Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse

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

Cambridge University, Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, United Kingdom, and Lehigh University, Department of Chemical Engineering, Bethlehem, Pennsylvania 18015, United States

Received: September 9, 2010; Revised Manuscript Received: October 12, 2010

A recognized shortcoming in current protein simulations is that most force fields are parametrized with relatively primitive three-site water models. Since the deficiencies of the common three-site water models in reproducing the phase diagram of water are well-known, an improved description of the solvent will be required, for example, to study proteins in molecular simulations at thermodynamic conditions other than standard temperature and pressure. Here, we combine a protein force field derived from Amber ff03 together with the highly optimized TIP4P/2005 water model, with a small backbone modification to match the population of helical states obtained with the new water model to experiment. Remarkably, we find that the resulting force field, Amber ff03w, produces a more cooperative helix−coil transition, compared with the similarly "backbone-corrected" Amber ff03* model with TIP3P water, with calculated helix propagation parameters in good agreement with the experiment. The radius of gyration for nonhelical conformations is significantly larger for Amber ff03w than for Amber ff03* and shows a collapse with increasing temperature as found in single-molecule experiments on longer proteins. The origin of the collapse appears to be a more favorable enthalpic component of the peptide−solvent interaction and is correlated with increasing turn formation, in accord with the experiment. In addition to this enhanced cooperativity, we verify that, with the new force field, replica exchange folding simulations of the GB1 hairpin and Trp cage result in folded structures, starting from completely unfolded initial conditions; simulations of folded proteins are also stable. These results together suggest that Amber ff03w (with TIP4P/2005) will be well suited for studying protein folding and properties of unfolded state and intrinsically disordered proteins over a wide range of thermodynamic conditions.

Introduction

A large contribution to the stability of proteins comes from hydrophobic interactions.[1−4] Therefore, a faithful description of biomolecular self-assembly processes such as protein folding and binding, aggregation, and micellization in molecular simulations must include an accurate description of water behavior over the range of conditions for which these processes are observed. Currently, most protein force fields are optimized for reproducing protein properties under conditions of standard temperature and pressure.[5−10] For such applications, the commonly used three-site water models, TIP3P,[11] SPC,[12] and SPC/E[13] are adequate. However, there is much interest in understanding protein behavior at low temperatures and high pressures, where interesting phenomena occur because of changes in solvent properties and solvent−protein interactions.[14−21]

While sufficient for simulations at standard conditions, the simple three-site water models (TIP3P, SPC) reproduce neither the temperature-dependence of pure water properties nor hydrophobic hydration even for simple solutes.[22−25] In the last five years, however, there has been remarkable progress in optimizing water models to reproduce experimental observations.[22,26] In fact a recent variant of four-point water models, TIP4P/2005,[26] can reproduce water's phase diagram at low temperature and high pressure, the dependence of density on temperature and specific heat,[27−31] and the hydration free energy of methane.[25,32] Indeed, the broad range of properties accurately predicted by this model suggests that it is close to the optimum that can be achieved with nonpolarizable models. It seems appropriate to use protein force fields in conjunction with this water model to improve their predictive power, particularly at state points which differ significantly from standard temperature and pressure.

Of course, the use of an existing force field with a nonstandard water model may require some fine-tuning in the energy function parameters to account for changes in the solvent properties. The justification for combining a protein force field with different water models depends to some extent on how closely the original water model was integrated with the force-field development. For example, in the case of the CHARMM force field, partial charges are explicitly optimized to match binding energies with TIP3P in various geometries;[6] for GROMOS, the parameters were optimized to reproduce solvation free energies in SPC water.[9] In these cases a new water model would clearly have to be included in similar calculations and the protein parameters altered, as necessary. On the other hand, the Amber force field charges were fit to reproduce a quantum mechanically derived electrostatic potential. The effects of polarization in the condensed phase are included either implicitly by using gas phase HF/6-31G* calculations which overestimate the polarization, or, in the case of Amber ff03, using more sophisticated calculations with an implicit solvent.[8] Since a particular water model is not considered explicitly, altering the type of water is more easily justified. On the other hand, the choice of water

* To whom correspondence should be addressed. E-mail: jeetain@lehigh.edu; rbb24@cam.ac.uk.
† Cambridge University.
‡ Lehigh University.

10.1021/jp108618d
Published on Web 11/01/2010

model will alter the properties of the protein,[33] and so it is necessary to validate the results.

Our recent work with Amber ff03*,[34] a modified version of Amber ff03,[8] has shown that a simple correction to backbone dihedral potentials can result in a more transferable force field able to fold proteins from α (Trp cage, Villin) and β (GB1 hairpin, Pin WW domain) structural classes.[35,36] We therefore use this as the starting point. We use experimental data to reparameterize the backbone dihedral potential correction for Amber ff03* so that the fraction helix for the 15 residue peptide Ac-$(AAQAA)_3$-NH_2 is correctly reproduced in the optimized TIP4P/2005 water.[26] In this modified force field, which we refer to as Amber ff03w, the balance among α, β, and PPII structures is found to be in agreement with scalar couplings measured for the short peptide, Ala_5. We have investigated the effect of the water model on the helix−coil transition, a fundamental element of protein folding.[37−49] We find that the cooperativity of the helix−coil transition of the helix-forming peptide Ac-$(AAQAA)_3$-NH_2 with Amber ff03w is much improved compared to Amber ff03* and the other fixed-charge force fields we have previously studied in TIP3P water.[34]

The temperature dependence of the helix propagation parameter w from a fit to the Lifson−Roig (LR) model for the helix−coil transition is in good agreement with experimental data. The nucleation parameter v is still somewhat larger than in the experiment, although there is also uncertainty in the experimental value. To obtain a direct comparison with the experiment, we have also computed the temperature-dependent carbonyl chemical shifts for the same peptide using a recent empirical shift-prediction algorithm, SPARTA+.[50] The results are in good agreement with the experimental curves at all measured temperatures.

Amber ff03w results in an expanded unfolded state (nonhelical conformations) for Ac-$(AAQAA)_3$-NH_2 compared to its parent ff03*, which is also manifested in a greater number of hydrogen bonds to solvent than within the peptide. Furthermore, for this peptide, we observe unfolded state collapse with increasing T as seen in recent single-molecule experiments on larger proteins.[51] Enthalpic and entropic contributions obtained from a fit of the temperature-dependence of the Flory interaction parameter to a thermodynamic model suggests that the transfer enthalpy for the unfolded peptide is more favorable in TIP4P/2005 water than in TIP3P. To evaluate the suitability of the force field for more general applications, we have run long simulations of folded ubiquitin and replica exchange molecular dynamics (REMD) folding simulations of Trp Cage and the GB1 hairpin. Residual dipolar couplings (RDCs) and generalized order parameters calculated from the ubiquitin simulations are in good agreement with the experiment, while both Trp cage and GB1 fold to their native structures, starting from completely unfolded initial conditions. Our results suggest that Amber ff03w together with TIP4P/2005 gives a good description of protein folding thermodynamics and a more accurate description of the unfolded state. This opens the way for investigating folding and the properties of unfolded states under a much wider range of thermodynamic conditions, facilitating the comparison with experiments.[51,52]

Results and Discussion

Parameter Fitting. As a starting protein force field, we use Amber ff03,[8] which compared favorably to NMR scalar coupling data for short alanine peptides.[53] Simulations with the same force field with TIP3P water indicated too high a helical propensity

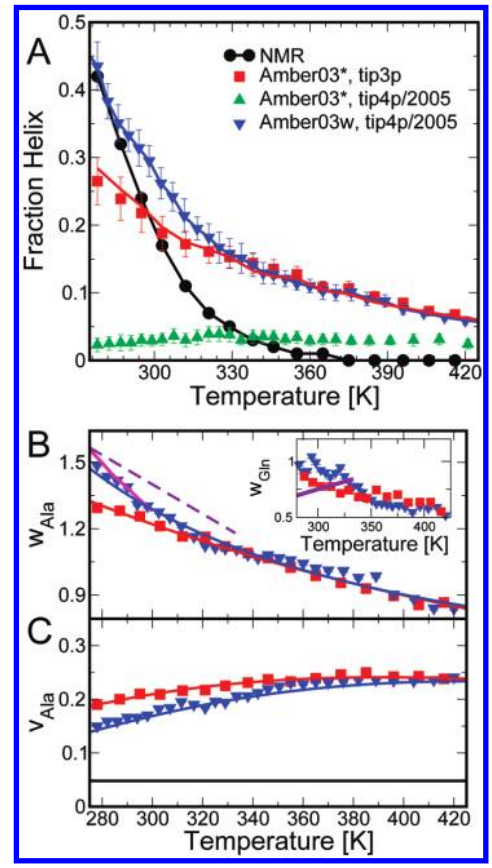

Figure 1. Thermodynamics of the helix−coil transition. (A) The fraction helix is shown as a function of temperature, as determined from REMD simulations. The temperature-dependent Lifson−Roig (LR) parameters[38] (B) w_{Ala} describing helix propagation and (C) v_{Ala} describing helix nucleation are shown below. The solid magenta and broken violet lines in B represent w_{Ala} determined experimentally by Rohl and Baldwin[60] and Moreau et al.,[49] respectively. The inset in B shows the helix propagation parameter w_{Gln} for Gln (symbols as in the main figure). The solid lines through the simulation data in A represent the fraction helix calculated from the LR model. In B and C the solid lines give the fit of a thermodynamic model to the LR parameters (see text).

in longer peptides, motivating a simple cosine correction $V(\psi)$ to the ψ torsion angle using,

$$V(\psi;k_\psi,\delta_\psi) = k_\psi[1 + \cos(\psi - \delta_\psi)] \qquad (1)$$

We refer to the optimal parameter values of $k_\psi = 1.5$ kJ/mol and $\delta_\psi = 285.5°$ as Amber ff03*.[34] The target experimental data in that optimization were scalar couplings measured for Ala_5 and the fraction of helix in a 15 residue helix-forming peptide, Ac-$(AAQAA)_3$-NH_2, at 300 K using the TIP3P water model. We apply the same methodology to modify the backbone potential for use with TIP4P/2005. Initial simulations of Ac-$(AAQAA)_3$-NH_2 with Amber ff03* and TIP4P/2005 water (Figure 1) indicate that the new water model destabilizes the helix: the helix population is too low at all temperatures relative to the experimental curve derived from NMR chemical shifts.[54] To adjust the helical propensity of Ac-$(AAQAA)_3$-NH_2, we modify the parameters k_ψ in eq 1 in an iterative procedure, finding optimal values of $k_\psi = 0.75$ kJ/mol, $\delta_\psi = 285.5°$ (Figure 1). We have also checked that simulations of Ala_5 with a protonated C-terminus are consistent with NMR scalar coupling data.[55] We quantify the agreement with experiment, using a parameter χ^2, defined as $\chi^2 = N^{-1}\sum_{i=1}^{N}[(J_{i,calc} - J_{i,meas})^2/\sigma_i^2]$, where

the sum runs over the N scalar couplings $J_{i,\mathrm{calc}}$ calculated from simulation and $J_{i,\mathrm{meas}}$ observed in experiment; the σ_i are estimates of the uncertainty in the scalar couplings calculated with the Karplus equation, as previously described.[34,53] Using the "DFT2" set of Karplus equation parameters from Case and Brüschweiler[56] for calculating the scalar couplings from simulation, we obtain a χ^2 value of 0.9, that is, the deviation is comparable to the uncertainty in the predicted $J_{i,\mathrm{calc}}$. At this point we should indicate why so much emphasis is placed on "tuning" the backbone, given the thousands of other parameters in the force field. While it is certainly important that the errors in the other parameters are small, errors in the backbone secondary structure propensity have a cumulative effect since they apply to every residue. For example, if the free energy of the α minimum relative to other states in the Ramachandran map is altered by $0.2k_BT$, this will result in a change of $\sim 3k_BT$ in the energy of the fully helical state of Ac-$(AAQAA)_3$-NH_2. A better balance between α and β secondary structures has been shown to be advantageous for folding simulations in the explicit solvent[35,36] and implicit solvent.[57−59] Matching the fraction helix to the experiment also allows us to make useful comparisons between simulations in TIP3P and TIP4P/2005, which would otherwise be difficult (e.g., if no helix is formed at all). The same correction is assumed to apply to all residues except for Gly and Pro, since these are sufficiently different that a separate optimization would be necessary. We have taken a minimalist approach to tuning the force field for the water model by changing only the backbone; this can be partially motivated by the fact that the specific water model was not considered in the initial protein force field parametrization. Nonetheless, other parameters in the protein force field may certainly be refined against experiment, in particular the nonbonded parameters. In the future it may be worthwhile to optimize the nonbonded parameters to match solvation thermodynamics, as is done in the GROMOS force fields,[9] or to reproduce binding energies with water molecules to tune the peptide−water interactions, as is done in the CHARMM force field.[6]

Helix−Coil Transition. Alanine-based helical peptides[40,41] have served as useful benchmarks for both experiments[61,62] and simulations[34,47,48] in protein folding. One of the deficiencies of Amber ff03* identified in previous work was the weak temperature dependence of helix formation, reflecting a lower cooperativity in the helix−coil transition than inferred experimentally (Figure 1A). The data for Amber ff03w in Figure 1 suggest that the cooperativity of helix formation in this force field may be much improved. To check this at a microscopic level, we fit our simulation data to a LR model for the helix−coil transition.[38] This model describes helix formation via a partition function defined by parameters v for helix nucleation and w for helix elongation (see Methods). The optimal LR parameters are derived by directly fitting the partition function to the observations from simulation using a Bayesian approach as described before.[34] In the model, we have used separate parameters for Ala and Gln but have not included specific capping effects.

Figure 1B shows the fitted helix nucleation parameter for Ala v_{Ala} and helix propagation parameter w_{Ala} as a function of temperature. As noted in a previous work, w_{Ala} obtained with Amber ff03* and TIP3P water shows too weak a dependence on temperature; fits to a thermodynamic model indicated that this arises from the change in enthalpy and entropy associated with adding a helical residue being about half of the values estimated from experimentally determined w.[34] We find that Amber ff03w with TIP4P/2005 results in a w_{Ala} parameter whose temperature dependence is in better agreement with the available experimental estimates[60] (Figure 1B). The fit of a thermodynamic model to the $w(T)$ for ff03w (Table 1) gives ΔH and ΔS which are larger than for ff03* and within the range of values estimated from the experiment. We note that there are differences between the w parameters estimated by different experimental groups which may be due to subtle differences in data analysis, and a direct comparison of observables computed from simulation (e.g., chemical shifts, scalar couplings) with the experiment is the ultimate test. Although there are only three Gln residues in our peptide, we do also obtain an estimate for w_{Gln} (Figure 1B, inset). In agreement with the experiment, this is smaller than w_{Ala}, although we do not find the increase in w_{Gln} with temperature observed experimentally[49] for either force field. Interestingly, there is only a slight difference in the nucleation parameters v_{Ala} between Amber ff03* and ff03w, and in neither case is there a good match with the commonly used experimental values of ~ 0.05.[63,64] Although there is still some uncertainty in the determination of v from experimental data, as noted previously by several groups,[61,63,65] this suggests that the entropic penalty for nucleating helix formation remains too small. We also note that one of us did not find a significant difference in the cooperativity of helix formation in Ac-$(AAQAA)_3$-NH_2 using a combination of Amber ff03 with TIP4P-Ew.[22] It is most likely due to the fact that, even with the altered water model, the transition midpoint of Amber ff03 is still 80 K higher than the experiment, and a backbone correction is needed to correct for this.

TABLE 1: Thermodynamic Parameters for Helix−Coil Transition[a]

parameter	ΔH_0 kJ·mol^{-1}	ΔS_0 J·mol^{-1}·K^{-1}	ΔC_p J·mol^{-1}·K^{-1}
w_{Ala}			
ff03*	−2.4 (0.1)	−6.4 (0.4)	−13.4 (1.2)
ff03w	−3.7 (0.1)	−10.1 (0.3)	2.4 (2.7)
Rohl and Baldwin[60]	−5.4	−15.7	
Moreau et al.[49]	−3.7	−9.7	
v_{Ala}			
ff03*	2.8 (0.1)	−3.8 (0.4)	−29.9 (−1.1)
ff03w	4.9 (0.3)	1.4 (1.0)	−38.4 (6.1)
Rohl et al.[63]		−24.9 (1.0)	

[a] Note that the parameters for ff03* differ slightly from published values[34] due to the separate treatment of Ala and Gln in the present work.

An influence on the cooperativity which we have not considered is the temperature dependence of the water density, since we use constant volume simulations. Sampling at constant volume was motivated first by the desire to make a direct comparison with the ff03* force field and second by results from an earlier $V-T$ REMD study of helix formation suggesting that the change in density had a small effect.[20] Our own earlier work on the helix−coil transition[34] also showed a weak dependence of helicity on density. Nonetheless, it will be important in future studies, to sample with constant pressure molecular dynamics or a two-dimensional $V-T$ REMD scheme to capture the variation of density with temperature with the more accurate water model.[20,66]

Chemical Shifts. To avoid the aforesaid uncertainty when comparing simulation results indirectly with models fitted to experimental data, we calculate experimental observables to facilitate a direct comparison. Specifically, we have derived chemical shifts from our simulated structures using the empirical SPARTA+ algorithm[50] from the Bax group. Figure 2 shows the comparison between the available experimental data (black

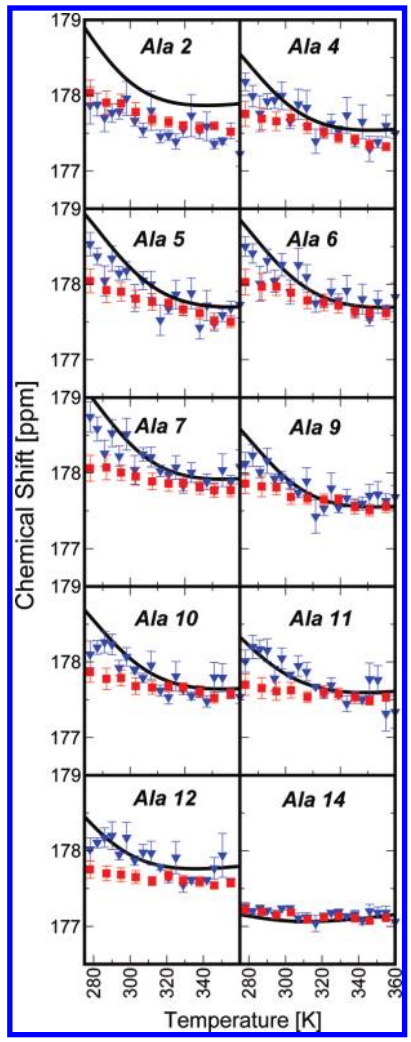

Figure 2. Direct comparison with experimentally observed chemical shifts. Temperature-dependent experimental chemical shifts (black lines)[54] are compared with simulated ensembles for Amber03* (with TIP3P; red ■) and Amber03w (with TIP4P/2005; blue ▼). We have used SPARTA+ to obtain chemical shifts from simulated peptide structures.[50]

lines) as a function of temperature for Ala residues with Amber03* (TIP3P; red squares) and Amber03w (TIP4P/2005; blue down triangles). Overall, we find a very good match between our simulation results and the experiment and a significant improvement in the temperature dependence with Amber03w. The noise in the simulation results arises because the reported values are averages of very broad distributions of predicted shifts. The discrepancy for Ala2 is probably explained by the fact that the shift prediction algorithm does not account for the end-capping residues (Ace- and-NH_2). We also note that, while the method used to parametrize the chemical shift prediction is purely empirical, similar empirical approaches are often necessary to obtain parameters in predictions of other observables (e.g., scalar couplings from a Karplus equation). The accuracy of the shift prediction schemes is now clearly at a level where they can be used for quantitative comparison of simulation with experiment.

Properties of the Unfolded Chain. Water is known to have an effect on the propensity of helix formation,[67] due to both hydrophobic and hydrophilic interactions with the peptide.[68] To investigate how the water model may affect the properties of the peptide, we calculate the number of hydrogen bonds between different components of the system: peptide−peptide, N_{PP},

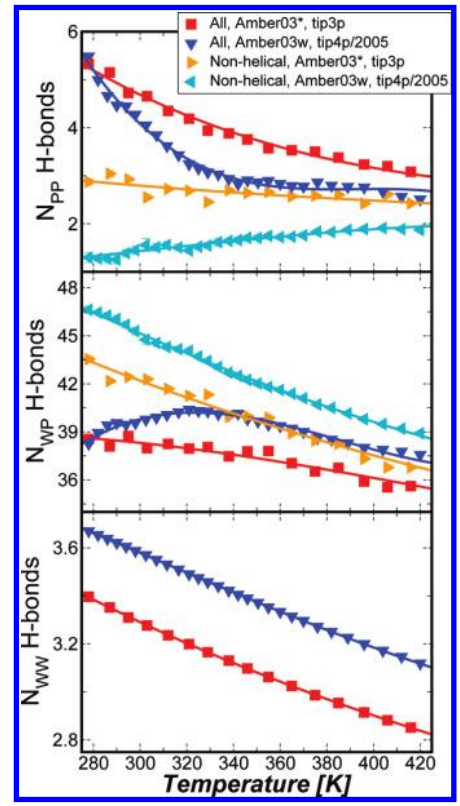

Figure 3. Hydrogen bonding. The average number of hydrogen bonds between peptide−peptide (PP: upper panel), water−peptide (WP: middle panel), and water−water (WW: lower panel) is shown as a function of temperature. Symbols are the simulation data, and lines are guides for the eye (cubic fits).

water−peptide, N_{WP}, and water−water, N_{WW}, atoms (Figure 3). Hydrogen bonds are defined as having a donor−acceptor distance within 0.35 nm and a donor-hydrogen-acceptor angle less than 30°. Figure 3A shows that PP H-bonds appear to be lost more rapidly in TIP4P/2005 water than in TIP3P water; however, this clearly has a large contribution from the more cooperative helix formation in the former. Therefore, we have repeated the analysis of the average number of hydrogen bonds, but considering nonhelical conformations only—defined as those containing zero helical segments. We find that N_{PP} averaged only over nonhelical structures is lower and *increases* with T for TIP4P/2005, compared with a decrease for TIP3P. The larger number of water−protein hydrogen bonds, N_{WP}, (Figure 3B) for nonhelical states suggests that the peptide in TIP4P/2005 water may be less hydrophobic than in TIP3P water. We also observe a maximum in the total number $N_{WP}(T)$ for TIP4P/2005 water as a result of two effects: an increase at low temperature which compensates the decrease in $N_{PP}(T)$ as the helix unfolds, and a decrease at high T when states with fewer hydrogen bonds become entropically favored. The larger number of water−water hydrogen bonds for TIP4P/2005 compared with TIP3P is consistent with the stronger hydrogen bonding in the former model.[30]

The differences between TIP3P and TIP4P/2005 appear to be related to a more favorable enthalpy of solvation. The hydrogen bond analysis suggests that this arises partly from a stronger interaction with the solvent. Interestingly, a recent study calculated hydration enthalpies and entropies for amino acid side-chain analogues using a number of different combinations of force fields and water models.[69] While the various combinations give almost identical hydration free energies, there are significant differences in the enthalpic and entropic contribu-

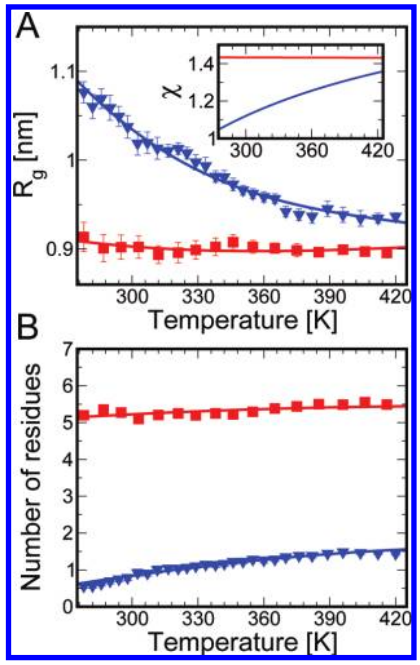

Figure 4. Protein size. (A) The radius of gyration averaged over nonhelical conformations (defined as zero helical segments) is shown as a function of temperature. Symbols are the simulation data, and lines are a fit to an adaptation of the model in ref 51. The fit yields thermodynamic parameters for the Flory interaction parameter χ of $\Delta H_\chi = -1.45$ (0.07) kJ·mol^{-1} and $\Delta S_\chi = -11.53$ (0.19) J·mol^{-1}·K^{-1} for ff03w and $\Delta H_\chi = 0.01$ (0.05) kJ·mol^{-1} and $\Delta S_\chi = -8.53$ (0.15) J·mol^{-1}·K^{-1} for ff03* (see text). The Flory parameter χ itself is shown in the inset as a function of temperature. (B) Number of residues which are structured, based on DSSP criteria, as a function of temperature.

tions. In that case, it was found that TIP4P-Ew, an optimized four-site model similar to TIP4P/2005, resulted in solvation enthalpies between 3−10 kJ·mol^{-1} more favorable than those obtained with TIP3P, particularly for hydrophobic residues.

The observation of increasing PP H-bonds with increasing temperature for nonhelical peptide conformations in TIP4P/2005 suggests that the peptide may actually be becoming more compact. A collapse in the unfolded state of proteins with temperature has previously been observed using Förster resonance energy transfer (FRET) in single molecule[51] and ensemble[70] experiments and is correlated with an increase in secondary structure content. In fact combining Amber ff03* with the TIP4P/Ew water model in previous simulations on cold shock protein did show a temperature-induced collapse, but the overall size of the protein was found to be much smaller than in the experiment.[51] Figure 4 shows the radius of gyration, R_g, as a function of temperature averaged over all of the nonhelical conformations for both Amber ff03* (with TIP3P) and Amber ff03w (with TIP4P/2005). Although the R_g of the 15 residue Ac-$(AAQAA)_3$-NH_2 peptide has not been measured, we can compare it to the value of 0.98 nm obtained for the 16 residue alanine-based peptide AK-16 in 4 M urea at room temperature.[71] This is in reasonable agreement with the 1.02 nm obtained with Amber ff03w.

To characterize the collapse of the chain with temperature, we use a simple polymer model to describe the effect of the solvent on the radius of gyration: the Flory−Huggins theory gives the radius of gyration of a polymer of n residues, each with volume ν, as[72]

$$R_g = \left[\frac{2n\nu}{3(2\chi - 1)}\right]^{1/3} \quad (2)$$

where χ, the Flory parameter, measures the affinity of monomers for each other relative to the solvent (specifically, the free energy for tranferring a monomer from a hypothetical pure monomer phase to solvent and for transferring a solvent molecule from a pure solvent phase to the monomer phase). The temperature dependence of χ can be described by a thermodynamic model,

$$\chi(T) = \frac{1.4}{k_B T}[\Delta H_\chi - T(\Delta S_\chi)] \quad (3)$$

where ΔH_χ and ΔS_χ are the enthalpy and entropy changes associated with the solvent−monomer exchange. We fitted 2 and 3 to the data in Figure 4 using the volume of a sphere of radius 3.2 Å for the average residue volume (most of the residues are Ala); the data were not sufficient to determine a separate heat capacity. We find for ff03w a favorable enthalpy change, $\Delta H_\chi = -1.4$ (0.1) kJ·mol^{-1}, and unfavorable entropy change, $\Delta S_\chi = -11.5$ (0.2) J·mol^{-1}·K^{-1}, which is in reasonable agreement with values estimated from the collapse of proteins using a similar model ($\Delta H_\chi = -1.3$ kJ·mol^{-1}, $\Delta S_\chi = -8.5$ J·mol^{-1}·K^{-1}).[51] For ff03*, the entropy change is similar ($\Delta S_\chi = -8.53$ (0.15) J·mol^{-1}·K^{-1}), but there is essentially no enthalpy difference ($\Delta H_\chi = 0.01$ (0.05) kJ·mol^{-1}). This explains at a thermodynamic level why the unfolded peptide is collapsed at lower temperatures for TIP3P, suggesting that the differences between TIP3P and TIP4P/2005 are largely due to the more favorable enthalpy of solvation.

Since an increase in the secondary structure with temperature was observed in the experiment,[51,70] we also calculate the average number of residues in the coil and structured (beta + turn + helix + bridge) states based on DSSP criteria,[73] considering only nonhelical conformations. Figure 4B shows that, for Amber ff03* with TIP3P, there is almost no variation with temperature. The large total number of structured residues may be related to the observation of many structured misfolded states in simulations of the GB1 hairpin with this energy function.[35] For Amber ff03w (with TIP4P/2005), the number of structured residues increases with T, concomitant with peptide collapse, consistent with experimental observations on other proteins. A distinction in the types of secondary structure observed is that we find a nonzero population (∼1−3 residues) in bend states for Amber ff03w (with TIP4P/2005) but zero population with Amber ff03* (with TIP3P).

Transferability. Since we have so far focused primarily on the helix−coil transition and the unfolded state behavior, it is important to test the proposed combination of protein force field and water model on other problems. First, we use Amber ff03w with the TIP4P/2005 water model for ab initio protein folding simulations of two miniproteins belonging to prototypical α and β structural classes, Trp cage and GB1 hairpin, respectively. We have run 150 ns REMD simulations starting from an initial condition of completely unfolded proteins in all of the replicas. Although we did not run the simulations for long enough to reach converged equilibrium populations, we observe folding events in both of the cases, reaching structures with a backbone root-mean-square deviation (rmsd) from the native structure of 0.5 Å. In Figure 5 we present folded structures from REMD simulations for both the proteins overlaid with the native pdb structures. This suggests that the newly optimized force field should be transferable to protein folding, including proteins from different structural classes, as was found for ff03*.[35,36] A more detailed analysis of miniprotein data will be provided in a future publication.

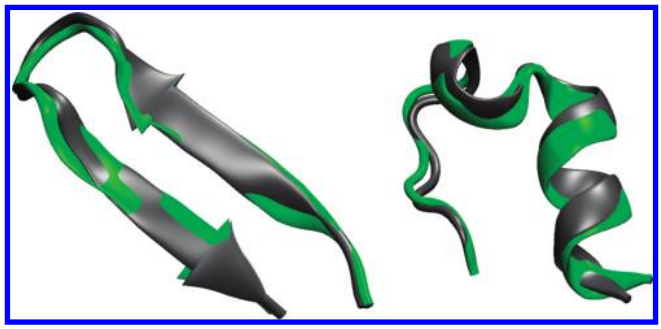

Figure 5. Transferability. Folded structures for the GB1 hairpin (left) and Trp cage (right) obtained from REMD simulations (gray) starting from completely unfolded state in all of the replicas are compared with the native structure from the PDB database (green).

TABLE 2: *Q*-Values for RDCs[a]

RDC set	ff03 (30 ns)	ff03* (30 ns)	ff03w (50 ns)
Hus et al.[76] Q_{ave}	0.28	0.26	0.29
Cornilescu et al.[77] Q_{methyl} (a)	0.59	0.56	0.55
Cornilescu et al.[77] Q_{methyl} (b)	0.31	0.27	0.31
Ottiger and Bax[78] Q_{methyl}	0.75	0.76	0.66

[a] Q_{ave} is the sum over Q for backbone amide RDCs in 10 different alignment media recorded by Hus et al.;[76] results are shown separately for the two alignment media used by Cornilescu et al.[77]

We have also tested the ability of the new force field combination to reproduce the properties of folded proteins, by running a 50 ns simulation of folded ubiquitin. We find a low backbone rmsd for the structured residues (1−72) of less than 1 Å. However, this is a somewhat arbitrary and not particularly discriminating criterion for evaluating force fields. We have therefore also calculated RDCs, a very sensitive probe of relative bond vector orientations,[74] from the simulations. By assuming that the contributions to the dipolar coupling from intramolecular motion and molecular tumbling can be factorized, we calculate the dipolar couplings using an established ensemble-averaging procedure.[34,75] The agreement between the experimental and the calculated RDCs is quantified by the Q value, defined as

$$Q = \left[\frac{\sum_i (D_{i,\mathrm{calc}} - D_{i,\mathrm{meas}})^2}{\sum_i (D_{i,\mathrm{meas}})^2} \right]^{1/2} \quad (4)$$

where the sum runs over the calculated and experimentally measured dipolar couplings, $D_{i,\mathrm{calc}}$ and $D_{i,\mathrm{meas}}$, respectively. Q values for dipolar couplings for different types of bond vectors, measured in different alignment media, are summarized in Table 2 for native state simulations of ubiquitin with Amber ff03 and ff03* in TIP3P water and for Amber ff03w in TIP4P/2005. The Q_{ave}, which is the first entry in the table, is an average over Q-values for backbone NH bonds obtained in 10 different alignment media. The Q_{ave} for Amber ff03w is comparable to the values obtained for the other two force fields. We also report Q values for side-chain methyl groups measured by several groups. Here also, the results for ff03w are comparable to previous simulations and possibly a little better in one case. Overall, the strength of Amber ff03w in combination with TIP4P/2005 appears to be its representation of the unfolded state, but these results demonstrate that it also gives a good description of folded proteins.

Conclusions

Simulations of proteins with atomistic force fields have proved to be remarkably accurate and able in several cases to fold proteins ab initio. From a folding perspective, the main deficiencies in existing energy functions pertain to the unfolded state (too collapsed) and the weak cooperativity and temperature dependence of folding. Neither of these aspects was a priority in the development of current force fields with three-site water models, and these are certainly adequate for simulations of folded proteins under standard conditions. However, if unfolded states, or intrinsically unstructured proteins, are of interest, it will be important to have a more accurate description of the solvation of the unfolded chain. Similarly, to study state points other than standard temperature and pressure, a solvent model is needed which accurately reproduces the phase diagram of water. We have shown that the combination of a modified version of Amber ff03 and Amber ff03w, with the recent TIP4P/2005 water model, provides an improved description of unfolded state properties and of the thermodynamics of the helix−coil transition. It is therefore likely that this model will be a good starting point for exploration of the folding under nonstandard conditions, including unfolded or unstructured proteins.

Methods

Simulation Details. REMD simulations of the blocked peptide Ac-$(AAQAA)_3$-NH_2 were run using the simulation package Gromacs[79,80] using a protocol similar to that used in our previous work[53] and with the implementation of the Amber force fields by Sorin and Pande.[81] We use 40 replicas spanning a temperature range of 250−452 K using Gromacs.[79,80] The peptide was solvated in a truncated octahedron simulation cell of 2186 TIP4P/2005 water molecules with an initial distance of 45 Å between the nearest faces of the cell. This cell was equilibrated for 500 ps at 300 K and a constant pressure of 1 atm, following which all REMD simulations were done at constant volume, with long-range electrostatics calculated using particle mesh Ewald (PME) with a 1.2 Å grid spacing and 9 Å cutoff. Dynamics was propagated with a Langevin integration algorithm using a friction of 1 ps^{-1} and replica exchange attempts every 10 ps (every 5000 steps with a time step of 2 fs). All replica exchange runs used the same set of initial configurations. The simulations were run for at least 200 ns per replica, of which the first 30 ns were discarded (an aggregate of ≈8 μs). The data for Ac-$(AAQAA)_3$-NH_2 solvated in TIP3P water and Amber ff03* were taken from previous work.

Simulations of Ala_5 were run using a similar protocol with 28 replicas spanning a temperature range of 278−421 K. The peptide was unblocked and protonated at both N and C termini, corresponding to the experimental conditions of pH 2.[55] Molecular dynamics simulations of the peptide in a 30 Å cubic simulation box of explicit TIP4P/2005 water[11] were performed for 50 ns per replica, of which the first 10 ns were discarded. REMD simulations of GB1 hairpin and Trp cage utilized 32 and 36 replicas, respectively, spanning a temperature range of 278−595 K (see our previous paper for more details[35]).

Simulations of folded ubiquitin with Amber03 and Amber03* in TIP3P water were as reported previously.[34] Simulations of ubiquitin with Amber03w in TIP4P/2005 were performed using a Langevin integrator (friction of 1 ps^{-1}) and a Parrinello−Rahman barostat (time constant of 5 ps) at 300 K temperature and 1 bar pressure. The protein (PDB ID: 1ubq) is solvated in a truncated octahedron box with 2750 TIP4P/2005 water molecules, 9 sodium ions, and 10 chloride ions. Other details

such as the treatment of electrostatics and cut-offs are exactly the same as described previously for Ac-$(AAQAA)_3$-NH_2.

Definition of Helical States. We define helical structure in terms of which regions of the Ramachandran map are occupied. The α_h region of the (ϕ, ψ) map is defined as[48] $\phi \in [-100°, -30°]$ and $\psi \in [-67°, -7°]$. This definition is intended to capture the narrower (ϕ, ψ)-angle distribution of residues within actual α-helices. Residues which lie within the α_h region of the Ramachandran map are denoted as helical (h) for the purposes of LR analysis. All residues outside the α_h region are defined as "coil" (c). A helical segment is one which has at least three consecutive residues whose (ϕ, ψ) angles fall within the α_h boundaries (i.e., the smallest helix is chhhc); since the peptide Ac-$(AAQAA)_3$-NH_2 is blocked, all (ϕ, ψ) pairs are properly defined. The fraction of helix for a given residue in the simulation $\langle h_i \rangle_{sim}$ is calculated as the fraction of time spent by that residue within helical segments.

Calculation of Helix−Coil Parameters. We have analyzed the data using the LR model for the helix−coil transition. In this model, the statistical weight of a given conformation is the product of weights for individual residues. A residue in the coil conformation is assigned a relative weight 1, and a residue of type i in a helical conformation, but not within a helical segment, is assigned a weight v_i. Residues in helical segments receive a weight of w_i, except for the terminal residues in the segment which have weight v_i (since usually $v \ll w$ this creates a "nucleation penalty" for initial helix formation). The subscripts i indicate that each of these parameters is potentially dependent on the residue types concerned. The partition function of this model for a peptide with N amino acid residues can be written in a compact matrix form as:[39,42]

$$Z = (0 \quad 0 \quad 1) \prod_{i=1}^{N} \mathbf{M}_i \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \tag{5}$$

in which the matrix $\mathbf{M}_i$ is defined as:

$$\mathbf{M}_i = \begin{pmatrix} w_i & v_i & 0 \\ 0 & 0 & 1 \\ v_i & v_i & 1 \end{pmatrix} \tag{6}$$

The average "fraction of helix" for a given residue i is defined as the relative population of conformations where it has weight w_i, given by $\langle h_i \rangle_{LR} = \partial \ln Z / \partial \ln w_i$. The average fraction of helix for the whole peptide, $\langle f_h \rangle_{LR}$, is the average over these for the nonterminal residues $\langle f_h \rangle_{LR} = (N-2)^{-1} \sum_{i=2}^{N-1} \langle h_i \rangle$ (since terminal residues cannot have a w weight by definition). We have determined the posterior distribution of most probable parameters w and v given the configurations observed in the simulations using a Bayesian formalism, as described previously.[34]

Acknowledgment. J.M. thanks Dr. Ad Bax for sending a preprint of the paper on SPARTA+ and making the computer code available. R.B. is supported by a Royal Society University Research Fellowship. This study utilized the high-performance computational capabilities of the Biowulf PC/Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov).

References and Notes

(1) Kauzmann, W. *Adv. Protein Chem.* **1959**, *14*, 1–63.
(2) Tanford, C. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*; Wiley: New York, 1973.
(3) Dill, K. A. *Biochemistry* **1990**, *29*, 7133–7155.
(4) Chandler, D. *Nature* **2005**, *437*, 640–647.
(5) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(6) Mackerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(7) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
(8) 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.
(9) Oosterbrink, C. A.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
(10) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.
(11) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926–935.
(12) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Intermolecular Forces*, 1st ed.; Reidel: Dordrecht, 1981.
(13) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
(14) Hawley, S. A. *Biochemistry* **1971**, *10*, 2426–2442.
(15) Panick, G.; Malessa, R.; Winter, R.; Rapp, G.; Frye, K. J.; Royer, C. A. *J. Mol. Biol.* **1998**, *275*, 389–402.
(16) Hummer, G.; Garde, S.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1552–1555.
(17) Patel, B. A.; Debenedetti, P. G.; Stillinger, F. H.; Rossky, P. J. *Biophys. J.* **2007**, *93*, 4116.
(18) Cheung, J. K.; Shah, P.; Truskett, T. M. *Biophys. J.* **2008**, *91*, 2427–2435.
(19) Sarupria, S.; Ghosh, T.; Garcia, A. E.; Garde, S. *Proteins* **2010**, *78*, 1641.
(20) Paschek, D.; Gnanakaran, S.; Garcia, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6765–6770.
(21) Dumont, C.; Emilsson, T.; Gruebele, M. *Nat. Methods* **2009**, *6*, 515–519.
(22) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665–9678.
(23) Paschek, D. *J. Chem. Phys.* **2004**, *120*, 6674–6690.
(24) Krouskop, P. E.; Madura, J. D.; Paschek, D.; Krukau, A. *J. Chem. Phys.* **2006**, *124*, 016102.
(25) Ashbaugh, H. S.; Collett, N. J.; Hatch, H. W.; Staton, J. A. *J. Chem. Phys.* **2010**, *132*, 124504.
(26) Abascal, J. L. F.; Vega, C. *J. Chem. Phys.* **2005**, *123*, 234505.
(27) Fernandez, R. G.; Abascal, J. L. F.; Vega, C. *J. Chem. Phys.* **2006**, *124*, 144506.
(28) Vega, C.; Abascal, J. L. F.; Nezbeda, I. *J. Chem. Phys.* **2006**, *125*, 034503.
(29) Aragones, J. L.; Noya, E. G.; Abascal, J. L. F.; Vega, C. *J. Chem. Phys.* **2007**, *127*, 154518.
(30) Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragones, J. L. *Faraday Discuss.* **2009**, *141*, 251–276.
(31) Pi, H. L.; Aragones, J. L.; Vega, C.; Noya, E. G.; Abascal, J. L. F.; Gonzalez, M. A.; McBride, C. *Mol. Phys.* **2009**, *107*, 365–374.
(32) Docherty, H.; Galindo, A.; Vega, C.; Sanz, E. *J. Chem. Phys.* **2006**, *125*, 074510.
(33) van der Spoel, D.; van Maaren, P. J.; Berendsen, H. J. C. *J. Chem. Phys.* **1998**, *108*, 10220–10230.
(34) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
(35) Best, R. B.; Mittal, J. *J. Phys. Chem. B* **2010**, *114*, 8790–8798.
(36) Mittal, J.; Best, R. B. *Biophys. J.* **2010**, *98*, L26–L28.
(37) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *11*, 526–535.
(38) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963–1974.
(39) Poland, D.; Scheraga, H. A. *Theory of helix-coil transitions in biopolymers*, 1st ed.; Academic Press: New York, 1970.
(40) Marqusee, S.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8898–8902.
(41) Marqusee, S.; Robbins, V. H.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5286–5290.
(42) Qian, H.; Schellman, J. A. *J. Phys. Chem.* **1992**, *96*, 3987–3994.
(43) Chakrabartty, A.; Baldwin, R. L. *Adv. Protein Chem.* **1995**, *46*, 141–176.
(44) Munoz, V.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 399–409.
(45) Munoz, V.; Serrano, L. *J. Mol. Biol.* **1995**, *245*, 275–296.
(46) Munoz, V.; Serrano, L. *J. Mol. Biol.* **1995**, *245*, 297–308.
(47) Hummer, G.; Garcia, A. E.; Garde, S. *Phys. Rev. Lett.* **2000**, *85*, 2637–2640.
(48) Garcia, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782–2787.

(49) Moreau, R. J.; Schubert, C. R.; Nasr, K. A.; Torok, M.; Miller, J. S.; Kennedy, R. J.; Kemp, D. S. *J. Am. Chem. Soc.* **2009**, *131*, 13107–13116.

(50) Shen, Y.; Bax, A. *J. Biomol. NMR* **2010**, *48*, 13–22.

(51) Nettels, D.; Müller-Spath, S.; Küster, F.; Hofmann, H.; Haenni, D.; Regger, 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.

(52) Robic, S.; Guzman-Casado, M.; Sanchez-Ruiz, J. M.; Marqusee, S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11345–11349.

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

(54) Shalongo, W.; Dugad, L.; Stellwagen, E. *J. Am. Chem. Soc.* **1994**, *116*, 8288–8293.

(55) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. *J. Am. Chem. Soc.* **2007**, *129*, 1179–1189.

(56) Case, D. A.; Scheurer, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2000**, *122*, 10390–10397.

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

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

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

(60) Rohl, C. A.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 8435–8442.

(61) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. *Biochemistry* **2000**, *104*, 9200–9210.

(62) Thompson, P. A.; Muñoz, V.; Jas, G. S.; Henry, E. R.; Eaton, W. A.; Hofrichter, J. *J. Phys. Chem. B* **2000**, *104*, 378–389.

(63) Rohl, C. A.; Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biochem.* **1992**, *31*, 1263–1269.

(64) Kennedy, R. J.; Walker, S. M.; Kemp, D. S. *J. Am. Chem. Soc.* **2005**, *127*, 16961–16968.

(65) Scholtz, J. M.; Hong, Q.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biopolymers* **1991**, *31*, 1463–1470.

(66) Paschek, D.; Puhse, M.; Perez-Goicochea, A.; Gnanakaran, S.; Garcia, A. E.; Winter, R.; Geiger, A. *ChemPhysChem* **2008**, *9*, 2742–2750.

(67) Avbelj, F.; Luo, P.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10786–10791.

(68) Garcia, A. E.; Hummer, G.; Soumpasis, D. M. *Proteins* **1997**, *27*, 471–480.

(69) Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, *110*, 17616–17626.

(70) Sadqi, M.; Lapidus, L. J.; Munoz, V. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12117.

(71) Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12491–12496.

(72) Chan, H. S.; Dill, K. A. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 447–490.

(73) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.

(74) Bax, A.; Kontaxis, G.; Tjandra, N. *Methods Enzymol.* **2001**, *339*, 127–174.

(75) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128–132.

(76) Hus, J.-C.; Peti, W.; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2003**, *125*, 5596–5597.

(77) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836–6837.

(78) Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1999**, *121*, 4690–4695.

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

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

(81) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472–2493.

JP108618D