JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Thermodynamics of Decaalanine Stretching in Water Obtained by Adaptive Steered Molecular Dynamics Simulations

Gungor Ozer,[†] Stephen Quirk,[‡] and Rigoberto Hernandez*,[†]

[†]Center for Computational and Molecular Science and Technology, School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, United States

[‡]Kimberly-Clark Corporation, Atlanta, Georgia 30076-2199, United States

ABSTRACT: The nonequilibrium stretching of decaalanine in vacuum using steered molecular dynamics and Jarzynski's relation led to the landmark determination of its potential of mean force by Park and Schulten (*Chem. Phys.* 2004). In so doing, the relative thermodynamics of the hydrogen-bond contacts and the entropy of the chain were quantified through the reversible work, the potential of mean force (PMF). A recently developed adaptive steered molecular dynamics algorithm (Ozer et al. *J. Chem. Theory Comput.* 2010) has now made it possible to determine the thermodynamics, PMF, of the stretching of decaalanine in a model solvent of TIP3P water molecules. The loss of internal hydrogen bonds and the formation of hydrogen bonds between the peptide and the solvent has also been tracked with the corresponding stabilization in the PMF. As in the vacuum, most of the thermodynamic penalty to unravel the chain in solvent occurs during the regime when the internal hydrogen bonds are broken. The formation of hydrogen bonds with the solvent provides a significant stabilization not seen in vacuum, reducing the total energy cost to unravel by nearly a factor of 2.

I. INTRODUCTION

An improved understanding of the thermodynamics and structure of polypeptides is critical to the characterization and prediction of peptide and protein structure and function. The simplest of these, homopolypeptides, are of interest because they provide useful benchmarks. Indeed, the thermodynamics of the stretching of polyalanines[1−5] in a vacuum has received a lot of attention using computational methods. A remaining critical question is how such thermodynamics (and motion) is affected by a water solvent.

There is, however, a question regarding the solubility of decaalanine in water. The problem is, of course, the hydrophobicity of alanine, but it can be overcome by mutating the ends by adding either charges or soluble amino acids. Several computational and experimental studies have suggested that decaalanine is soluble in water to varying degrees.[6−13] Interestingly, in a recent replica exchange molecular dynamics study, Pettitt et al.[14] have performed unfolding simulations of decaalanine in water, which presume the solubility of the helix, indicating that the initial structure is indeed α-helical as presumed in this study. Regardless, the present work does not prove or disprove the thermodynamics of the solubility of the unmodified decaalanine in water. We merely place decaalanine in a compact (α-helical) form in water (mechanically) where it has no choice but to be as we hold it there with invisible tweezers at both ends. What we then obtain is the relative energy to pull the helix apart and thereby obtain the relative stability of the solvated closed and open forms under a constrained end-to-end distance. We find that under these constraints, the α-helix is the more stable form in agreement with the recent experimental findings in ref 15.

A stable protein can be opened along a chosen pathway by a corresponding force. If applied sufficiently slowly (adiabatically), the minimum required work (exerted energy) corresponds to the reversible work. Amazingly, Jarzynski[16] and others[17−19] have shown that the corresponding reversible potential of mean force can be obtained through an appropriately weighted average of the applied work from a series of nonequilibrium trajectories. Schulten and Park[1,2] further showed that it could be computed numerically as implemented in steered molecular dynamics (SMD). Unfortunately, the convergence of the nonequilibrium trajectories is slow and gets worse as the nonequilibrium disturbance is lengthened. We have recently introduced an adaptive steered molecular dynamics (ASMD)[5,20] approach so as to mitigate the wandering of the work trajectories far from the average and thereby lead to faster convergence. ASMD relies on a staging algorithm sharing similarities with other SMD studies[21] and a more recent approach implementing steering within transition path sampling.[22]

This article obtains the potential of mean force for pulling decaalanine from the solvated α-helix to a nearly fully stretched coil using ASMD. In ref 20, the accuracy of the ASMD methodology, briefly reviewed in section II, could only be inferred from indirect experimental evidence. Recently, in ref 5, we have demonstrated that this approach can reproduce the free energy profile of decaalanine stretching obtained earlier in a vacuum[1,2] using significantly less CPU time than the standard SMD implementation. Specifically, we found that ASMD reduced the number of trajectories required for convergence approximately 10-fold. In section III, we obtain the PMF for the stretching of decaalanine in solvent and add to the debate[7−13,23] over the relative thermodynamics of the α-helix and coil. In so doing, we also track the number of hydrogen bonds within the peptide and between the peptide and solvent. The results provide

Received: August 12, 2012
Published: October 11, 2012

new insight into the unraveling of a peptide and the role of hydrogen bonding therein.

II. MODELS AND METHODS

Steered molecular dynamics[1,2] (SMD) is a nonequilibrium MD method that mimics force-probing experiments such as atomic force microscopy. Within the simulation, external forces are applied on the system of interest so as to steer it along a predefined "reaction" path. By integrating the forces applied on the system by auxiliary harmonically coupled particles, the work done on the system is calculated as the steering proceeds. This leads to a distribution of work values over the ensemble of configurations of the system associated with the initial constraints. The nonequilibrium work distribution can then be averaged using Jarzynski's equality, detailed below, to calculate the constrained free energy profile, that is, the potential of mean force (PMF), along the reaction path. Although SMD, as compared to unconstrained MD, effectively reduces the computational cost in modeling large conformational changes of bimolecular systems, the amount of force applied (ranging typically from 500 pN to several thousands pN) is far larger than that applied in AFM experiments (up to a couple hundred pN) for which SMD aims to reproduce the results. As the applied force (thus work) is set at such high values, the distribution of work gets distorted from Gaussian behavior and ultimately causes the calculated PMF to be dominated by the lowest energy trajectories. These shortcomings can be surmounted, but at significant computation cost, either by increasing the size of the ensemble, requiring on the order of millions of realizations, or by lowering the stretching velocity. Instead, we have recently introduced an adaptive SMD methodology[20] that allows one to obtain converged PMF values over shorter stages and then to link them together to obtain the overall PMF. The stages are taken to be short enough that the spreading of the trajectories is not as severe and hence allows for convergence with fewer trajectories. It also addresses possible concerns with the averaging of the exponentially weighted work functions in the Jarzynski relation as the spread is somewhat narrower. The excellent agreement between the direct averaging of the exponentials and the second-order cumulants expansion (seen in, for example, Figures 8 and 11 of ref 20) indicates both the avoidance of the possible numerical error in such averages as well as convergence with the underlying distribution of trajectories entering the average.

Staging the reaction path into many segments may resemble, at first glance, the weighted histogram analysis of a series of umbrella sampling[24] (WHAM-US) simulations. Regardless, it is a natural alternative to the use of SMD or ASMD for obtaining the PMF along a reduced dimensional coordinate, and it serves as an important benchmark for this work. In WHAM-US, it is preferable to use smaller bin (window) size, that is, use more bins, to minimize the statistical error and hence to obtain more accurate results such as the potential of mean force albeit with the additional cost of more sampling. Similarly, in ASMD, as the number of stages increased, the stretching velocity must be slowed to ensure that there is sufficient sampling within a given stage (see, for example, Figure 5 of ref 5 in which the convergence detiorates with increasing number of stages for a fixed stretching velocity). Although we have not explicitly compared the efficacy of ASMD to that of umbrella sampling, we had previously benchmarked ASMD as compared to the standard SMD in a series of decaalanine stretching simulations in vacuum and found that adaptive integration of SMD yields converged PMF, requiring fewer trajectories than standard SMD.[5] Meanwhile, Schulten's group compared the SMD method to umbrella sampling[1] and showed that SMD and US produce comparable results when the same amount of calculations are utilized in both. As such, we expect that ASMD offers a 10-fold computational enhancement in converging the PMF in the vacuum, and possibly more in solvent.

A. Adaptive Steered Molecular Dynamics. The basis for the determination of the potential of mean force along a chosen thermodynamic path using nonequilibrium dynamics is Jarzynski's equality:[16,17]

$$G(\xi_t) = G(\xi_0) - \frac{1}{\beta}\ln\langle e^{-\beta W_{\xi_t \leftarrow \xi_0}}\rangle_0 \tag{1}$$

in which the work along the ensemble of trajectories is averaged through the exponential weight. This relation lends itself both to experimental[25] and to computational[17−19] studies. The primary difficulties in employing this relation are (i) the determination of a relevant thermodynamic path $\lambda(t)$ for some order parameter or reaction path coordinate $\xi(\vec{r})$ and (ii) the implementation of a constraint to impose that the trajectories follow the path. Park and Schulten[1,2] suggested the use of an auxiliary particle to smoothly steer the system along the path $\lambda(t)$ at a constant velocity so as to couple to the order parameter through a harmonic potential:

$$h_\lambda(r) = \frac{k}{2}(\xi(r) - \lambda)^2 \tag{2}$$

The accumulated work applied by this constraint is then necessarily the work $W_{\lambda(t)\leftarrow\lambda(0)}$. To the extent that $\xi(t)$ for a given trajectory agrees with $\lambda(t)$, then the latter can be approximately[18] equated to $W_{\xi_t\leftarrow\xi_0}$ or exactly[1] in the limit of a stiff spring. Corrections to this approximation have been implemented,[19,26] but are generally ignored if the pulling rates are sufficiently slow that the deviations are small, and steered molecular dynamics (SMD) has become a useful tool for determining free energy differences between closely spaced structures.

The difficulty in applying SMD more generally is that the work values spread dramatically as the system is steered along longer paths. One class of solutions to solve this problem involves contraction of the trajectories by way of retaining only those that are deemed to continue to satisfy a funnel hypothesis,[21] and has even been applied in a staged procedure.[21] The adaptive SMD (ASMD) algorithm takes this narrowing to a relative extreme using only one structure at the beginning of each stage that is obtained from the nonequilibrium sample at the end of the previous stage. Specifically, the overall reaction coordinate ξ is partitioned into N segments bounded by the $N + 1$ constrained points, $\xi_0,\xi_1,\ldots,\xi_N$. Within each segment, the work is obtained using eq 1 assuming that an appropriate initial configuration is known. The iterated average free energy at any given point ξ_t along the path takes the form:

$$\Delta G_{\xi_t\leftarrow\xi_0} = -\frac{1}{\beta}\sum_{i=1}^{N'}\ln\langle e^{-\beta W_{\xi_i'\leftarrow\xi_{i-1}'}}\rangle_{i-1} \tag{3}$$

for N' chosen such that $\xi_t \in (\xi_{N'-1}, \xi_{N'}]$, $\xi_i' \equiv \xi_i$ for $i < N'$, and $\xi_{N'}' \equiv \xi_t$. The unconstrained system and Langevin bath variables (denoted as Γ and Θ, respectively, in ref 20) associated with the initial point ξ_0 are initialized by equilibration as in a standard SMD simulation.

In a typical SMD simulation, one constrains the initial path variable ξ to $\lambda(0)$ using an infinitely stiff spring and allows the unconstrained variables Γ to relax providing an ensemble of initial configurations. Alternatively, a single such configuration can be used to initiate all of the trajectories as long as the Langevin bath variables are sampled statistically and distinctly for each trajectory given a sufficiently slow external driving of the system across the constrained variable. This latter relaxation, leading to a resampling of the ensemble, also suffices for each stage in the ASMD as discussed in ref 5. Thus, the critical component in ASMD is the determination of the initial configuration for each new stage i in the ASMD given the configurations reached by the trajectories in the previous iteration $i - 1$. We found[5,20] that the JA (Jarzynski average) criterion is both efficient and leads to the correct overall PMF. According to the JA criterion, one chooses the structure $\{\xi_{i-1}(t),\Gamma_{i-1}(t)\}$ at the end of the previous iteration whose work was closest to the Jarzynski average energy across all of the trajectories. The ensemble of trajectories in the ith iteration are initiated at $\{\xi_{i-1}(t),\Gamma_{i-1}(t),\theta_\alpha\}$ for a statistical sampling of bath variables θ_α consistent with the thermal distribution. This choice is intuitive because it amounts to selecting a structure that is fully relaxed, taking advantage of the results from the trajectories that have already been calculated, without requiring an additional relaxation period before initiating the next batch of trajectories. The validity of the adaptive SMD methodology using the JA selection criteria has been shown in our earlier work.[5,20]

While the potential of mean force can be obtained by averaging the trajectories using eq 1, there is a question as to whether a similar expression can be used to obtain observables. For some observable $\hat{X}$, we define the average:

$$\bar{X}(\lambda(t);\{\xi_i'\}) = \frac{\langle \hat{X}(\xi_t)\, e^{-\beta W_{\xi_t \leftarrow \xi_{k_t}'}}\rangle_{k_t}}{\langle e^{-\beta W_{\xi_t \leftarrow \xi_{k_t}'}}\rangle_{k_t}} \quad (4)$$

$$= e^{\beta \Delta G_{\xi_t \leftarrow \xi_{k_t}'}} \langle \hat{X}(\xi_t)\, e^{-\beta W_{\xi_t \leftarrow \xi_{k_t}'}}\rangle_{k_t} \quad (5)$$

where k_t is chosen such that $\lambda(t) \in (\xi_{k_t}', \xi_{k_t+1}']$. This numerical average can be obtained readily during each ASMD stage. The question is whether it is equivalent to the expectation value $\bar{X}(\lambda(t))$. In the limit of a single-stage SMD simulation initiated with a distribution of unconstrained variables Γ and θ, much previous work has suggested that they are equivalent within reasonable limits.[18,26] If Γ is not initially distributed, as implemented at the start of each ASMD stage, then clearly the value of $\bar{X}(\lambda(t);\{\xi_i'\})$ will be highly dependent on the one structure used at a given initial point ξ_i'. However, once the steering has pulled λ away from this end point with sufficient time for Γ to relax (or at least respond to the nonequilibrium driving), there will be a statistical sampling, and the expectation value should result in the same structure as that found in the one-stage case. If instead a single such structure does not suffice to fully sample the steered distributions, then additional initial structures would need to be implemented in ASMD. Such a multiple-branching ASMD was discussed formally in ref 5 and will be further developed in future work.

B. Simulation Parameters. The NAMD[27] code is used to carry out all-atom adaptive SMD simulations. Atomic interaction parameters are taken from the CHARMM force field.[28] Decaalanine is treated as a 104-atom model, and all hydrogens are defined explicitly. Water molecules have been treated using the TIP3P[29] potential as described within NAMD. The use of more sophisticated water models may affect the results, but they would also require more computational resources. The initial coordinates of decaalanine peptide are taken from a compact helical structure modeled in a vacuum, which has been studied elsewhere.[1,2,5] Park and Schulten[1] selected this initial configuration so as to be more compact than the minimum energy configuration in vacuum. The end-to-end distance specified by that between the N-terminal nitrogen and the C-terminal capping nitrogen for this structure is 13.0 Å. This same structure is selected in the current case of a solvated stretching of decaalanine to provide direct comparison with our earlier ASMD work[5] in the vacuum. The fixed peptide, with all 104 atoms held at the positions of the initial structure, is solvated with 5138 water molecules in a cuboid box with dimensions 50 Å × 50 Å × 64 Å. This longest side of this square cuboid is chosen so as to ensure that there is sufficient water between the peptide ends at the final elongation. All MD simulations are propagated with time steps set to 2 fs. Temperature is controlled using Langevin dynamics as implemented in the built-in NAMD integrator. The water molecules within the simulation box are thermally equilibrated by first (pseudoadiabatic) heating the bath from 0 to 300 K in 5 ps at constant volume, and then correcting the solvent density by propagating the solvent under constant pressure and temperature (NPT) for 5 ps. An additional equilibration at constant volume and temperature is then performed for 10 ps while the peptide remains fixed.

In the production stage of the ASMD simulations, decaalanine is stretched so that the end-to-end distance between the nitrogen atom of the N-terminus (NN) and the nitrogen atom of the cap at the C-terminus (NC) is gradually increased from 13 to 33 Å (Figure 1). Specifically, the nitrogen at NN is kept fixed at the origin, (0,0,0). The nitrogen at C-terminus is pulled along the longest axis of the square cuboid box (defined to be the z-axis)

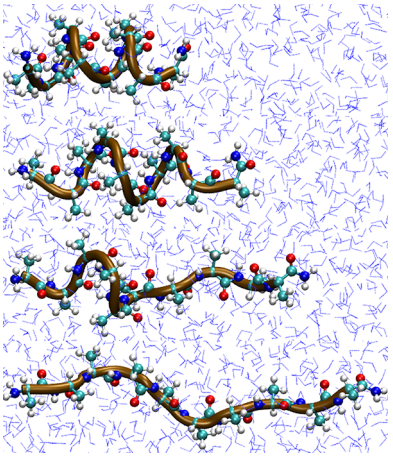

Figure 1. Illustration of the unraveling of decaalanine helix as N-terminal nitrogen (NN) is fixed and C-terminal capping nitrogen (NC) is pulled. Four different snapshots of the peptide structure are shown in a single cartoon embedded in water molecules (drawn using stick models) so as to remind the reader that all of the simulations were performed in all-atom water solvents. The top image represents the initial fixed structure[1] of the decaaline α-helix for which the NN−NC distance is 13 Å as used earlier in a vacuum by Park and Schulten. The bottom image is a representative coil structure randomly chosen from the non-equilibrium ensemble with an NN−NC distance of 33 Å. The second and third structures from the top are again randomly selected snapshots of stretched decaalanine when the NN−NC distance is 20 and 32 Å, respectively.

dx.doi.org/10.1021/ct300709u | J. Chem. Theory Comput. 2012, 8, 4837−4844

from (0,0,13) to (0,0,33) while harmonically attached to a pseudo particle using the biased Hamiltonian in eq 2. ASMD simulations of the decaalanine in solvent were also carried out at 300 K in the cuboid of dimensions 50 Å × 50 Å × 64 Å. The harmonic force constant, k, is set to be 7.2 kcal/mol as used earlier in SMD.[1,2] All other simulation parameters are the same as in the equilibration protocol described above.

The forced stretching of decaalanine is implemented as an ASMD experiment in which the reaction coordinate, that is, the end-to-end distance between NN and CN, is divided into 10 stages. Multiple trajectories at each stage have been generated over the equilibrated initial configuration by varying the random forces in the Langevin bath. To obtain good statistics and to observe the dependence of the experiment on the sample size, four sets of simulations, using 50, 100, 200, and 400 trajectories per step, were implemented. The resulting trajectories are then analyzed both numerically and visually. PMFs along the stretching coordinate are calculated with in-house scripts, while empirical and visual analyses are done using the VMD package.[30] Figure 2 illustrates in detail the generated work distribution and

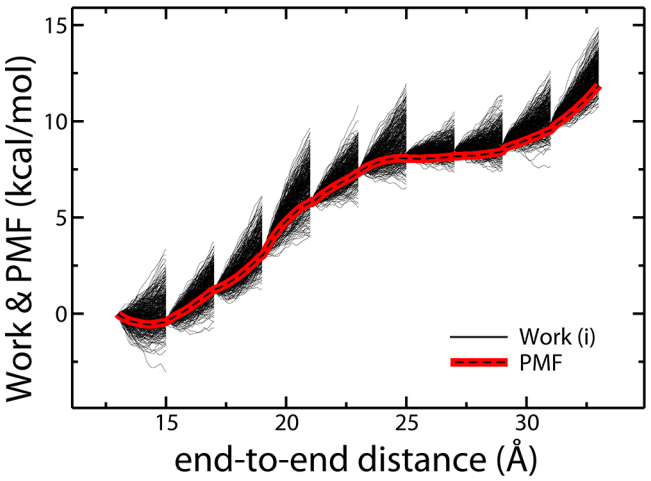

Figure 2. The work distribution at each stage and the overall PMF obtained for the forced stretching of decaalanine in solvent are displayed for 10 Å/ns pulling simulations. The figure illustrates the use of ASMD divided into 10 stages with 400 trajectories per step. As specified in the legend, the solid black lines represent the work for each trajectory, whereas the thick-red-highlighted black-dashed curve represents the overall PMF.

the resulting PMF obtained from 10 Å/ns pulling simulations with 400 trajectories sampled within each stage. The PMFs for other pulling velocities and number of trajectories generated per step are reported in sec. III.A. The effect of the solvent is also quantified through a count of the hydrogen bonds within the α-helix and those between the α-helix and water molecules around it. In the latter case, hydrogen bonds are formed when donor and acceptor atoms are found to be within 4 Å and the angle spanned by hydrogen-donor and hydrogen-acceptor rays is greater than 140°.

III. RESULTS AND DISCUSSION

The forced unraveling of decaalanine in a vacuum was investigated earlier in ref 5. The ASMD PMFs at two stretching velocities, 100 and 10 Å/ns, agreed well with those obtained by Schulten and Park.[1,2] In both cases, the converged reversible PMF was obtained at the slow stretching velocity, 10 Å/ns. The PMF for decaalanine stretched in a vacuum is reproduced as a black curve in the bottom panel of Figure 3. The end-to-end

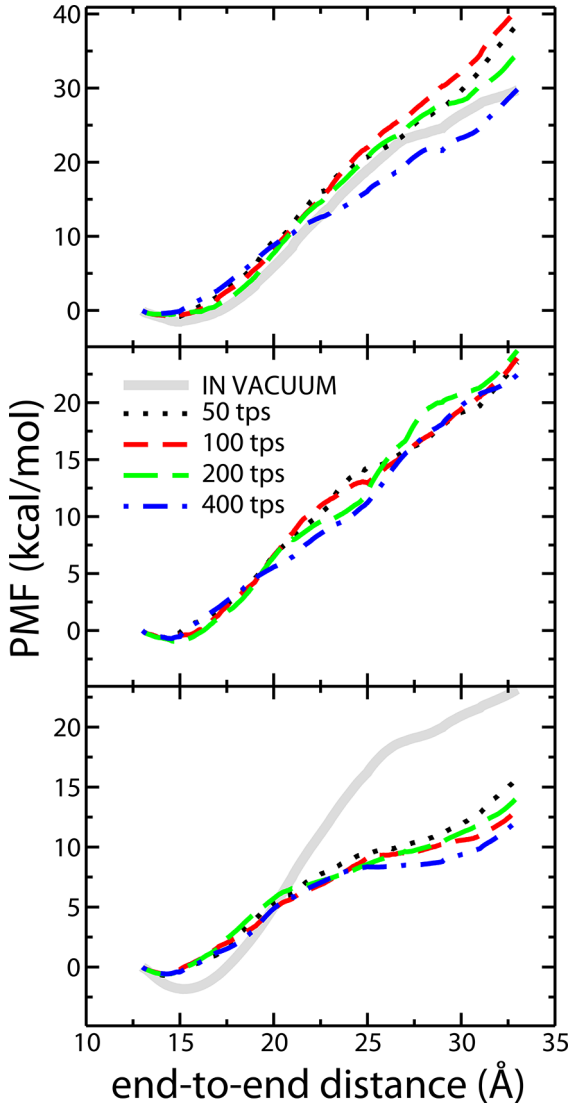

Figure 3. The PMFs obtained for the forced stretching of decaalanine in solvent using the adaptive SMD method are displayed for three stretching velocities: 100 Å/ns (top panel), 33 Å/ns (middle panel), and 10 Å/ns (bottom panel). The series of PMFs obtained using the adaptive SMD method with 50, 100, 200, and 400 trajectories per step (tps) are specified in the legend. The solid gray curves in the bottom and top panels represent the PMF obtained from the corresponding adaptive SMD simulations in a vacuum as shown in ref 5.

distance, that is, the distance from the nitrogen at the N-terminus to the carbon at the C-terminus, is stretched for 20 Å from an initial value of 13 Å to a final value of 33 Å. The initial structure is a compact helix with coordinates chosen to match those used earlier by Schulten and Park.[1,2] It is not, however, the lowest energy structure as is evident in the PMF in Figure 3. The constrained free energy along the unraveling coordinate for decaalanine in a vacuum goes down in energy from the initial (arbitrary) compact form to a minimum at ca. 15.2 Å in extension, whereas in explicit water the minimum energy configuration of the decaalanine is reached at ca. 14.3 Å in extension.

dx.doi.org/10.1021/ct300709u | J. Chem. Theory Comput. 2012, 8, 4837−4844

A. Thermodynamics. The steered unraveling of decaalanine in solvent has been investigated at three stretching velocities (10, 33, and 100 Å/ns). For each stretching velocity, 50, 100, 200, and 400 trajectories have been simulated at each step. The addition of 5138 water molecules into the system box dramatically increases the required CPU time per trajectory from 200 min to 65 h (when stretching at the slowest velocity, 10 Å/ns) on the HPC resources accessed in this work. (Specifically, the primary resource was the Steele Cluster Pool on the XSEDE running an array of 2.33 GHz dual quad-core Intel 5410 Xeon processors.) Assuming constant access to 96 cores, it would take over 282 days to acquire 10 000 trajectories in solvent. For the adaptive SMD calculations, it took less than 40 days to integrate all four sets of simulations (50, 100, 200, and 400 trajectories per step) at the slowest stretching velocity (i.e., 10 Å/ns) using 96 cores. The calculated PMFs for the solvated decaalanine stretch are shown in Figure 3. The best calculated value, shown as the blue dot-dashed curve in the bottom panel, is somewhat lower in energy than the vacuum result, shown as a solid black curve. Such a lowering is to be expected because of the stabilization provided by the water molecules as the peptide is unraveled. Meanwhile, the thermodynamically accessible states remain structured in the sense that there is an initial fast rise, possibly two, and a subsequent slower rise after 25 Å. The near agreement between the PMFs found in the initial stages of the stretching should not be surprising because they are near the original configuration, and hence the underlying Gaussian behavior in the work distribution is followed well. At longer distances, however, the calculated PMFs do not converge well with increasing number of trajectories per step at the fast stretching velocities, 100 Å/ns (top panel) and 33 Å/ns . At the lowest stretching velocity, 10 Å/ns in the bottom panel, the fluctuations in the work values are never larger than 1−2 kcal/mol, and this provides an error estimate in the calculated PMF within this narrow range. Also, at the lowest stretching velocity, the structural elements of the PMFs appear to be less dependent on the sampling size within a few kcal/mol and exhibit a monotonic converging trend in the free energies.

The initial choice for the decaalanine helix structure was obtained seemingly arbitrarily from a compact model with the C terminus capped with an amide group ($-NH_2$). The fact that this compact conformation is not the actual minimum energy structure, even along the direction of the unwinding of the chain, is visible in the offset in position (at ca. 14.3 and 15.2 Å) for the global minima, both in the vacuum and in the solvated systems at $\Delta G = -1.83$ kcal/mol and $\Delta G = -0.58$ kcal/mol, respectively. Meanwhile, both cases exhibit a near flattening in the free energy presumably in the region of the open form of the protein before experiencing sharp increases as the protein is further straightened at a high entropy cost. This occurs at $\Delta G =$ 18.34 kcal/mol and $\Delta G =$ 8.48 kcal/mol in the vacuum and solvated cases, respectively. Using these values along the steered coordinate, we can estimate the free energy difference in the helix to coil transition of decaalanine as $\Delta G =$ 20.17 kcal/mol and $\Delta G =$ 9.06 kcal/mol for the vacuum and solvated systems, respectively. That is to say, there is a $\Delta\Delta G$ (=11.11 kcal/mol) relative stabilization of the coil due to solvent interaction. The additional stabilization upon hydration can be compared to the free energy of solvation of alanine and alanine-based peptides in the literature. Mezei et al.[31,32] estimated the Helmholtz free energy of solvation of a single residue of polyalanine α-helix as $\Delta\Delta A_{12\leftarrow 11} = -2.0$ kcal/mol/residue. This per residue estimation of Helmholtz free energy along with the observation that $T\Delta\Delta S$ is negligible can be used to approximate the free energy of decaalanine peptide upon hydration. The α-helix is stabilized by $i \rightarrow i + 4$ contacts through hydrogen bonding. Therefore, in decaalanine, the inner six residues will contribute to the overall free energy difference of solvation for a total of $\Delta\Delta G \approx -12$ kcal/mol, which compares well with our observation of $\Delta\Delta G \approx -11.1$ kcal/mol. In a recent replica exchange molecular dynamics study, Pettitt et al.[14] calculated the solvation free energies of the helix and extended coil conformations of decaalanine as −220 and −290 kcal/mol, respectively. This suggests a dramatically different $\Delta G \approx -70$ kcal/mol in disagreement with our observed ΔG, which is likely due to the presence of salt effects not included in our present work.

Previous work by Baldwin and co-workers[33] on the relative stability of the helix and the coil reported that the helix formation of alanine-based peptides is enthalpy driven. They have measured the net per residue enthalpy change of the coil to helix transformation in solvent as $\Delta H = -3.8$ kcal/mol. The loss of conformational entropy for α-helix formation is estimated to be −2.2 cal/mol/K per alanine residue.[34] One can calculate the free energy of the helix formation of the decaalanine by $\Delta G = (n - 4)\Delta H - (n - 2)T\Delta S$ according to Baldwin and Scholtz[35] as −17.5 kcal/mol at 300 K as compared to our result of −9.06 kcal/mol.

B. Hydrogen Bonding. The potentials of mean force shown in the bottom panel of Figure 3 reveal important structural properties of the helix−coil transition of decaalanine in water. The initial structure is evidently not the minimum energy structure as the energy minimum appears at around 14.9 Å in nearly all of the simulations roughly independent of stretching velocity. A small shift in the slope of the rise in the PMF appears near 20 Å, which is not seen in the vacuum limit. A more pronounced flattening in the PMF is seen near 25 Å both in vacuum and in solvent. It is notable that the forced stretch does not exhibit these features for the high velocity cases. We hypothesize that this is due to insufficient relaxation of the water molecules around the nonequilibrium decaalanine structure in the fast forced stretching cases. This hypothesis is corroborated by the hydrogen-bond counts displayed in Figure 4. The average number of intrapeptide hydrogen bonds in a vacuum and solvent is shown in the top and middle panels, respectively. The average number of interpeptide hydrogen bonds to the solvent is shown in the bottom panel. Initially, there are 15 hydrogen bonds formed between neighboring water molecules and decaalanine. This model is consistent with previous work in the sense that even when the adopted conformation is an α-helix (i.e., peptide is stabilized with intrapeptide hydrogen bonds), water interacts strongly with solvent exposed helices.[36] At the fully stretched configuration, there is an average of 27 hydrogen bonds between neighboring water molecules and the decaalanine solute. This number is consistent with the average hydrogen-bond count between the "extended" decaalanine and TIP3P water molecules, 27, which Pettitt and co-workers[14] have calculated from a 20 ns MD simulation at 298 K. This shows that stretching velocity is small enough that the bath is allowed to equilibrate as the molecules is being stretched to the final denatured configuration. When decaalanine unravels into a solvated coil (with no intrapeptide hydrogen bonds), the total number of hydrogen bonds between it and water is 27. This is the same as the number of initial hydrogen bonds: 15 hydrogen bonds between decaalanine and water molecules and 12 degenerate hydrogen bonds within decaalanine. The degeneracy arises because each hydrogen-bonded pair within decaalanine breaks into two

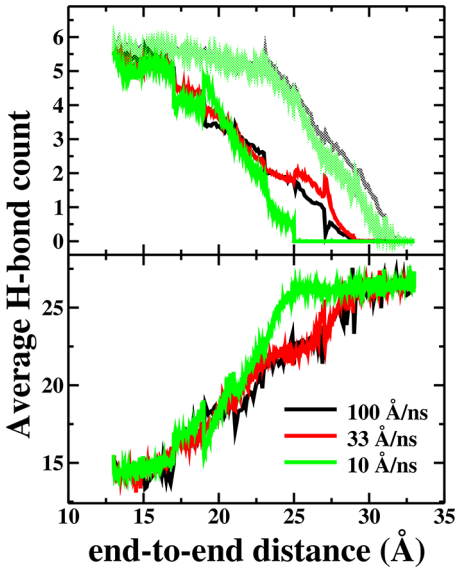

Figure 4. The average number of intrapeptide hydrogen bonds and hydrogen bonds formed between the decaalanine and the surrounding water molecules as a function of decaalanine end-to-end distance is shown in the top and bottom panels, respectively. In both panels, the results are obtained for three stretching velocities: 100 Å/ns (black), 33 Å/ns (red), and 10 Å/ns (green). Semitransparent black and green curves in the top panel represent the average number of intrapeptide hydrogen bonds in a vacuum as shown in ref 5.

hydrogen-bonding sites that are accessible to the solvent upon the stretching of the peptide. Hence, there are, on average, only 6 nondegenerate intra peptide hydrogen bonds in the initial helical structures in agreement with the initial averaged values shown in the top and middle panels of Figure 4. As the peptide is stretched, there is a marked difference in the loss of hydrogen bonds between the vacuum and water solvated cases. In vacuum, decaalanine maintains most of the intrapeptide hydrogen bonds even when it is stretched by as much as 12 Å, whereas in solvent, it nearly loses them all by this point. As indicated in the bottom panel, this is is due to the establishment of hydrogen bonds in the solvent, which are obviously not available in the vacuum. In a molecular dynamics study of ALA_{13}, Doruker and Bahar[37] observed the distribution of backbone dihedral angles along the MD trajectories and similarly concluded that a lack of hydration favors intrapeptide hydrogen bonding.

In the solvent, the interpeptide hydrogen bonds appear to exhibit four regimes. The first of these, during the first few angstroms of the forced stretch, exhibits a relatively constant number of hydrogen bonds. This is consistent with the thermodynamics in Figure 3, which display a relaxation of the structure from its initial to a minimum energy value during which the structure undergoes intramolecular reorganization without breaking the hydrogen bonds. The second and third regimes correspond to the loss of the first three and last three intrapeptide hydrogen bonds, respectively. The slope of the loss of hydrogen bonds appears to be slower in the second regime than in the third regime and corresponds to the small change in the slopes in the PMF. Finally, the fourth regime exhibits little change in hydrogen bonding and corresponds to the relatively flat thermodynamics in the PMF. These observations are thus consistent with the hypothesis that the thermodynamics of the stretching of neuropeptide Y are primarily correlated with the making and breaking of hydrogen bonds within the peptide and to the solvent.[20]

A more detailed view of the loss of structure (hydrogen bonds) as a function of the stretching of the ends can also be obtained from the ASMD simulations. The initial dihedral angle distribution, the Ramachandran plot, exhibits highest population in the α-helical regions (Figure 6). Echeverria and Amzel[4] hypothesized, however, that as decaalanine is stretched, the dihedral angles can sample other regions of the configuration space mainly due to the broken hydrogen bonds. As decaalanine is stretched, we can track which pairs tended to be involved in hydrogen bonds quantitatively. Specifically, the average number of hydrogen bonds between residues i and $i + n$ for all i and various residue separations n was calculated along the ASMD trajectory. The 3_{10}-helix, α-helix, and π-helix should exhibit large hydrogen bonding at n equal to 3, 4, and 5, respectively. Figure 5

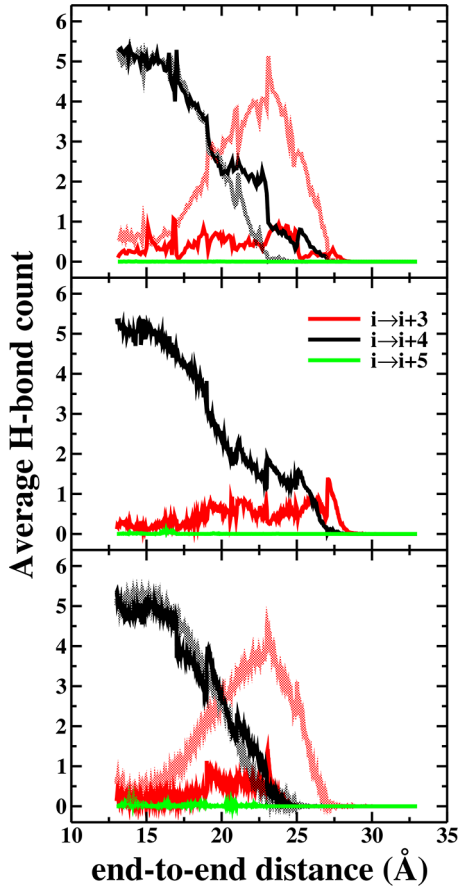

Figure 5. The average number of α-, 3_{10}-, and π-helical contacts as a function of decaalanine end-to-end distance is shown. Top, middle, and bottom panels represent data from 100, 33, and 10 Å/ns simulations, respectively. Black curves represents $i \rightarrow i + 4$ (α-helix), red represents $i \rightarrow i + 3$ (3_{10}-helix), and green represents $i \rightarrow i + 5$ (π-helix) contacts, respectively. Semitransparent black and red curves in the top and bottom panels are reproduced data shown in ref 5, and illustrated here for comparison.

displays the average hydrogen bonding in vacuum and in solvent as a function of the stretching distance. In the vacuum, the α-helix contacts are replaced by 3_{10}-helix contacts before completely unraveling into a coil. The vacuum results do not exhibit any π-helix ($i \rightarrow i + 5$ hydrogen bonding) formation. In solvent, similarly, the α-helical structure disappears roughly at the same end-to-end distance, that is, 24 Å. These lost contacts are,

dx.doi.org/10.1021/ct300709u | J. Chem. Theory Comput. 2012, 8, 4837−4844

however, mostly not replaced by the $i \rightarrow i + 3$ contacts of a 3_{10}-helix. Instead, these broken hydrogen bonds are immediately replaced by hydrogen bonds to the neighboring water molecules as seen in Figure 4. The coexistence of small amounts of 3_{10}-helix with the dominant α-helix conformation is also consistent with Daggett and co-workers' earlier report.[38] The relative populations of α-helix ($\phi = -57°, \psi = -47°$), 3_{10}-helix ($\phi = -49°, \psi = -27°$), π-helix ($\phi = -57°, \psi = -70°$), and extended-coil ($\phi \approx -90°, \psi \approx 120°$) structures can be seen through snapshots of the Ramachandran plots shown in Figure 6. The growth of 3_{10}-helix

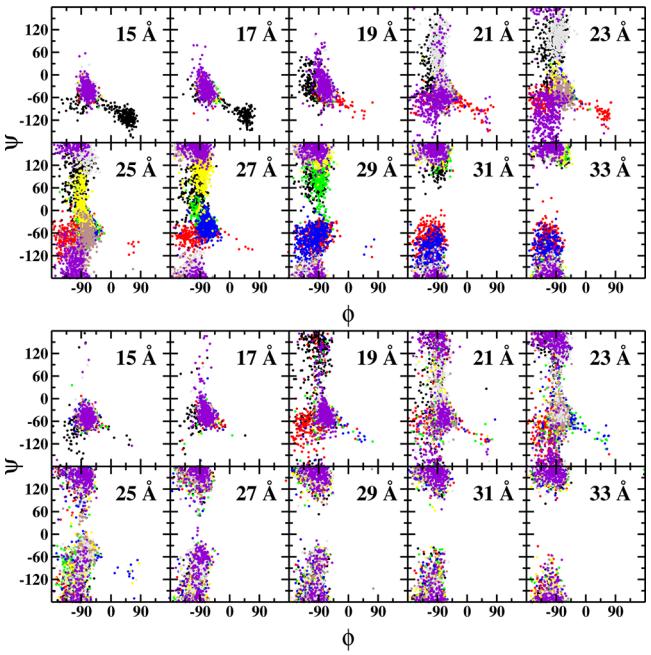

Figure 6. Ramachandran plots of the middle eight residues (excluding the termini residues because they do not have a pair of phi and psi angles) along the direction of the stretching in TIP3P solvent are displayed for the 100 Å/ns stretching simulations (top) and for the 10 Å/ns stretching simulations (bottom). Each small diagram above has a total of 3200 data points coming from the ϕ–ψ angle pairs of the eight middle residues at the end-to-end distance labeled inside them. Coloring scheme is as follows: ALA2, black; ALA3, red; ALA4, green; ALA5, blue; ALA6, yellow; ALA7, brown; ALA8, gray; ALA9, purple.

structure in the first 14 Å of stretching in the vacuum is visible in the broadening of the distribution toward larger ψ. Figure 6 displays some population toward the left of the initial α-helical region that corresponds to the small amounts of π-helix formation observed in the initial hydrogen-bonding population of Figure 5 especially when the decaalanine is stretched at the slow rate of 10 Å/ns. The small amount of 3_{10}-helix formation is observed when end-to-end distance is between 14 and 22 Å in Figure 5 and visible in the increased population in the angles just above the α helix region in Figure 6. Figure 6 also shows how solvent stabilization on the peptide is dependent on the pulling rate. The torsional angles of decaalanine peptide are distributed over broader regions as the pulling velocity is larger. This is true both in the 100 Å/ns simulations (Figure 6) and in the 33 Å/ns simulations (although not shown). In 10 Å/ns pulling simulations, however, the distribution is confined in narrower ranges, yielding a less disordered Ramachandran plot. This may confirm that, when stretching is slow enough, the decaalanine peptide slowly equilibrates with neighboring water molecules so that residual torsional angles are adjusted toward that of the minimum energy configuration. On the contrary, when the stretching rate is fast, so that the decaalanine residues do not relax enough, the tosional modes simply cannot adjust toward lower energy configurations. The difference is not quite visible in vacuum simulations (see Figure 9 of Ozer et al.[5]), which further suggests that water molecules stabilize the peptide during the forced helix–coil transformation.

IV. CONCLUSIONS

The recently formulated ASMD algorithm[20] had already been used to determine the potential of mean force for the stretching of decaalanine in vacuum.[5] They were found to be in agreement with the earlier SMD simulations in a vacuum of Park and Schulten[1,2] and serve as a conformation of the approach. The present work employs the ASMD algorithm to stretch a solvated decaalanine peptide from a helix to a coil. In both vacuum and solvent, the adaptive SMD approach converges with relatively little CPU cost as it requires the simulation of less than 1000 trajectories rather than tens of thousands or more. The latter calculation would render the solvent PMF essentially computationally intractable with current computer resources.

Comparison between the vacuum[5] and solvated stretching of decaalanine presented herein provides substantial insight on the role of the solvent. The free energy cost to unraveling the chain drops from ca. 20 to ca. 9 kcal/mol. The first half of the forced stretch in a vacuum leads primarily to deformation of the intramolecular contacts. In solvent, this stretch leads instead to the breaking of intrapeptide hydrogen bonds with no new intrapeptide hydrogen-bond formation. The latter are replaced by hydrogen bonds to the solvent, which accounts for the energy stabilization described above. Consistent with these findings, Echeverria and Amzel[4] previously claimed that in aqueous solvent peptide hydrogen bonds are bridged by the water molecules as an intermediate step toward the deformation of the initial intrapeptide hydrogen bonds. This suggests that the unraveling of decaalanine, whether stretched by an AFM or optical tweezers, should be sensitive to the hydrogen-bonding character of a solvent to a measurable degree.

One possible criticism to the use of ASMD for studying the thermodynamics of decaalanine stretching in solvent concerns the question of whether the α-helix structure of decaalanine is stable therein. Indeed, this question has been debated over many experimental and computational reports with somewhat inconclusive results. Blondelle et al.[8] and Williams et al.[9] observed very little helical content experimentally in studying solubilized and Ac– and $-NH_2$ capped LYS-$(ALA)_{14}$-LYS employing a "host–guest" technique using circular dichroism. Employing the same technique, Marqusee et al.[13] found that 16 or 17 residue long peptides containing only alanine and a small number of lysine or glutamine strikingly exhibit helix formation in water. Baldwin and co-workers[6] later showed that alanine is the only amino acid that forms stable helix in water without the aid of any helix-stabilizing interactions. Similarly, Spek et al.[7] reported that ALA side chains intrinsically stabilize the α-helix structure. The following computational studies have addressed the possible existence or absence of helix propensity in polyalanines capped with Ac– and $-NH_2$. A quantum mechanical analysis by Park and Goddard[11] showed that polyalanine peptides retain the α-helical structure in solvent. Increasing helix propensity with increasing dielectric constant of the solvating medium is also reported by Wales and co-workers[12] via a master equation dynamics approach. On the other hand, Jortner and co-workers[23] found that dodecaalanine prefers a β-

dx.doi.org/10.1021/ct300709u | J. Chem. Theory Comput. 2012, 8, 4837–4844

hairpin structure over an α-helix structure in a molecular dynamics study. Scheraga and co-workers[10] reported decreased helical content with increasing solvent polarizability via a Monte Carlo investigation. The decrease in the helical propensity of alanine-based peptides upon hydration had also been reported earlier by Doruker and Bahar.[37]

The present work providing information about the relative thermodynamics of the α-helix to the coil structure provides some insight to this debate. In particular, the stabilization of the α-helix in solvent is approximately one-half of that as in vacuum. This is a result of the stabilization of the coil in the water solvent due to the hydrogen bonds that are formed therein. The fact that the tight α-helix structure at the one end of the ASMD stretch is not the most stable helical form is also reflected by the decrease in the free energy with initial stretching. However, the (local) minimum in the free energy is seen for a relatively tight α-helical structure. Whether this local minimum is a global minimum has not been resolved in this work, however, as there was no attempt to compare its thermodynamics to other closed forms such as the β-hairpin structure.

AUTHOR INFORMATION

Corresponding Author

*Fax: (404) 894-0594. E-mail: hernandez@chemistry.gatech.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

An insightful reading of the manuscript by Dr. Eliezer Hershkovits is gratefully acknowledged. This work has been partially supported by the National Science Foundation (NSF) through Grant No. CHE 1112067. The computing resources necessary for this research were provided in part by the National Science Foundation through TeraGrid resources provided by the Purdue Dell PowerEdge Linux Cluster (Steele) under grant number TG-CTS090079, and by the Center for Computational Molecular Science & Technology through Grant No. CHE 0946869.

REFERENCES

(1) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. *J. Chem. Phys.* **2003**, *119*, 3559.
(2) Park, S.; Schulten, K. *J. Chem. Phys.* **2004**, *120*, 5946.
(3) Minh, D. D. L.; McCammon, J. A. *J. Phys. Chem. B* **2008**, *112*, 5892.
(4) Echeverria, I.; Amzel, L. M. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1302.
(5) Ozer, G.; Quirk, S.; Hernandez, R. *J. Chem. Phys.* **2012**, *136*, 215104.
(6) Rohl, C. A.; Chakrabartty, A.; Baldwin, R. L. *Protein Sci.* **1996**, *5*, 2623.
(7) Spek, E. J.; Olson, C. A.; Shi, Z.; Kallenbach, N. R. *J. Am. Chem. Soc.* **1999**, *121*, 5571.
(8) Blondelle, S. E.; Forood, B.; Houghten, R. A.; Pérez-Payá, E. *Biochemistry* **1997**, *42*, 489.
(9) Williams, L.; Kather, K.; Kemp, D. S. *J. Am. Chem. Soc.* **1998**, *120*, 11033.
(10) Vila, J. A.; Ripoll, D. R.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13075.
(11) Park, C.; Goddard, W. A., III. *J. Phys. Chem. B* **2000**, *104*, 7784.
(12) Mortenson, P. N.; Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2002**, *117*, 1363.
(13) Marqusee, S.; Robbins, V. H.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5286.
(14) Kokubo, H.; Hu, C. Y.; Pettitt, B. M. *J. Am. Chem. Soc.* **2011**, *133*, 1849.
(15) Reiner, A.; Wildemann, D.; Fischer, G.; Kiefhaber, T. *J. Am. Chem. Soc.* **2008**, *130*, 8079.
(16) Jarzynski, C. *Phys. Rev. E* **1997**, *56*, 5018. Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690. Jarzynski, C. *Phys. Rev. E* **1997**, *56*, 5018.
(17) Crooks, G. E. *J. Stat. Phys.* **1998**, *90*, 1481.
(18) Hummer, G.; Szabo, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3658.
(19) Rodriguez-Gomez, D.; Darve, E.; Pohorille, A. *J. Chem. Phys.* **2004**, *120*, 3563. Geissler, P. L.; Dellago, C. *J. Phys. Chem. B* **2004**, *108*, 6667. Ytreberg, F. M.; Zuckerman, D. M. *J. Chem. Phys.* **2004**, *120*, 10876. Hatano, T.; Sasa, S. I. *Phys. Rev. Lett.* **2001**, *86*, 3463. Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2004**, *121*, 8742. Zimanyi, E. N.; Silbey, R. J. *J. Chem. Phys.* **2009**, *130*, 171102.
(20) Ozer, G.; Valeev, E.; Quirk, S.; Hernandez, R. *J. Chem. Theory Comput.* **2010**, *6*, 3026.
(21) Lu, N.; Wu, D.; Woolf, T. B.; Kofke, D. A. *Phys. Rev. E* **2004**, *69*, 057702. Ytreberga, F. M.; Zuckerman, D. M. *J. Chem. Phys.* **2004**, *120*, 10876. Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *122*, 204104. Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *123*, 084109. Cuendet, M. A.; Michielin, O. *Biophys. J.* **2008**, *95*, 35753590.
(22) Guttenberg, N.; Dinner, A. R.; Weare, J. *J. Chem. Phys.* **2012**, *136*, 234103.
(23) Levy, Y.; Jortner, J.; Becker, O. M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2188.
(24) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenburg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
(25) Liphardt, J.; Dumont, S.; Smith, S. B.; T., I., Jr.; Bustamante, C. *Science* **2002**, *296*, 1832. Douarche, F.; Ciliberto, S.; Petrosyan, A.; Rabbiosi, L. *Europhys. Lett.* **2005**, *70*, 593. Berkovich, R.; Klafter, J.; Urbakh, M. *J. Phys.: Condens. Matter* **2008**, *20*, 354008.
(26) Paramore, S.; Ayton, G. S.; Voth, G. A. *J. Chem. Phys.* **2007**, *126*, 051102.
(27) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *28*, 1781.
(28) Brooks, B. R.; Bruccoleri, R. E.; Olafson, R. E.; Sates, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(30) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.
(31) Mezei, M.; Fleming, P. J.; Srinivasan, R.; Rose, G. D. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 502.
(32) Kentsis, A.; Mezei, M.; Gindin, T.; Osman, R. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 493.
(33) Lopez, M. M.; Chin, D.-H.; Baldwin, R. L.; Makhatadze, G. I. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1298.
(34) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9190.
(35) Scholtz, J. M.; Baldwin, R. L. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 95.
(36) Baldwin, R. L. *Biophys. Chem.* **2002**, *101−102*, 203.
(37) Doruker, P.; Bahar, I. *Biophys. J.* **1997**, *72*, 2445.
(38) Armen, R.; Alonso, D. O.; Daggett, V. *Protein Sci.* **2003**, *12*, 1145.