Free energy surface of the helical peptide Y (MEARA) 6

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · NOVEMBER 2000

Impact Factor: 3.3 · DOI: 10.1021/jp002207k

4 AUTHORS, INCLUDING:

Amedeo Caflisch
University of Zurich
237 PUBLICATIONS **10,946** CITATIONS

SEE PROFILE

Free Energy Surface of the Helical Peptide Y(MEARA)$_6$

André Hiltpold, Philippe Ferrara, Jörg Gsponer, and Amedeo Caflisch*

Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received: June 16, 2000; In Final Form: August 17, 2000

The folding of the helical peptide Y(MEARA)$_6$ was studied by a series of molecular dynamics simulations with an implicit solvation model that allowed sampling of a total of more than 4 μs. In the 44 runs at 360 K started from all-coil conformations the peptide assumed an α-helical structure within the first 30 ns, with an average folding time of 10 ns. The free energy surface shows that the coil to helix transition has a small barrier at the helix nucleation step which consists of two to three $i, i + 4$ hydrogen bonds and does not show a strong preference along the sequence. On the helix side of the barrier, there is a very broad basin corresponding to conformations having more than one helical turn. Although the α-helical content is predominant, there is a nonnegligible percentage of conformations with one or more π-helical turns stabilized in part by interactions between Met side chains. Control simulations with two different helical sequences, a 31-residue polyalanine and A_5 $(AAARA)_3$ A, did not reveal a significant π-helix population, which indicates that the π-helical content of Y(MEARA)$_6$ is not an artifact of the force field and solvation model. The folding mechanism and free energy surface presented here are in agreement with previous theoretical models and experimental data on different helical sequences, which suggest that they may be valid for the folding of helical peptides, in general.

1. Introduction

In the majority of globular proteins, the native state contains regular elements of secondary structure, α-helices and/or β-sheets. Therefore, a common working hypothesis is that elucidating the formation of secondary structure will improve the understanding of the protein folding reaction.[1−3] α-Helices have a regular backbone conformation whose simplicity has made them very attractive for statistical mechanical descriptions since more than forty years.[4−7] Furthermore, during the past decade a large number of simulation studies have been published; the methods used are based on either energy minimization,[8] molecular dynamics[9−13] (MD) or Monte Carlo sampling techniques.[14] More recently, the use of implicit solvation model has allowed to simulate at an atomic level of detail the helix−coil transition at equilibrium.[15−17] On the experimental front, advances in laser-induced temperature jump methods have made possible the investigation of the kinetics of the helix−coil transition,[18−20] while the thermodynamics have been studied by circular dichroism (CD) and differential scanning calorimetry (DSC).[21,22]

Recently, Richardson et al.[22] have analyzed the structure and stability of the synthetic peptide Y(MEARA)$_6$ by CD and DSC. This repetitive sequence was "extracted" from a 60 amino acid domain of the human CstF-64 polyadenylation factor which contains 12 nearly identical repeats of the consensus motif MEAR(A/G). The CD and DSC data were insensitive to concentration indicating that Y(MEARA)$_6$ is monomeric in solution at concentrations up to 2 mM. The far UV−CD spectra indicates that the peptide has a helical content of about 65% at 1 °C. The DSC profiles were used to determine an enthalpy difference for helix formation of 0.8 kcal/mol per amino acid.

In this paper, the folding mechanism and energy surface of Y(MEARA)$_6$ are investigated by MD simulations with an implicit model for the solvent. The folding to the α-helical conformation is demonstrated by fifty-four runs started from the extended (7 runs at temperature values ranging from 330 to 390 K) and all-coil (3 and 44 runs at 300 and 360 K, respectively) conformations. The main goal of this study was to determine the free energy profile of the helix−coil transition and to compare it with the folding free energy surface of an antiparallel β-sheet.[23] The present simulation results indicate that the helix nucleation event, which consists of the formation of two to three α-helical hydrogen bonds, is not localized on a distinct region of the sequence. This implies that there are multiple pathways for the coil to helix transition, in agreement with MD simulations with explicit water[24] and recent experimental studies.[20]

2. Methods

2.1. Model. All peptides were modeled by explicitly considering all heavy atoms and the hydrogen atoms bound to nitrogen or oxygen atoms.[25] The aqueous solvent was approximated by an implicit model based on the solvent accessible surface:[26]

$$V_{\mathrm{solv}}(\mathbf{r}) = \sum_{i=1}^{N} \sigma_i A_i(\mathbf{r}) \qquad (1)$$

for a molecule having N heavy atoms with Cartesian coordinates $\mathbf{r} = (\mathbf{r}_1, ..., \mathbf{r}_N)$. $A_i(\boldsymbol{r})$ is the solvent accessible surface area of heavy atom i, computed by an approximate analytical expression[27] and using a 1.4 Å probe radius. Furthermore, ionic side chains were neutralized[28] and a linear distance-dependent dielectric function ($\epsilon(r) = 2r$) was used for the electrostatic interactions.[17] Since a cutoff is used, there is no significant difference between a linear distance-dependent dielectric function and a more sophisticated one, like a sigmoidal function,[29,30] because the deviation from linearity is small for distances

* Corresponding author. Phone: (41 1) 635 55 21. Fax: (41 1) 635 57 12. E-mail: caflisch@bioc.unizh.ch.

10.1021/jp002207k CCC: $19.00
Published on Web 10/11/2000

TABLE 1: Simulations Performed

Starting conformation	temp (K)	no. of simulations	length (ns)	av folding time[a] (ns)
folded[b]	275	1	100	
folded	300	1	100	
folded	360	1	100	
extended[c]	330	2	100	15.5
extended	360	2	100	11.2
extended	375	1	100	3.1
extended	390	2	100	7.8
all-coil[d]	300	3	100, 200[e]	80.3
all-coil	360	26	100	9.8
all-coil	360	18	30	8.9

[a] The definition of the folding time is given in section 2.5. [b] The α-helical conformation was generated with backbone dihedral values of $\phi = -57°$ and $\psi = -47°$. This conformation was minimized by 500 steps of steepest descent. [c] All backbone dihedral angles were set to 180°. [d] Three 10 ns molecular dynamics runs at 1000 K were started from the extended structure with different initial assignments of the velocities. Conformations were saved every 0.5 ns and quenched by 100 steps of steepest descent followed by a maximum of 10 000 steps of conjugate gradient minimization. A convergence criterion of 0.01 kcal/(mol Å) was reached in all minimizations. In the 44 conformations used for the runs from the all-coil, there was not any $i, i + 4$ hydrogen bond and the average value of the C_α RMSD from the α-helical state was 9.0 ± 1.8 Å. The all-coil conformation used as starting structure in Tr1 is shown in the lower right corner of Figure 2. [e] In one of the three runs, Y(MEARA)$_6$ did not reach the helical conformation within 100 ns; hence, the simulation was continued for another 100 ns.

smaller than 10 Å.[31,32] The CHARMM PARAM19 default cutoffs for long-range interactions were used, i.e., a shift function[25] was employed with a cutoff at 7.5 Å for both the electrostatic and van der Waals terms. This cutoff length was chosen to be consistent with the parametrization of the force field. The model contains only two σ parameters: one for carbon and sulfur atoms ($\sigma_{C,S} = 0.012$ kcal/(mol Å^2)), and one for nitrogen and oxygen atoms ($\sigma_{N,O} = -0.060$ kcal/(mol Å^2)).[33] It is important to note that there is no bias in the model toward any particular secondary structure type. In fact, exactly the same force field and implicit solvation model have been recently used to reversibly fold to the correct conformation by standard molecular dynamics a β-hairpin of 12 residues,[17] and two triple-stranded antiparallel β-sheet peptides whith sequence identity of only 15% (ref 23 and Ferrara and Caflisch, manuscript in preparation). Furthermore, the same force field and solvation energy were used to demonstrate with an atomistic model the non-Arrhenius behavior of the temperature dependence of the folding rate.[17]

2.2. Simulations. All simulations and most of the analysis of the trajectories were performed with the CHARMM program.[25] Constant temperature MD simulations were carried out by weak coupling to an external bath with a coupling constant of 5 ps.[34] The SHAKE algorithm[35] was used to fix the length of the covalent bonds having hydrogen atoms at one end. The Newton equation of motion was integrated with the leapfrog algorithm and an integration time step of 2 fs. The nonbonded interactions were updated every 10 dynamics steps and coordinate frames were saved every 10 ps. Table 1 contains a list of the simulations. A 100 ns run requires approximately 10 days on a 500 MHz Pentium III processor.

2.3. Progress Variables. An adequate choice of progress variables has to take into account the folded state conformation and its symmetries.[17,36] The full α-helical (π-helical) structure of Y(MEARA)$_6$ has 27 $i, i + 4$ (26 $i, i + 5$) hydrogen bonds. Q_α and Q_π are the fractions of formed backbone hydrogen bonds of type $i, i + 4$ and $i, i + 5$, respectively. Because of the regularity of the helical conformation, the following subsets of hydrogen bonds are particularly appropriate to project the energy surface: Q_N and Q_C are the fraction of $i, i + 4$ hydrogen bonds involving the carbonyl oxygen of residues 1−14 and 15−27, respectively. Following criteria were used for the definition of a hydrogen bond: the distance between hydrogen and acceptor atom has to be smaller than 3.0 Å and the angle between donor, hydrogen, and acceptor larger than 120°. With these criteria, it is very unlikely that the carbonyl of residue i is involved in a bifurcated hydrogen bond with the NH groups of residues $i + 4$ and $i + 5$. Although a hydrogen bonding distance of 3.0 Å is rather large, it yields about 57% α-helical content at 300 K in agreement with the value obtained by the DSSP[37] analysis (see below). A hydrogen bond cutoff of 2.5 Å yields an α-helical content of about 40% at the same temperature.

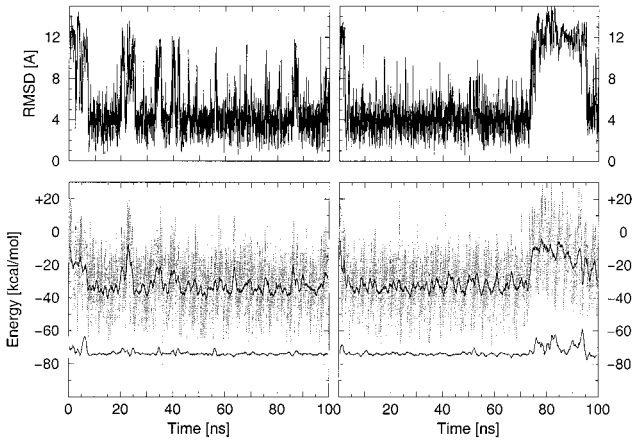

Figure 1. Time series of trajectories 1 (Tr1, left panels) and 8 (Tr8, right panels) started from all-coil conformations at 360 K. (Top panels) Evolution of the C_α RMSD (Å) from the α-helical conformation as a function of time. (Bottom panels) The gray lines represent the evolution of the effective energy (upper curves) and solvation energy (bottom curves) as a function of time. The black lines are running averages over 100 ps intervals.

2.4. Effective Energy and Free Energy. For the understanding of protein folding, the important role of effective energy and free energy surfaces, determined by simulations and experiments, has been reviewed recently.[38] The effective energy is the sum of the intramolecular energy (CHARMM PARAM19 force field energy) and the solvation free energy. The latter is approximated by the solvent accessible surface term of eq 1 and contains the entropic contribution of the solvent within the approximations of an implicit model of the water molecules. The effective energy does not include the configurational entropy which consists of conformational and vibrational entropy contributions.[28] The plots on the left of Figures 3 and 4 show the values of the effective energy averaged within a bin defined by discretizing the (Q_N,Q_C) space (Figure 3) and the (Q_α,Q_π) space (Figure 4).

For a system in thermodynamic equilibrium, the difference in free energy in going from a state A to a state B is proportional to the natural logarithm of the quotient of the probability of finding the system in state A divided by the probability of state B. The free energy surface is projected on the aforementioned two-dimensional spaces of progress variables (plots on the right of Figures 3 and 4) by using an arbitrarily chosen reference bin as the denominator of the probability quotient.

2.5. Folding Time. The coil to helix transition is considered completed when the C_α RMSD from the minimized α-helical conformation reaches a value smaller than 2.0 Å and its running average over 100 ps is smaller than 3.0 Å. The latter condition is important to avoid counting transient folding events.

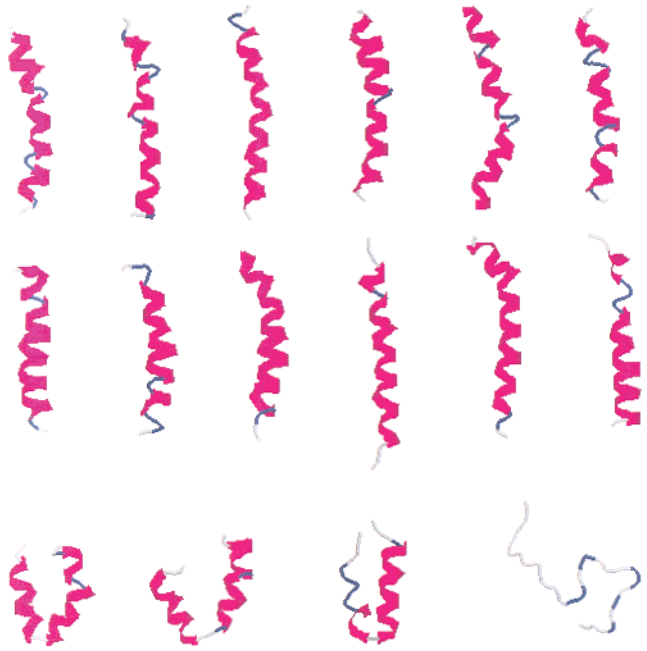

Figure 2. Cluster analysis of the 5000 conformations saved every 20 ps along Tr1. A C_α RMSD cutoff of 3.0 Å and the procedure explained in section 2.6 were used for clustering. The first two rows show the representatives of the 12 largest clusters. They contain from left to right the following percentages of population: 5.1, 4.6, 4.5, 3.3, 3.1, 2.7, 2.7, 2.5, 2.0, 1.9, 1.8, and 1.7. The third row contains three snapshots saved at about 40 ns (they are representatives of clusters with population smaller than 0.2%) and the initial all-coil conformation obtained by MD at 1000 K. Helical stretches are colored in red, while reverse turns are in blue. This figure was made with RasMol.

2.6. Cluster Analysis. The method for cluster analysis is based on structural similarity, i.e., C_α RMSD after optimal superposition. The procedure is iterative and simple. The first conformation is the representative of the first cluster. For each remaining conformation, the cluster representative with the smallest C_α RMSD deviation is identified. If the deviation is smaller than a given cutoff value (3 Å is used in this study) the conformation is assigned to that cluster, otherwise a new cluster is created.[39] This cluster analysis procedure is very fast and allows us to cluster several thousands of conformations in few minutes. It is implemented in the molecular modeling program WITNOTP (A. Widmer, Novartis Pharma, unpublished).

3. Results and Discussion

3.1. Control Runs and Folding from the Extended Conformation. Three control simulations at temperature values of 275, 300, and 360 K were started from the α-helical structure (Table 1). The system was stable and visited mainly helical conformations with a percentage of formed i, $i + 4$ hydrogen bonds of 59%, 57% and 51% at 275, 300, and 360 K, respectively. This is in agreement with the 65% helical content measured by CD at 1 °C.[22] The relatively high stability at 360 K is probably an artifact due to the approximations inherent to the force field and solvation model.[17]

Sampling a statistically significant number of folding events at 300 K is too time-consuming since three simulations started from the high-temperature denaturated state reached the helical conformation in 31, 40, and 170 ns, respectively. To determine a temperature value for the folding studies, seven simulations were started from the extended conformation with temperatures ranging from 330 to 390 K (Table 1). The two runs at 360 K reached a stable helical state at 8.3 and 14.1 ns, respectively. Therefore, the thermodynamics and kinetics of folding were investigated at 360 K, with 44 simulations started from all-coil conformations. The following analysis focuses on the results obtained at 360 K.

3.2. Folding from All-Coil Conformations. The $Y(MEARA)_6$ peptide assumed an α-helical structure within the first 30 ns in all runs at 360 K. The time series of the C_α root-mean-square deviation (RMSD) from the α-helical structure show large fluctuations (Figure 1). The system samples a large amount of different conformations in the energy basin of the helical state (see below). These include conformations which are mainly helical but have one or both terminal regions frayed (Figure 2).

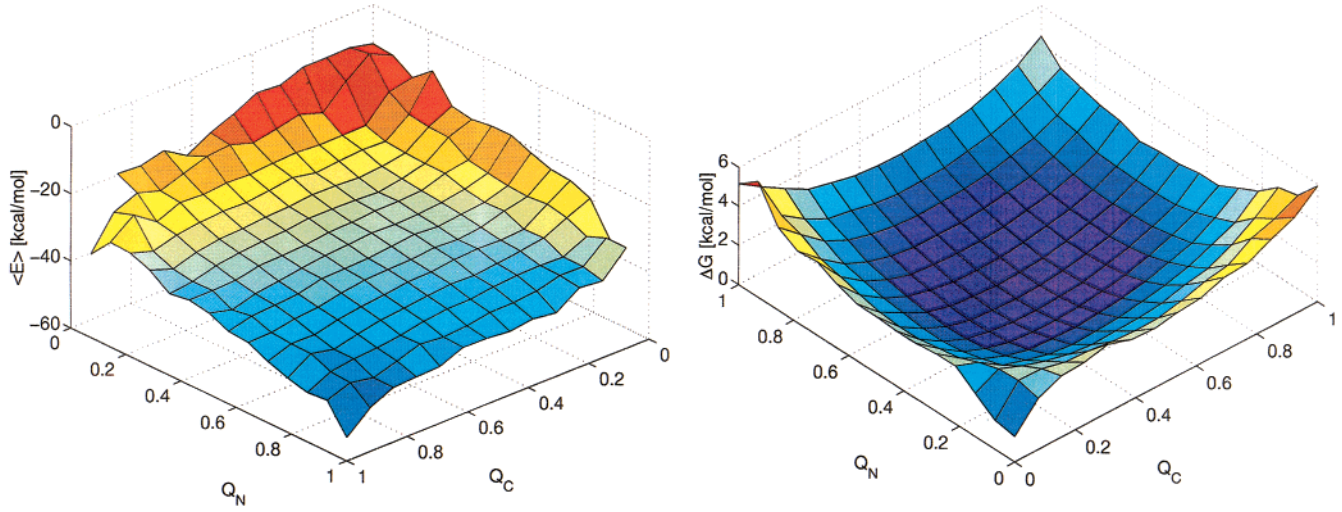

Figure 3. Average effective energy ($\langle E \rangle$, left) and free energy surface (ΔG, right) at 360 K as a function of the fraction of i, $i + 4$ hydrogen bonds in residues 1−14 (Q_N) and 15−27 (Q_C). A total of 1.82×10^5 conformations were used; these were sampled during the 30−100 ns intervals of the 26 100 ns simulations started from the all-coil conformations. To make it more clear, the plot of ΔG is rotated with respect to the one of $\langle E \rangle$ by about 180° around an axis going through the center of the horizontal plane. $\langle E \rangle$ was evaluated by averaging the effective energy values of the conformations within a bin without minimizing them. Because of insufficient statistics, only $\langle E \rangle$ values for data points with $M_{n,c} > 4$ are included in the plot, where $M_{n,c}$ denotes the number of conformations with n and c α-helical hydrogen bonds formed between residues 1−14 and 15−27, respectively. ΔG was computed as $-k_B T \ln(M_{n,c}/M_{7,6})$. The minimum and maximum values of $M_{n,c}$ are 1 and 4018, respectively. The error is estimated by separating the 26 simulations into two sets of 13 simulations each. The average and maximal errors of $\langle E \rangle$ are 0.7 and 25.9 kcal/mol (bin with $n = 14$ and $c = 0$), respectively. The average and maximal errors of ΔG are 0.1 and 5.5 kcal/mol ($n = 13$, $c = 0$), respectively.

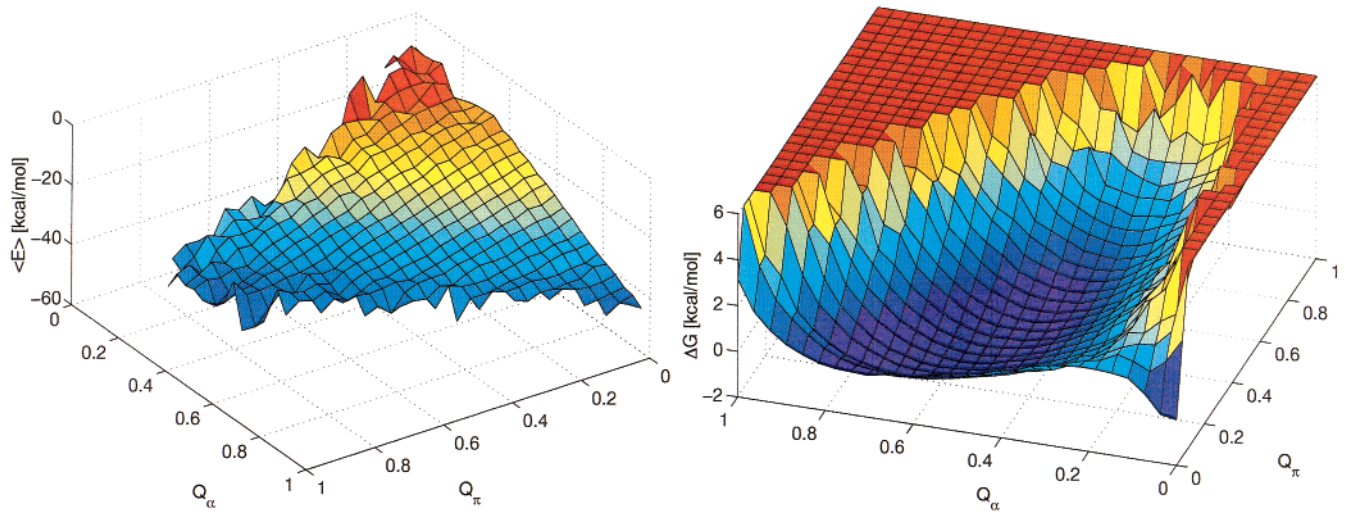

Figure 4. Same as in Figure 3 for the average effective energy ($\langle E \rangle$, left) and free energy surface (ΔG, right) at 360 K as a function of the fraction of α-helical and π-helical hydrogen bonds. Because of insufficient statistics, only $\langle E \rangle$ values for data points with $N_{a,p} > 4$ are included in the plot, where $N_{a,p}$ denotes the number of conformations with a and p α-helical and π-helical hydrogen bonds, respectively. ΔG was computed as $-k_B T \ln(N_{a,p}/N_{0,0})$. The minimum and maximum values of $N_{a,p}$ are 0 and 2203, respectively. ΔG was set to a value of 6.0 kcal/mol when $N_{a,p} = 0$. The error is estimated by separating the 26 simulations into two sets of 13 simulations each. The average and maximal errors of $\langle E \rangle$ are 1.5 and 55.0 kcal/mol (bin with $a = 26$ and $p = 2$), respectively. The average and maximal errors of ΔG are 0.2 and 5.0 kcal/mol (10 different bins along the rim), respectively.

Furthermore, the helix can be either slightly curved or completely bent in the middle with two helices folded together in an antiparallel arrangement. Only two events of complete unfolding to the all-coil conformation were sampled. In trajectory 39 (Tr39), Y(MEARA)$_6$ folded after 5 ns and unfolded to the all-coil state at 29 ns (not shown). The peptide then refolded at about 30 ns. In Tr8, the system crossed the barrier in the direction of the all-coil state at about 75 ns and remained completely unfolded for about 20 ns (Figure 1). A cluster analysis of the conformations sampled during the 75−95 ns interval of Tr8 was performed using a C_α RMSD cutoff of 3.0 Å, as explained in section 2.6. This analysis revealed that although the β-sheet content in the all-coil state is significant, there is not any predominant structure. The same was found by an analysis of the simulation intervals preceding helix formation. These simulation results indicate that for Y(MEARA)$_6$ the all-coil state corresponds to the random coil and there are not stable intermediates at 360 K.

Upon folding, the effective energy (intramolecular plus solvation) decreases from about −10 kcal/mol to values oscillating between −65 and 0 kcal/mol with an average of −32 ± 13 kcal/mol (Figure 1). The solvation energy in the helical state is almost constant and has much smaller oscillations around an average value of about −74 ± 4 kcal/mol. In the all-coil conformations sampled in the 75−95 ns interval of Tr8, the average values of the effective and solvation energy are −12 ± 14 and −69 ± 6 kcal/mol, respectively. Since the all-coil to helix effective energy difference is about −20 kcal/mol and the difference in solvation is only −5 kcal/mol, most of the stabilization of the helical state is due to the intramolecular energy which is about −15 kcal/mol more favorable than in the all-coil state. Hence, the 360 K simulations yield an enthalpic stabilization of about −0.5 kcal/mol per residue which is smaller in absolute value than the −0.8 kcal/mol contribution extracted from the calorimetric data at 275 K.[22] The discrepancy may be due to the different temperatures and the fact that the value from the simulations does not include the helix−water energy.

To evaluate the effects of the implicit solvation model, two test simulations of 100 ns each were performed at 360 K without the solvent accessible surface term. One was started from the completely extended conformation and the other from one of the all-coil conformations obtained by high-temperature molecular dynamics. In both simulations the peptide collapsed into random coil conformations but no folding event was observed. The average value of the radius of gyration in the two runs was 8.0 Å. As a basis of comparison, the radius of gyration of the α-helical conformation is 14.0 Å. Therefore, the SAS solvation energy has a nonnegligible effect on the free energy surface, although at 360 K it varies by only −5 kcal/mol on average between the all-coil and α-helical states. The small variation is due to the relatively small size of the system and the fact that a helix does not have a hydrophobic core.

3.3. Energy Surface. The average effective energy as a function of Q_N and Q_C has a downhill profile with a single minimum corresponding to the fully helical conformation (Figure 3a). The free energy surface shows a very broad minimum corresponding to the helical state and another minimum at the all-coil state separated by a low free energy barrier of about $k_B T$ in the direction of folding and about $5k_B T$ in the opposite direction (see also Figure 5). The barrier corresponds to the nucleation step and is therefore much closer to the all-coil state than the helical state. Since the effective energy decreases almost monotonically by increasing Q_N and/or Q_C, the all-coil side of the barrier originates from the loss of conformational entropy due to the formation of two to three i, $i + 4$ hydrogen bonds. Comparing the effective energy and free energy surfaces one notices that the broad minimum in the latter is significantly shifted away from the fully α-helical conformation. This is a mainly entropic effect; for a long helical peptide there are many favorable conformations with one or more frayed turns (Figure 2). The free energy surface shows a relatively high degree of symmetry with respect to the diagonal of the $Q_N Q_C$ plane, which is due to the regularity in the sequence and structure of Y(MEARA)$_6$. The symmetry and smoothness of the free energy surface indicate that the statistical error is small (see also caption of Figure 3). A symmetrical free energy surface was recently found for the folding of a 20-residue three-stranded antiparallel β-sheet at temperature of 330 and 360 K.[23] The free energy surfaces of Y(MEARA)$_6$ and the antiparallel β-sheet peptide differ mainly in the height and location of the folding

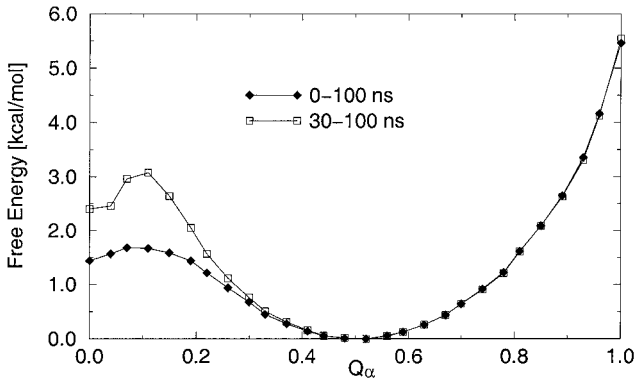

Figure 5. Free energy at 360 K as a function of the fraction of α-helical hydrogen bonds, calculated by $\Delta G = -k_B T \ln(N_a/N_{14})$. (Filled diamonds) A total of 2.60×10^5 conformations were used; these were sampled during the 26 100 ns simulations started from the all-coil conformations. (Empty squares) A total of 1.82×10^5 conformations were used; these were sampled during the 30−100 ns intervals of the 26 100 ns simulations started from the all-coil conformations.

barrier, which is much lower and closer to the fully unfolded state in the former. In the β-sheet peptide, the folding barrier measures about $5k_BT$ and its peak occurs just after the almost complete formation of one of the two β-hairpins.[23]

Because of the significant occurrence of π-helical turns (see below), it is interesting to project the energy values in the (Q_α,Q_π) plane. The average effective energy as a function of Q_α and Q_π has a downhill profile, apart from irregularities due to the insufficient sampling for small values of Q_α (Figure 4a). The free energy surface indicates that nucleation consists of formation of two to three α-helical hydrogen bonds (Figure 4b). After nucleation, there are no barriers for the interconversion between α- and π-helical content. Conformations with only π-helical hydrogen bonds are very unlikely because of the large entropy loss.[40] The presence of π-helical turns require that at least 35% of the i, $i + 4$ hydrogen bonds of Y(MEARA)$_6$ are formed. On the other hand, a significant amount of α-helical turns can be formed independently of the π-helical content. In fact, the broad minimum corresponding to the folded state encompasses conformations with up to about 75% α-helical content and less than 10% i, $i + 5$ hydrogen bonds.

The two curves in Figure 5 represent the 360 K equilibrium profile of the free energy as a function of Q_α (empty squares) and the corresponding profile for the equilibration kinetics starting from the all-coil state (filled diamonds). The curves overlap for $Q_\alpha > 0.3$ and have a global minimum corresponding to about 50% α-helical content. As mentioned above, conformations with more than 90% i, $i + 4$ hydrogen bonds are very unlikely and their free energy is higher than the all-coil state. The free energy of the all-coil state and the height of the coil to helix barrier are higher at equilibrium (30−100 ns simulation intervals) than for the complete trajectories started from the all-coil state (0−100 ns simulation intervals). This is due to the fact that the system samples the all-coil and transition state conformations almost exclusively within the first 30 ns of the simulations started from the all-coil state. Once the system has reached the helical basin, it almost never unfolds to the all-coil state (except for Tr8 and Tr39, as mentioned above).

3.4. Nucleation Mechanism and Kinetics. The 44 trajectories from the all-coil state at 360 K show that the nucleation is not localized in a distinct region of the sequence and that there are many different folding pathways (Figure 6) in agreement with previous simulation studies on the helical peptide (AAQAA)$_3$.[17] There is a significant content of β-sheet in many

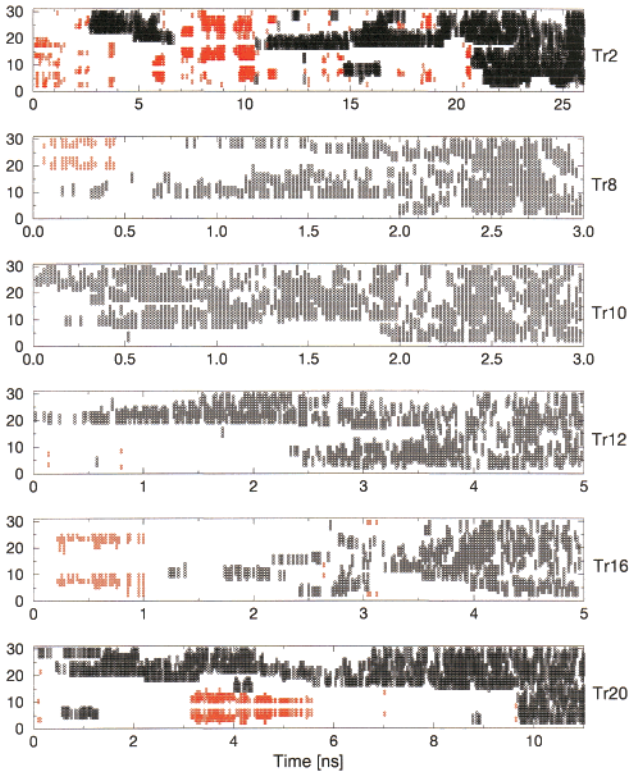

Figure 6. DSSP secondary structure content as a function of time for six runs from the all-coil conformations at 360 K. The *y*-axis indicates position along the sequence. Red diamonds, β-sheet; black circles, α-helix. Only the time interval preceding the folding event is shown.

of the runs where the peptide required more than 5 ns to fold (Figure 6). This indicates that conformations with some β-sheet content represent transient misfolded states of Y(MEARA)$_6$. The β-sheet content can be either spread over all of the sequence, as in the 6−10 ns interval of Tr2 and the first nanosecond of Tr16, or localized in a segment, as in the 3−6 ns interval of Tr20. In the latter, there is a two stranded antiparallel β-sheet involving residues 3 to 14 packed against a helical conformation in the C-terminal region of the sequence.

For each run, the last structure with two i, $i + 4$ hydrogen bonds sampled before the folding transition was analyzed. These conformations lead directly to higher Q_α values, i.e., into the helical basin. The analysis of the 44 conformations revealed that the two nucleating i, $i + 4$ hydrogen bonds are nonuniformly distributed along the sequence with a preference for the C-terminal region. In fact, there are about 50% more nucleations involving the CO group of residues 19−27 than either 1−9 or 10−18. This may be due in part to the presence of the Tyr residue at the N-terminus since mutagenesis experiments indicate that Tyr destabilizes a α-helix by about 0.6 kcal/mol with respect to Ala.[41] Moreover, the segment 19−27 contains four alanine residues while there are only three alanines in the segments 1−9 and 10−18. Only in 9/44 cases the two nucleating hydrogen bonds were contiguous. This is illustrated in Figure 6, which shows that the helix can nucleate at two sites far apart in the sequence. The nucleations at different sequence positions can be almost simultaneous as in the first nanosecond of Tr8, or separated by a time gap ranging from about 2 ns (Tr12) to 10 ns (Tr2). The most frequent nucleation site (seven events) consists of the CO of residue 23 and NH of residue 27, and there were 3, 7, and 0 i, $i + 4$ residue pairs with 2, 1 and 0, nucleation events. Since the nucleation barrier is located at about two to three i, $i + 4$ hydrogen bonds, the analysis was repeated for the last conformation with three α-helical hydrogen bonds

sampled before folding. A similar picture emerged, but now two of the three hydrogen bonds involve residues which are contiguous in the sequence in 29/44 cases.

The evolution of the population of all-coil conformations as a function of time is reflected in the percentage of simulations that have not reached the α-helical structure at a given time. A single-exponential fit yields a characteristic time of decay of 10.2 ns (correlation coefficient of 0.98), while the folding time (as defined in section 2.5) averaged over the forty-four simulations is 9.4 ns. No improvement in the χ^2 was observed when fitting a double exponential. Therefore, the decrease of the all-coil population is essentially monoexponential. Such a behavior was observed previously for an alanine-based α-helical peptide.[17] Although the folding rate of Y(MEARA)$_6$ has not been measured experimentally, the peptide folds probably faster in the MD simulations with implicit solvent than in the test tube. For 21-residue synthetic peptides containing mainly alanine, measurement of the helix−coil transitions initiated by laser temperature jumps indicate that the relaxation time ranges from about 20 to 300 ns at ambient temperature.[18,20] This is not inconsistent with the 80 ns value obtained from the three simulations at 300 K. The accelerated coil to helix transition observed at 360 K is due in part to the higher temperature. Nevertheless, it cannot be ruled out that part of the effect is due to the implicit solvation model because of the lack of friction exerted by the water molecules and the smooth solute−solvent potential energy provided by the mean solvent model.[17,23] Recently, the conformational transition between the A- and B-DNA forms was observed to be about 20 times faster when using the generalized Born implicit solvation model than explicit water molecules.[42]

3.5. *π*-Helical Content. The snapshots saved during the 100 ns simulations started from the α-helical conformation were submitted to a DSSP analysis.[37] The DSSP α-helical content is almost constant at about 55% in the temperature range 275 to 360 K. On the other hand, the DSSP *π*-helical content decreases at high temperature; it is 31%, 26% and 14% at 275, 300, and 360 K, respectively. The temperature dependence is probably due to the entropic penalty in fixing the side chains which is larger at elevated temperatures because of the higher conformational freedom in the unfolded state. The nonnegligible *π*-helical content is mainly due to the periodicity in the sequence which promotes the favorable contacts between Met side chains with an $i, i + 5$ spacing. The latter compete with the salt bridges between Arg and Glu at position i and $i + 3$, respectively. At 300 K, the percentage of Arg-Glu salt bridges for residues 5−8, 10−13, 15−18, 20−23 and 25−28 is 13.3, 8.0, 4.6, 5.6 and 6.5, respectively. Using a 6.0 Å threshold for the distance between the sulfur atoms, the percentage of Met-Met side chain contacts for residues 2−7, 7−12, 12−17, 17−22 and 22−27 is 2.1, 5.7, 7.7, 9.1 and 7.9, respectively. The maximum at 17−22 and the distribution along the sequence correlate with the *π*-helical content shown in Figure 7d. Furthermore, it is clear from a comparison of the time series of the C_α RMSD from the *π*-helix and the distance between sulfur atoms of Met side chains at positions $i, i + 5$ (not shown) that the formation of a *π*-helical turn is concomitant with the formation of the corresponding Met−Met side chain contact. The stretches of *π*-helix usually are not longer than two turns as in the few occurrences of *π*-helix in the PDB database (see Table 1 of ref 40). The present simulation results are not in disagreement with the CD data of Richardson et al.[22] since the dichroic spectrum of a *π*-helix might be indistinguishable from the α-helical one (G. I. Makhatadze, personal communication).

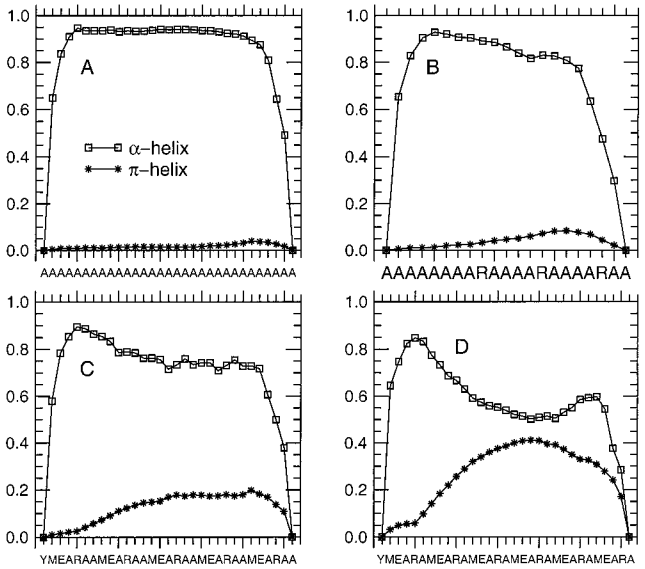

Figure 7. DSSP results of the 100 ns runs from the α-helical structure at 300 K. The *x*-axis indicates position along the sequence and the *y*-axis the percentage of α-helix (empty squares) and *π*-helix (asterisks). Key: (a) 31-residue polyalanine, (b) A_5 $(AAARA)_3$, (c) Y(MEARAA)$_5$, and (d) Y(MEARA)$_6$.

To check that the high *π*-helical content of Y(MEARA)$_6$ is not an artifact of the force field and solvation model, several 100 ns control simulations at 300 K were carried out on a 31-residue polyalanine and the peptide A_5 $(AAARA)_3$. These were started from either the minimized α-helical structure or random conformations. Both peptides show a negligible *π*-helical propensity and a very high α-helical content (Figure 7a,b) in agreement with experimental results.[7,18] To verify that the *π*-helical content of Y(MEARA)$_6$ might be a consequence of the spacing of 5 residues between Met side chains, a 300 K simulation of the sequence repeat MEARAA was performed. Although the distribution along the sequence is similar, the amount of *π*-helix is much smaller in Y(MEARAA)$_5$ than in Y(MEARA)$_6$ (Figure 7c,d).

4. Conclusions

The length of Y(MEARA)$_6$ makes it difficult to study helix formation by MD simulations with explicit water molecules. Therefore, multiple MD runs were performed with an implicit solvation model. The same model had been previously used to investigate the thermodynamics and kinetics of folding of three synthetic peptides: a 12-residue *β*-hairpin and 15-residue α-helix,[17] as well as a 20-residue three-stranded antiparallel *β*-sheet.[23] The absence of explicit water molecules facilitates the folding transition since the atoms of the solute do not feel any friction with the solvent but only the intrasolute friction. The present simulation results indicate that the synthetic peptide Y(MEARA)$_6$ assumes a mainly α-helical structure with a nonnegligible content of *π*-helix. This is not inconsistent with the currently available experimental evidence.[22] A significant *π*-helical content was found previously by explicit solvent molecular dynamics simulations of the peptides $(AAQAA)_3$ and $(AAKAA)_3$,[13] which provides further evidence that the *π*-helical content of Y(MEARA)$_6$ is not an artifact of the approximations inherent to the solvation model.

An exponential decay of the unfolded population is common to both Y(MEARA)$_6$ and the 20-residue three-stranded antiparallel *β*-sheet[43] previously investigated by MD at the same temperature (360 K).[23] The free energy surfaces of Y(MEARA)$_6$

and the antiparallel β-sheet peptide differ mainly in the height and location of the folding barrier, which in $Y(MEARA)_6$ is much lower and closer to the fully unfolded state. The main difference between the two types of secondary structure formation consists of the presence of multiple pathways in the α-helix and only two predominant pathways in the three-stranded β-sheet. The helix can nucleate everywhere, with a preference for the C-terminal third of the sequence in $Y(MEARA)_6$. Furthermore, two nucleation sites far apart in the sequence are possible. Folding of the three-stranded antiparallel β-sheet peptide started with the formation of most of the side chain contacts and hydrogen bonds between strands 2 and 3, followed by the 1−2 interstrand contacts. The inverse sequence of events, i.e., first formation of 1−2 and then 2−3 contacts, was also observed, but less frequently.[23]

The free energy barrier seems to have an important entropic component in both helical peptides and antiparallel β-sheets. In an α-helix, it originates from constraining the backbone conformation of three consecutive amino acids before the first helical hydrogen bond can form, while in the antiparallel β-sheet it is due to the constraining of a β-hairpin onto which a third strand can coalesce.[23] Therefore, the folding of the two most common types of secondary structure seems to have similarities (a mainly entropic nucleation barrier and an exponential folding rate) as well as important differences (location of the barrier and multiple vs two pathways). The similarities are in accord with a plethora of experimental and theoretical evidence[44] while the differences might be a consequence of the fact that $Y(MEARA)_6$ has about 7−9 helical turns whereas the three-stranded antiparallel β-sheet consists of only 2 "minimal blocks", i.e., two β-hairpins. In this context, it will be interesting to investigate if in a n-stranded β-sheet (with $n > 3$) the first β-hairpin can nucleate anywhere in analogy to the lack of positional preference for the nucleating turn in α-helices.

Acknowledgment. We thank U. Haberthür for interesting discussions and N. Budin for computer support. We also thank A. Widmer (Novartis Pharma, Basel) for providing the molecular modeling program WITNOTP which was used for visual analysis of the trajectories and cluster analysis. P.F. is a Fellow of the Roche Research Foundation. This work was supported in part by the Swiss National Science Foundation (grant no. 31- 53604.98 to A.C.).

References and Notes

(1) Lacroix, E.; Kortemme, T.; de la Paz, M. L.; Serrano, L. *Curr. Opin. Struct. Biol.* **1999**, *9*, 487−493.
(2) Imperiali, B.; Ottesen, J. J. *J. Peptide Res.* **1999**, *54*, 177−184.
(3) Crane, J. C.; Koepf, E. K.; Kelly, J. W.; Gruebele, M. *J. Mol. Biol.* **2000**, *298*, 283−292.
(4) Schellman, J. A. *J. Phys. Chem.* **1958**, *62*, 1485−1494.
(5) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526−535.
(6) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963−1974.
(7) Rohl, C. A.; Fiori, W.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3682−3687.
(8) Czerminski, R.; Elber, R. *J. Chem. Phys.* **1990**, *92*, 5580−5601.
(9) DiCapua, F. M.; Swaminathan, S.; Beveridge, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 6768−6771.
(10) Soman, K. U.; Karimi, A.; Case, D. A. *Biopolymers* **1991**, *31*, 1351−1361.
(11) Tirado-Rives, J.; Jorgensen, W. L. *Biochemistry* **1991**, *30*, 3864−3871.
(12) Daggett, V.; Levitt, M. *J. Mol. Biol.* **1992**, *223*, 1121−1138.
(13) Shirley, W. A.; Brooks, C. L., III. *Proteins: Struct., Funct. Genet.* **1997**, *28*, 59−71.
(14) Hansmann, U. H. E; Okamoto, Y. *J. Chem. Phys.* **1999**, *110*, 1267−1276.
(15) Sung, S.; Wu, X. *Proteins: Struct., Funct. Genet.* **1996**, *25*, 202−214.
(16) Schaefer, M.; Bartels, C.; Karplus, M. *J. Mol. Biol.* **1998**, *284*, 835−848.
(17) Ferrara, P.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. B* **2000**, *104*, 5000−5010.
(18) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. *Biochemistry* **1997**, *36*, 9200−9210.
(19) Lednev, I. K.; Karnoup, A. S.; Sparrow, M. C.; Asher, S. A. *J. Am. Chem. Soc.* **1999**, *121*, 8074−8086.
(20) Thompson, P. A.; Muñoz, V.; Jas, G. S.; Henry, E. R.; Eaton, W. A.; Hofrichter, J. *J. Phys. Chem. B* **2000**, *104*, 378−389.
(21) Taylor, J. W.; Greenfield, N. J.; Wu, B.; Privalov, P. L. *J. Mol. Biol.* **1999**, *291*, 965−976.
(22) Richardson, J. M.; McMahon, K. W.; MacDonald, C. C.; Makhatadze, G. I. *Biochemistry* **1999**, *38*, 12869−12875.
(23) Ferrara, P.; Caflisch, A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10780−10785.
(24) Takano, M.; Yamato, T.; Higo, J.; Suyama, A.; Nagayama, K. *J. Am. Chem. Soc.* **1999**, *121*, 605−612.
(25) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(26) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199−203.
(27) Hasel, W.; Hendrickson, T. F.; Still, W. C. *Tetrahedron Comput. Methodol.* **1988**, *1*, 103−116.
(28) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct. Genet.* **1999**, *35*, 133−152.
(29) Mehler, E. L.; Eichele, G. *Biochemistry* **1984**, *23*, 3887−3891.
(30) Ramstein, J.; Lavery, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7231−7235.
(31) Mehler, E. L. *Protein Eng.* **1990**, *3*, 415−417.
(32) Schaefer, M.; Bartels, C.; Karplus, M. *Theor. Chem. Acc.* **1999**, *101*, 194−204.
(33) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins: Struct., Funct. Genet.* **2000**, *39*, 252−260.
(34) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(35) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(36) Dinner, A. R.; Karplus, M. *J. Phys. Chem. B* **1999**, *103*, 7976−7994.
(37) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.
(38) Dinner, A. R.; Sali, A.; Smith, L. J.; Dobson, C. M.; Karplus, M. *TIBS* **2000**, *25*, 331−339.
(39) Spaeth, H. *Cluster-Analyse-Algorithmen zur Objektklassifizierung und Datenreduktion*; Oldenburg: Muenchen-Wien, 1977.
(40) Weaver, T. M. *Protein Sci.* **2000**, *9*, 201−206.
(41) Fersht, A. *Structure and Mechanism in Protein Science*; W. H. Freeman and Co.: New York, 1999.
(42) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489−2498.
(43) De Alba, E.; Santoro, J.; Rico, M.; Jiménez, M. A. *Protein Sci.* **1999**, *8*, 854−865.
(44) Karplus, M. *J. Phys. Chem. B* **2000**, *104*, 11−27.