspective.

An obvious concern with any simulation study is whether the length of the runs is sufficient to obtain a representative sampling of the phase space such that accurate equilibrium averages may be obtained. A demanding test of such "convergence" for protein folding is the comparison between the results of simulations started from folded and unfolded configurations. Previous studies have observed that REMD simulations on the order of 50 ns per replica were too short to obtain converged results in the low temperature replicas[67] for the miniprotein BBA5. Similarly, Juraszek and Bolhuis note that 36 ns per replica is insufficient to obtain converged results at relevant temperatures for Trp cage simulations.[54] By plotting the folded population in the 303 K replica, averaged over a moving "window" of 10 ns, we are able to assess how fast the averages for the simulations from different initial conditions converge (Figure 1C, D). Remarkably, we find that ∼200 ns is required before "converged" populations are obtained, much longer than the total length of many REMD simulations. Although it is possible that a different choice of replica exchange parameters may result in more efficient sampling, we did not obtain a significant improvement on varying the time interval between exchanges.

We discard the first 0.2 μs of each simulation out of 0.5 μs while calculating equilibrium averages and obtain well-converged results with similar folded populations starting from either a folded or unfolded structure, as shown in Figure 1A, B. We find folded populations of ∼65% for GB1 and ∼90% for Trp cage at 300 K, in reasonable agreement with the experimental values of ∼50%[16] and ∼75%,[18,21] respectively. We note that the GB1 hairpin structure is not stable in the parent ff03 force field. (SI Figure 2). Representative folded structures obtained from the simulations starting from unfolded configurations are shown in Figure 1. The folded Trp cage has a very well-defined structure, whereas the folded GB1 shows significant "fraying" at the ends and side-chain disorder, which is consistent with experimental findings.[16] The folded structures obtained in the simulations are in accord with the NMR data for each peptide (see below). However, we find that the temperature dependence of the folded population is too weak relative to experiment for both proteins. A two-state thermodynamic fit (Figure 1) gives enthalpies and entropies of unfolding that are about half of the experimental values for both proteins (parameters in Table 1).

A number of earlier studies have aimed to improve the "balance" between α and β structure in folding simulations, although these have mainly employed implicit solvent. Using a generalized Born (GB) implicit solvent, Kim et al. optimized the Amber ff99 force field to make the folded states of five miniproteins the global free energy minimum and obtained promising results for the proteins in the training set.[31] A somewhat different approach was taken by Sakae and Okamoto, who optimized Amber ff94 in conjuction with a GB solvent;

TABLE 1: Fit of Two-State Thermodynamic Model[a]

		GB1 hairpin		Trp cage	
parameter	unit	simulation	experiment	simulation	experiment
T_0	K	316 (14)	297^b, 295.3^c	366 (4)	317.1 (0.8)
ΔH_0	kcal mol^{-1}	5.75 (0.18)	11.6^b, 12.2^c	7.54 (0.09)	13.38 (0.24)
ΔS_0	cal mol^{-1} K^{-1}	18.2 (0.6)	39.0^b, 41.4^c	20.6 (0.2)	42.19 (0.11)
ΔC_p	cal mol^{-1} K^{-1}	24.0 (6.4)	$\equiv 0^{b,c}$	38.7 (3.0)	59.75 (0.01)

[a] We fit the equilibrium fraction folded p_F to $p_F = [1 + \exp(-\Delta G/k_B T)]^{-1}$, with the enthalpy and entropy of unfolding given by $\Delta H(T) = \Delta H_0 + \Delta C_p(T - T_0)$, $\Delta S(T) = \Delta S_0 + \Delta C_p \ln(T/T_0)$, with folding midpoint $T_0 = \Delta H_0/\Delta S_0$. The results are compared with the [b] spectroscopic[16] and [c] calorimetric[17] data for GB1 and calorimetric measurements for Trp cage.[21] The experimental fit for GB1 assumes a zero ΔC_p; however, we include it in our fit due to the large range of temperature covered by the REMD simulations.

however, they optimized the force fields on folded proteins, with the target of minimizing the root-mean-square force on each atom in each protein's native state.[68,69] Tests of the modified force fields on short peptides showed improved agreement with experimental results.[70] An alternative approach was pursued by Shell et al., who found optimal combinations of force fields with generalized Born solvent models by testing the different force fields on the folding of miniproteins.[32] Our work extends these results to the performance of all-atom force fields for folding with explicit solvent.

Comparison of Equilibrium Results with Experiment. Although we are able to obtain folded configurations that are close to the experimental structures and folded populations in reasonable agreement with experiment near 300 K, the most direct comparison with experiment involves the calculation of experimental observables from the simulations. In this way, we can also test whether the distribution of unfolded configurations sampled in the simulations is consistent with the available data. We focus on the NMR chemical shift measurements and the NOE distances available for both peptides, and the FRET measurements made on the GB1 hairpin.

NMR Data. NMR chemical shifts are a sensitive measure of peptide conformation, reflecting both backbone structure and side-chain packing. A number of accurate empirical algorithms have been developed for the calculation of shifts from structure, on the basis of correlations of chemical shifts with simple geometrical properties.[71−73] Here, we have calculated chemical shifts on the basis of SHIFTX[71] and CamShift algorithms[73] to assess our simulation results. We focus on chemical shift deviation (CSD) H^α shifts, for which the difference from standardized "random-coil" chemical shifts can be most accurately predicted by these approaches. At low temperature, we obtain excellent agreement with the experimental data using either of the algorithms (Figure 2A, B). We find that the shifts derived from the more recently developed CamShift algorithm are in slightly better agreement with the experimental data. The main discrepancies with experiment are for H^α of residues G11, P17, P18 in Trp cage (the shift prediction algorithms do not distinguish between the diastereotopic H^α of glycine). Since these residues are all immediately adjacent to W3, it is most likely that the chemical shifts of these residues are subject to Trp ring current effects that are imperfectly captured by the shift prediction algorithms.

These results mainly confirm the accuracy of the folded structures obtained, since these are the majority of the population at low temperature. We have therefore used a set of temperature-dependent H^α shifts for GB1[17] to assess the accuracy of the sampled configurations at higher temperatures (Figure 2C). Qualitatively, we obtain similar results for the CSD temperature-dependence, with residues having positive CSD decreasing, and those with negative CSD, increasing, as expected for convergence toward random coil shifts at high temperature. Some
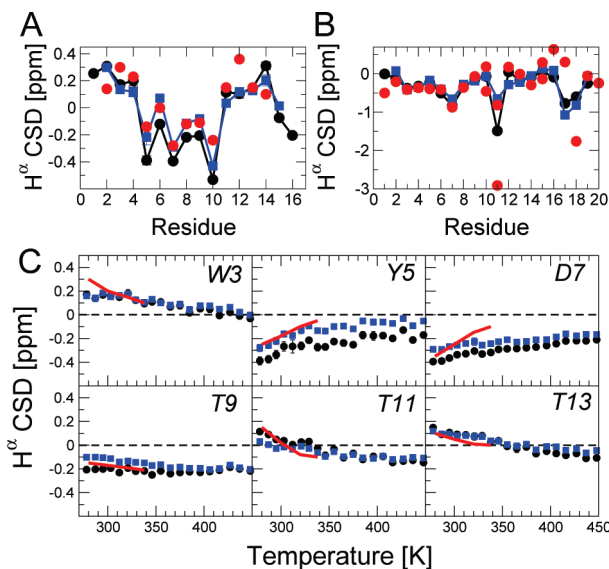
Figure 2. Comparison with CSD from experiment. The H^α CSD, obtained by subtracting the random coil shift, is shown for (A) GB1 and (B) Trp cage. Values calculated from the 278 K replicas with SHIFTX (black circles)[71] and CamShift (blue squares)[73] are compared with the experimental data at 278 K for GB1[88] and at 282 K for Trp cage[18] (red circles). α-Helical structure is usually associated with a negative CSD, and extended structures, with a positive CSD. Note that although two α proton shifts can be measured for Gly, both SHIFTX and CamShift report only a single value. (C) Temperature dependence of selected chemical shift deviations (CSD) for GB1. The experimental CSDs of Honda et al.[17] (solid lines) are compared with those calculated from simulation using SHIFTX (black circles) and CamShift (blue squares).

residues (W3, D7) may show a slightly weaker temperature dependence than in experiment.

Last, we note that the NOE-derived distance restraints are also well-satisfied by the low temperature replicas (SI Figure 3).

FRET Efficiencies. FRET efficiencies are initially calculated assuming that the conformational dynamics is much slower than the donor fluorescence lifetime (≈3 ns for Trp). Assuming, in addition, that either the donor or acceptor orientation decays on a time scale faster than the fluorescence lifetime, the FRET efficiency may be calculated from $\langle E(R) \rangle = \langle (1 + (R/R_0)^6)^{-1} \rangle$ where R is the distance between donor and acceptor chromophores and R_0 is the spectroscopically determined Förster radius (2.2 nm).[74,75] Although the donor Trp is explicitly present in our simulations, the acceptor (a dansylated Lys at the C terminus) is not, so we measure the distance between the Trp side-chain C6 and the C-terminal Glu Cδ and add a fixed extra distance, $\delta R = 0.2$ nm, to this to account for the displacement of the Dansyl chromophore from Glu 16.

The calculated FRET efficiencies are shown in Figure 3 and are in reasonable agreement with the experimental efficiencies[16]

α and β Structures in Ab Initio Protein Folding

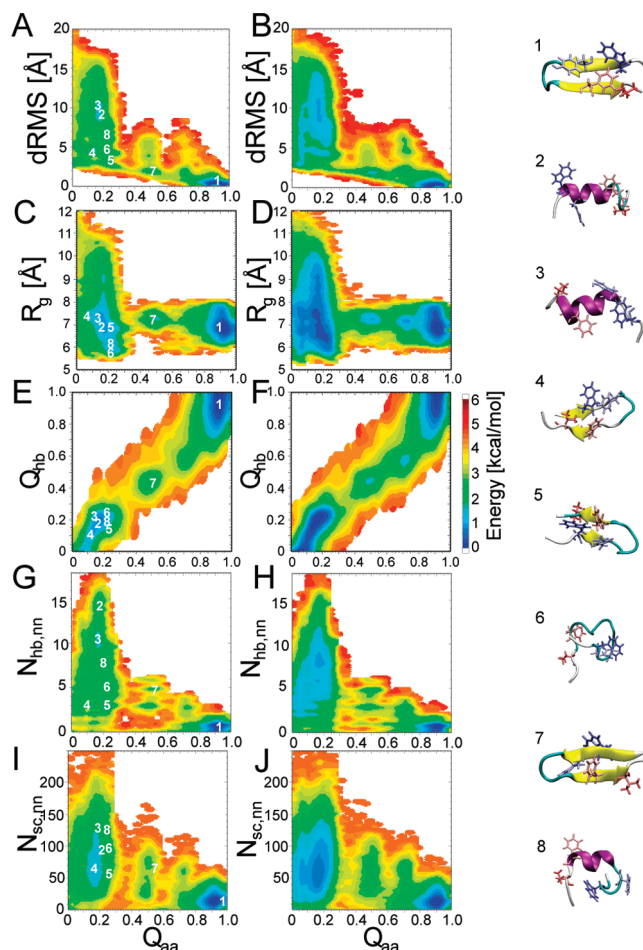*J. Phys. Chem. B, Vol. 114, No. 26, 2010* **8793**
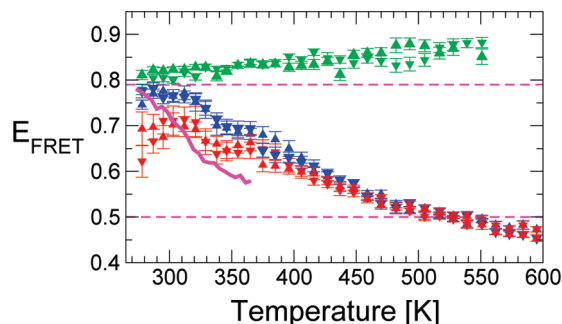
Figure 3. Trp-Dansyl FRET transfer efficiencies. FRET efficiencies were calculated from the simulations assuming a linker offset of 0.2 nm and an R_0 of 2.2 nm.[16] For slow chain dynamics relative to donor lifetime: blue symbols, ensemble average efficiency; green symbols, average over only folded state (backbone $d_{RMS} < 1.5$ Å); red symbols, average over only unfolded state. "Up" and "down" triangles refer to replica exchange starting from folded and unfolded, respectively. Solid magenta line, experimental data from Munoz et al.;[16] broken magenta lines, folded and unfolded efficiencies from two-state experimental analysis.

Figure 4. Folding free energy surfaces for GB1. Two-dimensional potentials of mean force have been calculated from projections onto the backbone distance matrix rms, d_{RMS}; the radius of gyration, R_g; the all-atom fraction of native contacts, Q_{aa}; the fraction of native hydrogen bonds, Q_{hb}; the number of non-native hydrogen bonds, $N_{hb,nn}$; and the number of non-native side-chain contacts, $N_{sc,nn}$. In A, C, E, G, and I are shown free energy surfaces at 303 K, with the corresponding surfaces at 355 K in parts B, D, F, H, and J. The centroid location of the clusters discussed in the text are indicated on the plots. Representative structures from the most populated clusters for the GB1 hairpin at 303 K are shown at right. The primary data have been smoothed using Gaussians of width comparable to the grid spacing.

near 300 K. Interestingly, the folded efficiency increases slightly with temperature, as a consequence of "fraying" of the hairpin termini at elevated temperatures. More noteworthy is the fact that the overall temperature dependence is weaker than observed experimentally. There are two contributions to this: first, the too-weak temperature dependence of the folded population as mentioned above; and second, an efficiency for the unfolded state that is apparently too high. This latter effect can be attributed to an unfolded state that is too "collapsed" or structured, as we discuss further below. Other values of δR do not change these qualitative conclusions, as shown in SI Figure 4: shorter values result in a similar overall picture, whereas longer values produce a FRET efficiency for the folded state that is inconsistent with the experimental value. Calculating FRET efficiencies in the limit of "fast" conformational dynamics (i.e., the peptide samples its full conformational space within the fluorescence lifetime), an unlikely scenario, results in too high mean efficiency at all temperatures (SI Figure 4).

Folding Free Energy Surfaces. We characterize the energy landscape of each peptide by calculating two-dimensional free energy surfaces for projections onto selected reaction coordinates chosen to capture both local and global structure formation. As a common reference coordinate, we use the global fraction of all-atom native contacts Q_{aa}, which is a good folding coordinate for minimally frustrated models.[76,77] We construct two-dimensional potentials of mean force as a function of Q_{aa}, and either distance matrix rms d_{RMS}, radius of gyration R_g, fraction of native hydrogen bonds Q_{hb}, number of non-native hydrogen bonds $N_{hb,nn}$, or number of non-native side-chain contacts $N_{sc,nn}$. Free energy surfaces for the GB1 hairpin shown in Figure 4 reveal two dominant minima: an "unfolded" state near $Q_{aa} = 0.2$ and a "folded" state near $Q_{aa} = 0.9$. We note that the free energy surfaces calculated from REMD starting from either folded or unfolded are very similar, even for the low temperature replicas (SI Figure 5). The unfolded state shows considerable heterogeneity with structures varying widely in compactness, as evident in the radius of gyration, R_g, distribution. All the unfolded structures contain a very low fraction of native hydrogen bonds, Q_{hb}, indicating the absence of native-like secondary structure in the unfolded state. However, many of the unfolded states do contain an appreciable number of non-native hydrogen bonds (Figure 4G, H) as well as non-native side-chain contacts (Figure 4I, J). This indicates that the

unfolded state for GB1 in our model is not just "random coil" but also contains some non-native secondary structure.[78,79]

To obtain insight into the structural origin of the various basins on the free energy landscape, we have performed a global cluster analysis of the equilibrium configurations at 303 K (Table 2). We observe a single dominant cluster with population 62.2%, corresponding to the correctly folded hairpin, with the remaining "unfolded" clusters all having population of ~3% or less. Notably, however, many of the unfolded clusters are quite narrowly defined when projected onto the various coordinates we consider (Table 2). We have indicated the positions of each of these clusters on the free energy surfaces in Figure 4 along with the centroid structure from each cluster. These structures confirm that in the most frequently visited clusters of the unfolded state, there is considerable population of non-native secondary structure, including both helical and non-native sheets and hairpins. Note that although short helices are present, these represent a small fraction of the total ensemble (*total* helical population = 8% at 303 K). This may reflect too great a propensity for the formation of local secondary structure in this

**TABLE 2: Most Populated Clusters of the GB1 Hairpin at 303 K[a]**

cluster	pop., %	$\langle Q_{aa}\rangle$	$\langle N_{hb,nn}\rangle$	$\langle d_{RMS}\rangle$, Å	$\langle R_g\rangle$, Å	backbone state
native						−EEEEEAAALEEEEE−
1	62.2	0.90 (0.05)	0.7 (0.4)	0.5 (0.2)	7.0 (0.4)	−EEEEEAAALEEEEE−
2	3.1	0.17 (0.02)	14.0 (1.3)	9.4 (0.8)	7.0 (0.4)	−EAAAAAAAAAAAEE−
3	2.4	0.17 (0.02)	10.9 (0.8)	9.9 (0.9)	7.2 (0.4)	−EEEAAAAAAAAAEE−
4	2.4	0.13 (0.02)	2.9 (0.3)	4.6 (0.3)	7.4 (0.2)	−EEEAEAAEEEEEEE−
5	2.3	0.23 (0.03)	3.1 (0.4)	3.0 (0.3)	6.9 (0.2)	−EEEEAEAAEEEAEE−
6	1.7	0.24 (0.05)	4.9 (0.5)	4.4 (0.2)	5.9 (0.1)	−EEAAEAAAAAAEEE−
7	1.3	0.50 (0.03)	5.0 (0.7)	2.0 (0.4)	7.2 (0.1)	−EEEAEAAALEEEEE−
8	1.2	0.23 (0.02)	7.2 (1.0)	6.4 (0.6)	6.0 (0.3)	−EAAAEAAAAAAEAE−
sequence						EEWTYDDATKTFTVTE

[a] Clusters were obtained by grouping conformations in the 303 K REMD replica of the GB1 hairpin with the single linkage method with a 1 Å cut-off; the total population, average native fraction of native contacts (Q_{aa}), number of non-native hydrogen bonds ($N_{hb,nn}$), distance matrix RMS (d_{RMS}), radius of gyration (R_g), and the most frequent backbone state are summarized for each cluster. Numbers in brackets give the standard deviation of each quantity over the cluster. The backbone states have been classified according to the region of the Ramachandran map populated by the 14 non-terminal residues, using "E", "A" and "L" to denote the extended (β and polyprolineII), right-handed helical α_R, and left-handed helical α_L regions, respectively. Residues that are in a non-native conformation are underlined.

force field, although this would need further testing against experimental data for weakly structured peptides.

Of the most unfolded states, those with predominantly helical conformations (e.g., clusters 2, 3, 8) generally appear the most "non-native", when using any of d_{RMS}, $N_{hb,nn}$, or $N_{sc,nn}$ as coordinates. In general, the 2D free energy surfaces involving these coordinates show a distinctive "L"-shape, indicating that, in general, non-native contacts must first be broken before native contacts can be formed. An interesting exception to this is an apparent intermediate near $Q_{aa} = 0.5$ (cluster 7), with a considerable fraction of both native and non-native hydrogen bonds. The structure of this apparent intermediate is revealing. It consists of a fully formed β-hairpin in which one of the strands has been "flipped" such that the hydrophobic side chains are now on opposite sides of the hairpin rather than on the same side. The difference in backbone conformation is only one residue, with Tyr5 in the α_R rather than the extended region of Ramachandran space (Table 2). In this particular case, Q_{aa} is not a good reaction coordinate[77] because this intermediate appears to be "on-pathway" (SI Figure 1); however, we never observe formation of the folded state directly from this structure (see, for example, the trajectory in SI Figure 1B), in which this state is populated twice around 0.15 μs.

The free energy surfaces for Trp cage shown in Figure 5 display a very well-defined minimum region for the folded state ($d_{RMS} < 1.5$ Å). The fully unfolded state is highly disordered, with Q_{aa} and Q_{hb} less than 0.1, but has negligible population at 303 K. Instead, a collapsed intermediate state is visible near $Q_{aa} \approx 0.6$, $Q_{hb} \approx 0.6$. Cluster analysis of Trp cage in the 303 K replica results in only three dominant clusters. Cluster 3 corresponds to the collapsed intermediate and has native-like helical structure but a distorted turn region and disordered hydrophobic core (Figure 5). CIDNP NMR experiments on unfolded Trp cage have suggested the presence of substantial native-like structure, which may be consistent with this intermediate.[80] We also find with our simulation model that the folded basin of the protein is split into two subminima with a combined population of 72.1%, having similar d_{RMS} from the experimental structure. The first, identified with cluster 1, has $Q_{aa} \approx 0.95$, and the second (cluster 2) has $Q_{aa} \approx 0.75$. Although the structures of these two substates are almost identical (Figure 5), there is, in fact, a subtle difference in backbone conformation: Gly15 is in the extended region of the Ramachandran map rather than the α_L region (Table 3). Analysis of the non-native side-chain contacts (Figure 5I, J) reveals that cluster 2 is also characterized by the formation of new side-chain contacts not
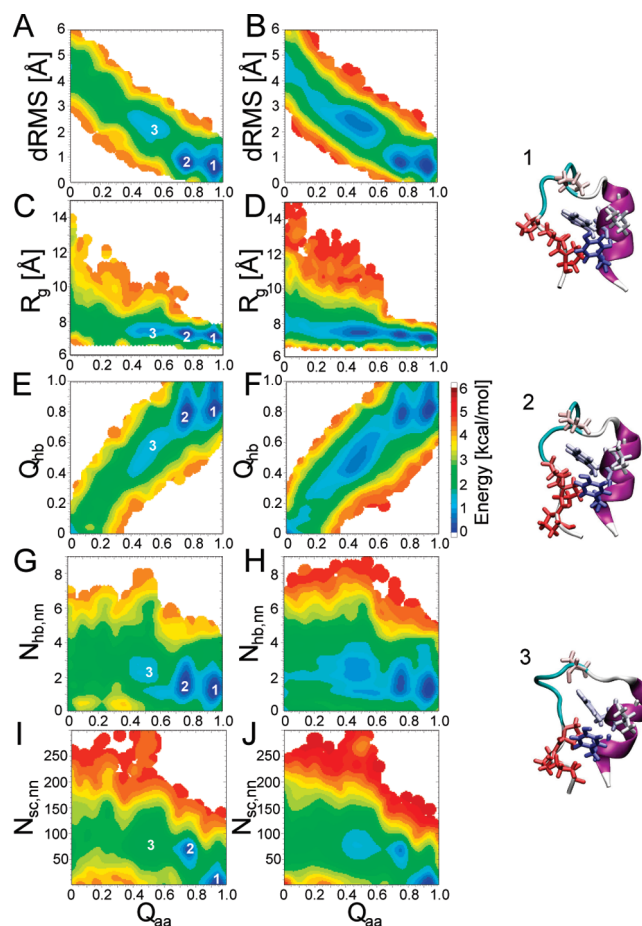
Figure 5. Folding free energy surfaces for Trp cage. Two-dimensional potentials of mean force were obtained using coordinates analogous to those for GB1; see the caption of Figure 4 for details. To the right are shown representative structures from the most populated clusters for Trp Cage at 303 K.

present in tthe native state. However, the hydrogen bonding in both clusters 1 and 2 is predominantly the native helical H-bonding, which is disrupted in the intermediate cluster 3.

Putative Folding Mechanism. Because of the exchanges of replicas, folding events do not occur at a single temperature in our simulations. Nonetheless, we can investigate the key events in folding by following the continuous trajectories through temperature space (with the caveat that the mechanism may vary with temperature). In Figure 6A, B we show typical folding transition paths for GB1 hairpin and Trp cage (additional paths

TABLE 3: Most Populated Clusters of Trp Cage[a]

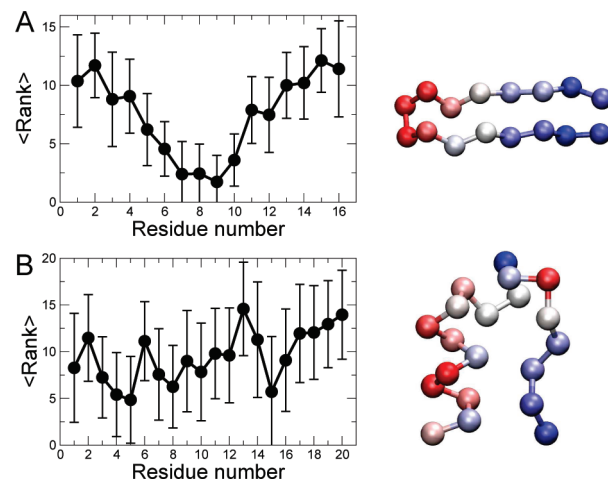| cluster	pop., %	$\langle Q_{aa}\rangle$	$\langle N_{hb,nn}\rangle$	$\langle d_{RMS}\rangle$, Å	$\langle R_g\rangle$, Å	backbone state
native						−AAAAAAAALEAAALEEEE−
1	41.0	0.94 (0.03)	1.4 (0.6)	0.7 (0.3)	7.7 (0.1)	−AAAAAAAALEAAALEEEE−
2	31.1	0.75 (0.03)	1.7 (0.6)	0.8 (0.2)	7.8 (0.1)	−AAAAAAAALEAAAEEEEE−
3	1.2	0.54 (0.03)	2.7 (0.6)	2.2 (0.1)	7.9 (0.1)	−AAAAAAAALEEAAAEEEE−
sequence						NLYIQWLKDGGPSSGRPPPS

[a] Clusters were obtained by clustering conformations from the 303 K replica with a 0.5 Å cut-off; other details as in the footnote to Table 2.
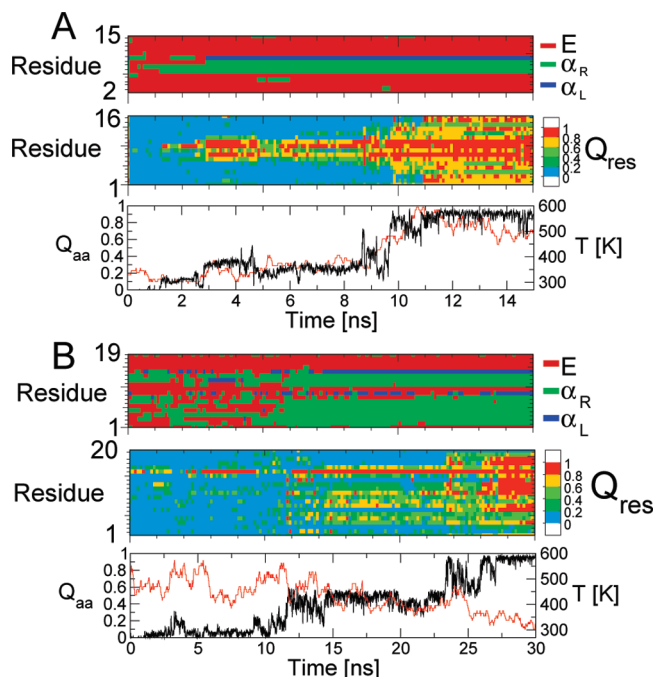
Figure 6. Folding pathways for the GB1 hairpin and Trp cage. Representative pathways are shown in A for GB1 and in B for Trp cage. The time evolution of backbone conformation (upper panels), fraction of native contacts per residue Q_{res} (center panels), and overall fraction of all-atom native contacts Q_{aa} (lower panels) is shown. On the right-hand axis (red curves) in the lower panels, we show the temperature of the replica as a function of time.

Figure 7. Average order of local structure formation for (A) the GB1 hairpin and (B) Trp cage. For each transition path, the residues were ranked according to the order in which they formed 50% of their native contacts (see the Supporting Information). The plots show the average and standard deviation of this ranking for each residue. To the right of each graph is shown a C_α representation of the corresponding native peptide, colored according to the order in which residues reach 50% nativeness (red, early; blue, late).

are given in SI Figure 6). In addition to the global parameter Q_{aa}, we monitor the local state of each residue by computing a residue-wise fraction of native contacts $Q_{res,i}(t)$ analogous to Q_{aa} but only for native contacts of residue i and region of the Ramachandran map populated by the residues. We emphasize that most of these transitions occur over a period of time during which the continuous transition path trajectories span a wide range of temperatures. As a result, the transition paths cannot be considered as samples from an equilibrium distribution at any given temperature. This is clear from inspection of the temperature plots in Figure 6, which show large variations in temperature. However, we believe that these trajectories can be used to obtain a qualitative picture of mechanism from the present data.

The pattern of formation of residue native contacts Q_{res} for GB1 hairpin is consistent with a hairpin formation being initiated from the turn: in almost all cases, we observe that native contacts form first in the turn region and propagate toward the hairpin termini in a "zipper"-like mechanism, as inferred from most experiments.[16,61−63] To quantify this statement, we have analyzed the order in which contacts are formed[89] in all of the folding trajectories and unfolding trajectories (viewed in the folding direction) obtained from our replica-exchange simulations. For each transition path, we determine the time τ at which each contact is ≈50% formed. As an operational definition of τ, we

fit the function $f(t) = A[1 + \exp(-(t - \tau))]^{-1}$ to the fraction of native contacts for that residue, $Q_{res,i}(t)$, where A reflects the final fraction of native contacts formed by that residue (see SI Figure 7 for example fits). We then rank the residues by these fitted τ, obtaining the order in which they each assume native-like structure. Applying this method to all the transition paths allows us to assess the variation in the order of structure formation.

The results for the GB1 hairpin (Figure 7) clearly show that the predominant mechanism of folding is "zipper"-like, starting from the hairpin, with relatively little variance over the different transition paths. A careful inspection of all the transition paths reveals that there are, indeed, some where the termini also form native contacts early (perhaps more reminiscent of the "hydrophobic collapse" mechanism); however, these are a small fraction, as is clear from Figure 7. Thus, adoption of the correct conformation in the turn region appears to be critical for folding, occurring prior to or during the main folding event. We also find that Lys 10 adopts its native α_L conformation (rarely populated in unfolded peptides and proteins) concomitant with folding (Figure 6A, top). We find that the backbone hydrogen bonds are formed at the same time as the side-chain contacts, rather than via a "hydrophobic collapse" mechanism in which hydrophobic core formation precedes backbone hydrogen bonding, as proposed by many other simulation studies.[26,33]

The free energy surfaces for Trp cage are very suggestive of a concerted formation of secondary and tertiary structure; that is, the fraction of (helical) backbone hydrogen bonds increases together with the overall fraction of native contacts. The transition path for Trp cage (Figure 6B), monitored by the residue-based contacts, $Q_{res}(i)$, shows a cooperative formation

8796 *J. Phys. Chem. B, Vol. 114, No. 26, 2010*

Best and Mittal

of the fully folded structure, reflected by an almost simultaneous increase in $Q_{res}(i)$ throughout the sequence. Interestingly, similar to the GB1 hairpin, we find that a key event in folding is the transition to an α_L conformation, in this case by Gly 10 (Figure 6B, top). Analyzed in terms of the order of formation of native structure (Figure 7), the mechanism of Trp cage seems to be more heterogeneous, with the order of contact formation less well-defined. Nonetheless, it appears that residues near the center of the helix tend to form their native contacts slightly earlier, on average, than the rest of the peptide, and residues that make mainly hydrophobic contacts form their native structure last. This indicates that, although tightly coupled, helix formation on average precedes tertiary packing.

Concluding Remarks

We have used the folding of two miniproteins belonging to different structural classes (α and β) to show that simple optimization to backbone dihedral potential can correct the intrinsic secondary structure propensity of biased force fields. Since the ff03* parameters used were derived with a specific focus on the helix−coil transition, it was not clear a priori that the modifications would also represent an improvement for protein folding, especially for nonhelical proteins. We find that both the helical Trp cage and the GB1 β-hairpin fold to the correct structure, and the folded population at 300 K is in agreement with experiment. By contrast, the β-hairpin is not stable in the original ff03 force field. The experimental NMR and FRET data are also well reproduced. These results strongly suggest that the current force fields provide a reasonable description of the folding energy landscape near ambient conditions.

Although the primary focus here was to determine whether the ff03* force field is able to reproduce the folded structure and thermodynamic properties of selected miniproteins, we have also presented a preliminary qualitative analysis of mechanism based on continuous REMD trajectories moving through temperature space. These suggest that the folding mechanism is that inferred from experiment, particularly for the GB1 hairpin. We note, however, that since the configuration of interest is actually moving through the temperature space in continuous REMD trajectories, it is not possible to assign a specific temperature to a given folding event. In future work, we intend to probe the mechanisms using methods that avoid this complication.

We find that temperature dependence of the folded population is too weak relative to experiment, with unfolding entropy and enthalpy around half of the experimental values. Deficiencies in the force-field hydrogen bonding, where the enthalpic and entropic components are approximately half of those from experiment, likely make a major contribution to the discrepancy.[13] We also find, on the basis of comparison with FRET data for the GB1 hairpin, that the unfolded structures are too compact, and the unfolded state appears to contain many states with ordered secondary structure, both α and β. This may be because the force field, together with TIP3P water, is generally too "hydrophobic", favoring more compact and ordered structures.[65] In the absence of multibody effects such as polarizability, this increased bias toward local structure may help force fields based on pair potentials to fold proteins. Nonetheless, we have shown that a simple balance between β and α structure results in a force-field that can be used to fold structures from different structural classes with a remarkable accuracy.

Models and Methods

Replica Exchange Molecular Dynamics. Replica exchange molecular dynamics (REMD) simulations of the GB1 hairpin and the Trp cage were performed using the Gromacs 4.0.5 simulation package.[81−83] In a typical REMD simulation, several copies of the system are propagated in parallel at different temperatures. Periodically, an attempt is made to swap the configuration of two replicas following Metropolis criteria in such a way that the canonical probability distribution is maintained at each temperature. By following this procedure over the course of a simulation, thermodynamic averages may converge much more rapidly at lower temperatures due to enhanced configurational sampling achieved by configurations sampled at higher temperatures.[84]

The Amber ff03* force field (Amber ff03[60] with a modified Ψ torsion potential[13]) was used to represent the protein except where noted, and the TIP3P model was used for water.[85] The structure of the 16-residue GB1 hairpin was taken from residues 41−56 of the full length GB1 protein (PDB: 1GB1) and solvated in a truncated octahedron simulation cell with 3.5 nm between the nearest faces, containing 984 water molecules, 6 sodium ions and 3 chloride ions to neutralize the charge. The 20-residue Trp cage (PDB: 1L2Y) was solvated in a truncated octahedron water box of side 3.75 nm containing 1201 water molecules, 7 sodium, and 8 chloride ions. The termini of both peptides were unblocked, corresponding to the experimental situation.[16,18] For the GB1 hairpin, the salt bridge between the charged termini is expected to be an important contribution to the stability, analogous to the introduction of additional salt bridges between the termini by mutation;[86] an opposing effect is expected for Trp cage, in which the positive charge at the N-terminus would destabilize the helix.[87]

All REMD simulations were performed at constant volume with long-range electrostatics calculated using PME with a 1.2 Å grid spacing and 9 Å cutoff. The system was propagated using Langevin dynamics with a friction of 1 ps^{-1}, and replica exchange was attempted every 10 ps (every 5000 steps with a time step of 2 fs). The temperatures of the replicas spanned a range of 278−595 K, with 32 replicas for GB1 and 36 replicas for Trp cage. For each peptide, one set of REMD simulations was initiated with all the replicas in the folded state. A second set of REMD simulations was initiated from "unfolded" configurations, picked at fixed intervals from the 595 K replica of the first set of simulations after the initial equilibration period. All REMD runs were propagated for 0.5 μs per replica, for a total of 16 and 18 μs for GB1 and Trp cage, respectively.

Definition of Order Parameters. Backbone d_{RMS} is defined as $d_{RMS} = [N_{bb}^{-1} \sum_{(i,j)} (r_{ij} - r_{ij}^0)^2]^{1/2}$, where the sum runs over the N_{bb} backbone native contacts (i, j) of native amino acid contacts that are separated by distances r_{ij} in the configuration of interest and by r_{ij}^0 in the native state.

The fraction of all-atom (excluding hydrogen atoms) native contacts is defined as $Q_{aa} = N_{aa}^{-1} \sum_{(i,j)} [1 + \exp(\beta(r_{ij} - \lambda r_{ij}^0))]^{-1}$, where the sum runs over the N_{aa} pairs (i, j) of native amino acid contacts that are separated by distances r_{ij} in the configuration of interest and by r_{ij}^0 in the native state ($\beta = 5$ Å^{-1}; $\lambda = 1.2$).

Residue-wise, $Q_{res}(i) = N_j^{-1} \sum_j [1 + \exp(\beta(r_{ij} - \lambda r_{ij}^0))]^{-1}$ is defined analogous to Q_{aa}, but where the sum runs over only the N_j residues in contact with residue i in the native state.

Definition of Ramachandran Regions. We have defined three regions of the Ramachandran map to classify structures in the simulations:

(1) Right-handed helix α_R is defined as $-160 < \phi < -20$; $-120 < \psi < 50$ for nonglycine and as $-160 < \phi < -20$; $-90 < \psi < 70$ for glycine residues.

(2) Left-handed helix α_L is defined as $20 < \phi < 110$; $-80 < \psi < 90$ for nonglycine and as $20 < \phi < 160$; $-70 < \psi < 90$ for glycine (i.e., the mirror image of glycine α_R).

(3) "Extended" conformations, comprising β and polyproline II, are defined as $-180 < \phi < -20$; $50 < \psi < 180$, $-180 < \phi < -20$; $-180 < \psi < -120$ or $160 < \phi < 180$; $110 < \psi < 180$ for nonglycine and as $-180 < \phi < -20$; $70 < \psi < 180$, $-180 < \phi < -20$; $-180 < \psi < -90$, $20 < \phi < 180$; $90 < \psi < 180$, or $20 < \phi < 180$; $-180 < \psi < -70$ for glycine.

These are illustrated in SI Figure 8.

Supporting Information Available: Supporting figures 1−8 are available online. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.

(2) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102–106.

(3) Ensign, D. L.; Kasson, P. M.; Pande, V. S. *J. Mol. Biol.* **2007**, *374*, 806–816.

(4) Freddolino, P. L.; Park, S.; Roux, B.; Schulten, K. *Biophys. J.* **2009**, *96*, 3772–3780.

(5) Chung, H. S.; Louis, J. M.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11837–11844.

(6) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740–744.

(7) Maragakis, P.; Lindorff-Larsen, K.; Eastwood, M. P.; Dror, R. O.; Klepeis, J. L.; Arkin, I. T.; Jensen, M. Ã.; Xu, H.; Trbovic, N.; Friesner, R. A.; Palmer, A. G.; Shaw, D. E. *J. Phys. Chem. B.* **2008**, *112*, 6155–6158.

(8) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. *Curr. Opin. Struct. Biol.* **2009**, *19*, 1719–1722.

(9) Garcia, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782–2787.

(10) Best, R. B.; Buchete, N.-V.; Hummer, G. *Biophys. J.* **2008**, *95*, L07–L09.

(11) Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulten, K. *Biophys. J.* **2008**, *94*, L75–L77.

(12) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.

(13) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.

(14) Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 584–590.

(15) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657–661.

(16) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196–199.

(17) Honda, S.; Kobayashi, N.; Munekata, E. *J. Mol. Biol.* **2000**, *295*, 269–278.

(18) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425–430.

(19) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952–12953.

(20) Qiu, L.; Hagen, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 3398–3399.

(21) Streicher, W. W.; Makhatadze, G. I. *Biochemistry* **2007**, *46*, 2876–2880.

(22) Ahmend, Z.; Beta, I. A.; Mikhonin, A. V.; Asher, S. A. *J. Am. Chem. Soc.* **2005**, *127*, 10943–10950.

(23) Doose, H. N. S.; Sauer, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16650–16655.

(24) Klimov, D. K.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2544–2549.

(25) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068–9073.

(26) Zagrovic, B.; Sorin, E. J.; Pande, V. S. *J. Mol. Biol.* **2001**, *313*, 151–169.

(27) Zhou, R.; Berne, B. J. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12777–12782.

(28) Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *121*, 1080–1090.

(29) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14766–14770.

(30) Andrec, M.; Felts, A. K.; Levy, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6801–6806.

(31) Kim, E.; Jang, S.; Pak, Y. *J. Chem. Phys.* **2007**, *127*, 145104.

(32) Shell, M. S.; Ritterson, R.; Dill, K. A. *J. Phys. Chem. B* **2008**, *112*, 6878–6886.

(33) Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9062–9067.

(34) Roccatano, D.; Amadei, A.; DiNola, A.; Berendsen, H. J. C. *Protein Sci.* **1999**, *8*, 2130–2143.

(35) Ma, B.; Nussinov, R. *J. Mol. Biol.* **2000**, *296*, 1091–1104.

(36) Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins* **2001**, *42*, 345–354.

(37) Zhou, R.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931–14936.

(38) Bolhuis, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12129–12134.

(39) Colombo, G.; DeMori, G. M. S.; Roccatano, D. *Protein Sci.* **2003**, *12*, 538–550.

(40) Paschek, D.; Garcia, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *93*, 238105.

(41) Wei, G.; Mousseau, N.; Derreumaux, P. *Proteins* **2004**, *56*, 464–474.

(42) Nguyen, P. H.; Stock, G.; Mittag, E.; Hu, C.-K.; Li, M. S. *Proteins* **2005**, *61*, 795–808.

(43) Bolhuis, P. G. *Biophys. J.* **2005**, *88*, 50–61.

(44) Daidone, I.; D'Abramo, M.; Dinola, A.; Amadei, A. *J. Am. Chem. Soc.* **2005**, *127*, 14825–14832.

(45) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435–13441.

(46) Yoda, T.; Sugita, Y.; Okamoto, Y. *Proteins* **2007**, *66*, 846–859.

(47) Bonomi, M.; Branduardi, D.; Gervasio, F. L.; Parrinello, M. *J. Am. Chem. Soc.* **2008**, *130*, 13938–13944.

(48) Snow, C. D.; Zagrovic, B.; Pande, V. S. *J. Am. Chem. Soc.* **2002**, *124*, 14548–14549.

(49) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587–7592.

(50) Zhou, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280–13285.

(51) Schug, A.; Herges, T.; Wenzel, W. *Phys. Rev. Lett.* **2003**, *91*, 158102.

(52) Chowdhury, S.; Lee, M. C.; Xiong, G.; Duan, Y. *J. Mol. Biol.* **2003**, *327*, 711–717.

(53) Chowdhury, S.; Lee, M. C.; Duan, Y. *J. Phys. Chem. B* **2004**, *108*, 13855–13865.

(54) Juraszek, J.; Bolhuis, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15859–15864.

(55) Paschek, D.; Nymeyer, H.; Garcia, A. E. *J. Struct. Biol.* **2007**, *157*, 524–533.

(56) Paschek, D.; Hempel, S.; Garcia, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17754–17759.

(57) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. *PLoS Comput. Biol.* **2009**, *5*, e1000452.

(58) Chebaro, Y.; Dong, X.; Laghaei, R.; Derremaux, P.; Mousseau, N. *J. Phys. Chem. B* **2009**, *113*, 267–274.

(59) Kannan, S.; Zacharias, M. *Proteins* **2009**, *76*, 448–460.

(60) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(61) Munoz, V.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5872–5879.

(62) Du, D.; Zhu, Y.; Huang, C.-Y.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15915–15920.

(63) Du, D.; Tucker, M. J.; Gai, F. *Biochemistry* **2006**, *45*, 2668–2678.

(64) Huang, X.; Bowman, G. R.; Pande, V. S. *J. Chem. Phys.* **2008**, *128*, 205106.

(65) Nettels, D.; Müller-Spath, S.; Küster, F.; Hofmann, H.; Haenni, D.; Rüegger, S.; Reymond, L.; Hoffmann, A.; Kubelka, J.; Heinz, B.; Gast, K.; Best, R. B.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 20740–20745.

(66) Best, R. B.; Hummer, G. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 1088–1093.

(67) Rhee, Y. M.; Pande, V. S. *Biophys. J.* **2003**, *84*, 775–786.

(68) Sakae, Y.; Okamoto, Y. *Chem. Phys. Lett.* **2003**, *382*, 626–636.

(69) Sakae, Y.; Okamoto, Y. *J. Theor. Comp. Chem.* **2004**, *3*, 339–358.

(70) Sakae, Y.; Okamoto, Y. *J. Theor. Comp. Chem.* **2004**, *3*, 359–378.

(71) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. *J. Biomol. NMR* **2003**, *26*, 215–240.

(72) Shen, Y.; Bax, A. *J. Biomol. NMR* **2007**, *38*, 289–302.

(73) Kohlhoff, K. J.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 13894–13895.

(74) Schuler, B.; Lipman, E. A.; Steinbach, P. J.; Kumke, M.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2754–2759.

(75) Best, R. B.; Merchant, K. A.; Gopich, I. V.; Schuler, B.; Bax, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18964–18969.

(76) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. *J. Chem. Phys.* **1996**, *104*, 5860–5868.

(77) Best, R. B.; Hummer, G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6732–6737.

(78) Kohn, J. E.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12491–12496.

(79) Fitzkee, N. C.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12497–12502.

(80) Mok, K. H.; Kuhn, L. T.; Goez, M.; Day, I. J.; Lin, J. C.; Andersen, N. H.; Hore, P. J. *Nature* **2007**, *447*, 106–109.

(81) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

(82) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.

(83) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(84) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(85) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926–935.

(86) Olsen, K. A.; Fesinmeyer, R. M.; Stewart, J. M.; Andersen, N. H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15483–15487.

(87) Munoz, V.; Serrano, L. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *20*, 301–311.

(88) Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. *J. Am. Chem. Soc.* **2004**, *126*, 7238–7243.

(89) Gin, B. C.; Garahan, J. P.; Geissler, P. L. *J. Mol. Biol.* **2009**, *392*, 1303–1314.