Thermodynamics and Kinetics of Folding of Two Model Peptides Investigated by Molecular Dynamics Simulations

Philippe Ferrara, Joannis Apostolakis,† and Amedeo Caflisch*

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

Received: November 23, 1999; In Final Form: March 9, 2000

The folding of an α-helix and a β-hairpin was studied by 862 molecular dynamics simulations with an implicit solvation model that allowed sampling of a total of 4 μs. The average effective energy is rather flat for conformations having less than about 50% of the folded state contacts formed, except for the α-helix at very high temperatures. For both peptides there is a smooth decrease of the effective energy close to the folded state. The free energy landscape shows that the helix−coil transition is not first order, while the β-hairpin has one or two minima, depending on the temperature. At low temperature ($T < 1.1T_m$) there is an increase in the folding rate with increasing temperature as expected from an activation energy limited process. At higher temperatures the rate decreases for both peptides which is consistent with an activation entropy dominated process. The unfolding rate, by contrast, shows an Arrhenius-like behavior; i.e., it increases monotonously with temperature. The β-hairpin peptide folds about 30 times slower than the α-helix peptide at 300 K. Multiple folding pathways are present for the α-helix, whereas the β-hairpin initiates folding mainly at the β-turn.

1. Introduction

The understanding of the mechanism of secondary structure formation is thought to shed light on the protein folding problem. Of the two elements of regular secondary structure present in proteins, i.e., α-helices and β-sheets, the former have been the subject of many experimental and theoretical studies in the past (see refs 1−5 for reviews). Despite the simplicity of an α-helix with respect to a globular protein, it still represents a challenge to simulate the folding of a helical peptide using an all-atom description of the solute−solvent system. The introduction of the self-guided molecular dynamics (MD) simulation method has allowed the folding of a 16-residue alanine-based peptide in explicit water.[6] Very recently, a stochastic approach for reaction path calculations has been proposed and applied to the folding of the C peptide of ribonuclease A in water solution.[7] The reversible folding of a helical β-heptapeptide has been studied by MD simulations in methanol.[8] Simulations with explicit water molecules[9] require more computational time because of the higher density of water than methanol. Much less has been done on β-sheets (see refs 3 and 4 for a review). The shortest β-structure is the β-hairpin, which is made of two antiparallel strands connected by a β-turn. β-Hairpins are thought to fold much slower than α-helices.[10] Partial refolding of a β-hairpin corresponding to the residues 85−102 of barnase has been observed in MD simulations in water.[11] Monte Carlo[12] and MD[13] simulations of synthetic β-hairpins with an implicit solvation model have been reported. Very recently, the C-terminal β-hairpin fragment of protein G B1 has been investigated by Monte Carlo folding simulations using an implicit solvent model[14] and independently by explicit water MD simulations of unfolding and refolding from partially unfolded conformations.[15]

The main goal of the present simulation study was to determine the temperature dependence of the folding reaction and at the same time investigate the effective energy and free energy profiles at several values of the temperature ranging from far below up to far above the melting point (T_m). We present equilibrium and kinetics data obtained from 862 MD simulations (each lasting between 30 ps and 100 ns) of two synthetic peptides at different temperatures with an implicit solvation model. The first peptide, whose sequence is $(AAQAA)_3$, has been the subject of experimental[16] and theoretical[17,18] studies. Both types of studies have pointed out that it folds into an α-helix with a percentage of helicity of about 50% at 300 K. From our simulations at 300 K, the α-helical conformation has a probability of about 60%, and on average 65% of the backbone hydrogen bonds are formed, in reasonable agreement with the experimental data.[16]

The second peptide has the sequence Val_5-D-Pro-Gly-Val_5. Short peptides containing a central D-Pro-Gly (^{D}PG) segment adopt a β-hairpin conformation with a two-residue loop at ^{D}PG.[19−22] In the MD simulations at 300 and 330 K, the β-hairpin structure has a probability of about 99% and 80%, respectively. Below T_m, the β-hairpin free energy profile has two minima corresponding to the folded and unfolded states which are separated by a barrier at about 30% of the native contacts formed.

Folding and unfolding rates as a function of temperature are obtained from MD runs started from random conformations and the folded state, respectively. For both peptides the unfolding rate shows an Arrhenius behavior, whereas for $T > 1.1T_m$ the folding rate decreases in agreement with experimental data on peptides and small globular proteins. At 300 K, the α-helix folds in about 3 ns and the β-hairpin in about 96 ns.

2. Model and Methods

2.1. Solvation Model. An implicit solvation model was used in conjunction with the CHARMM force field[23] and a united-atom description (PARAM19[24]) of the peptides to overcome the problem of the very demanding computational time of

* Corresponding author. Email: caflisch@bioc.unizh.ch.
† Current addresss is: GMD-SCAI, Schloss Birlinghoven, D-53754 St. Augustin, Germany.

10.1021/jp994157t CCC: $19.00
Published on Web 05/02/2000

explicit water simulations. The CHARMM PARAM19 default cutoffs for long-range interactions were used; i.e., a shift function[23] 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. Solvation effects were approximated as follows. The ionizable amino acids were neutralized.[25,26] Since a cutoff is used, there is no significant difference between a linear distance-dependent dielectric function and a more sophisticated one, such as a sigmoidal function,[27,28] because the deviation from linearity is small for distances smaller than 10 Å.[29] Hence, a linear distance-dependent dielectric function, $\epsilon(r) = 2r$, was used to approximate the screening effects of the electrostatic interactions. Furthermore, a solvent-accessible surface (SAS) term[30] was added to the CHARMM force field. In this frame, the mean solvation term is given by

$$V_{\text{solv}}(\mathbf{r}) = \sum_{i=1}^{M} \sigma_i A_i(\mathbf{r}) \tag{1}$$

for a protein having M atoms with Cartesian coordinates $\mathbf{r} = (\mathbf{r}_1, ..., \mathbf{r}_M)$. $A_i(\mathbf{r})$ is the solvent-accessible surface area of atom i, computed by an approximate analytical expression[31] and using a 1.4 Å probe radius. Previous studies have shown that the SAS model can be used in MD simulations of different proteins to avoid the main difficulties which arise in in vacuo simulations.[32] The model contains only two σ parameters: one for carbon and sulfur atoms ($\sigma_{C,S} > 0$) and one for nitrogen and oxygen atoms ($\sigma_{N,O} < 0$). They were determined by performing MD simulations on six proteins: crambin (1crn, 46 residues), trypsin inhibitor (1bpi, 58 residues), CI2 (2ci2, 64 residues), ubiquitin (1ubq, 76 residues), SH3 domain of the p85α subunit of bovine phosphatidylinositol 3-kinase (1pht, 83 residues), and histidine-containing phosphocarrier protein (1hdn, 85 residues). Optimization of the two σ_i yielded the same values as in a previous work[32] ($\sigma_{C,S} = 0.012$ kcal/(mol Å^2), $\sigma_{N,O} = -0.060$ kcal/(mol Å^2), and $\sigma_H = 0$). With these parameters, in a 1 ns MD simulation at 300 K, the root-mean-square deviation (rmsd) from the native structure for the C_α atoms averaged over the last 0.5 ns was 1.5 Å (1crn), 1.9 Å (1bpi), 1.9 Å (2ci2), 1.8 Å (1ubq), 1.6 Å (1pht), and 2.5 Å (1hdn). Moreover, the force field and solvation model were used to successfully simulate the reversible folding to the native state of two synthetic 20-residue peptides (Ferrara and Caflisch, manuscript in preparation). These are a peptide with Gly-Ser at the two turns[33] and a peptide with D-Pro-Gly at the two turns,[34] which were shown by NMR to adopt a monomeric three-stranded antiparallel β-sheet in aqueous solution.

2.2. Simulation Method. Constant temperature MD simulations were performed with a coupling constant of 5 ps.[35] The nonbonded interactions were updated every 10 steps. The SHAKE algorithm was used to fix the length of the covalent bonds where hydrogen atoms are involved, which allows an integration time step of 2 fs. Structures were saved every 10 ps.

The helical conformation of the synthetic peptide Ace-$(AAQAA)_3$-$NHCH_3$ (peptide A) (Ace:-$COCH_3$) was generated with backbone dihedral values of $\varphi = -57°$, $\psi = -47°$,[36] $\omega = 180°$, and $\chi_1 = \chi_2 = 180°$ and $\chi_3 = -90°$ for the Gln side chains. Two hundred steps of conjugate gradient minimization were then performed, and the final structure was taken as the folded state of peptide A (Figure 1c). As the most populated conformation of the valine-based peptide Ace-$V_5{}^DPGV_5$-NH_2 (peptide B) is not known in detail, MD simulations were used

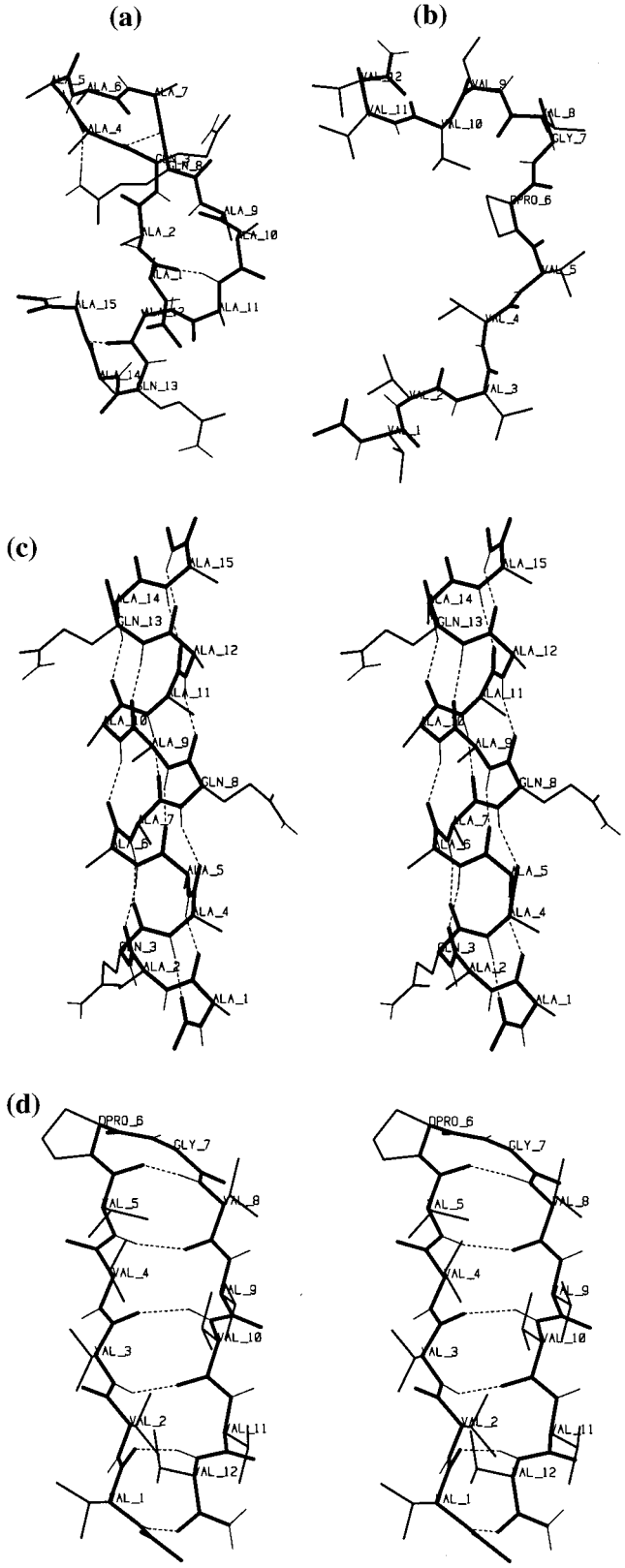

Figure 1. (a) An initial conformation used for the R runs of Ace-$(AAQAA)_3$-NHMe. (b) An initial conformation used for the R runs of Ace-$V_5{}^DPGV_5$-NH_2. (c) Stereo picture (relaxed eyes) of the folded state of Ace-$(AAQAA)_3$-NHMe. (d) Stereo picture (relaxed eyes) of the folded state of Ace-$V_5{}^DPGV_5$-NH_2 obtained after 50 ns MD at 300 K starting from a completely extended conformation. The backbone atoms and the carbonyl oxygens are shown in bold, the side chains in medium lines, and the hydrogen atoms in thin lines. Dashed lines represent hydrogen bonds.

to determine it. First, an extended conformation was generated with $\varphi = \psi = \omega = 180°$. This conformation of peptide B was submitted to a 50 ns MD simulation at 330 K (see Figure 3c). After about 7 ns it folded into a β-hairpin with a two-residue loop at ^{D}PG and was stable for nearly 12 ns. Many interconversions between type I′ and type II′ β-turn[37] were observed at the ^{D}PG site (see below). After it was unfolded for ~11 ns, it refolded and was stable until the end of the simulation, except between 34 and 37 ns. Transient unfolding events were also observed between 10 and 18 ns and 38 and 50 ns. The conformation obtained after 50 ns was submitted to 200 steps of conjugate gradient minimization, and the final structure, which has a type II′ turn, is considered as the folded state of peptide B (Figure 1d).

To investigate both the thermodynamics and kinetics of peptide folding, three sets of simulations were performed (Table 1). The first set (N) was started from the folded conformation. For the second set of simulations (R), 200 conformations were generated by randomizing the dihedral angles of the rotatable bonds, followed by 1000 steps of energy minimization. The 50 structures with the most favorable effective energy (sum of intrapeptide plus solvation eq 1) were retained as starting conformations. For peptide A (B), their average root-mean-square deviation of the C_α atoms (C_α-rmsd) from the folded state is 5.4 Å (5.1 Å) with a minimum value of 3.9 Å (3.1 Å) and a maximum of 7.6 Å (7.9 Å). Two random conformations are shown in Figure 1a,b. The R runs were stopped when the C_α-rmsd from the folded state reached a value smaller than 1.0 Å. This criterion, which is more stringent than the 1.5 Å cutoff used to estimate the population of folded state and in the cluster analysis (see below), was chosen to guarantee that the folded state, rather than a transient folded-like conformation, was reached. The mean folding time was obtained by averaging the simulation times for a given temperature, and the folding rate is defined as the inverse of the mean folding time.

The third set of simulations (U) was initiated from the folded conformation with different initial velocity values and stopped when the C_α-rmsd from the folded state was larger than 4.4 Å (4.1 Å) for peptide A (B), which corresponds to the average C_α-rmsd from the folded state of the random structures for the R simulations minus 1 Å. This criterion was chosen to be consistent with the R runs where the simulations were stopped when, as mentioned above, a C_α-rmsd from the folded state of less than 1 Å was reached. The average stopping time of the U runs yielded the mean unfolding time, whose inverse defines the unfolding rate. The total simulation time is 1.42 μs for peptide A (0.18, 1.06, and 0.18 μs for the N, R, and U runs, respectively) and 2.64 μs for peptide B (0.54, 1.94, and 0.16 μs for the N, R, and U runs, respectively).

2.3. Analysis Methods. For a system at equilibrium the free energy of a state with a reaction coordinate q is given by[38]

$$\Delta G(q) = -k_B T \ln p(q) \tag{2}$$

where k_B is the Boltzmann constant, T is the temperature, and $p(q)$ is the probability of finding a conformation whose reaction coordinate is q. If a sufficient number of transitions between states with different q are observed in the simulation, $p(q)$ can be taken as the number of conformations in state q divided by the total number of conformations.

The Lifson−Roig (LR) theory is used here to estimate the helical nucleation and propagation parameters.[39] It classifies the states of a residue according to its (φ, ψ) backbone angles, and a residue is considered helical if its (φ, ψ) values are within 30° from the ideal values of (−57°, −47°).[36] Each residue i in the helical conformation is assigned the statistical weight w_i', except for the two end residues of a helical segment which have a weight v_i'. The statistical weight of the nonhelical (coil) residues is u_i' (the subscript i is neglected henceforth). The LR theory defines the reference state as the coil state with a weight of $1 = u'/u'$. The nucleation $v^2 = (v'/u')^2$ and propagation $w = w'/u'$ parameters of the LR theory are defined as in ref 39. The free energies of helix nucleation and propagation are then computed from these two parameters:[39]

$$\Delta G_{nucl} = -k_B T \ln \frac{w v^2}{(1+v)^5} \tag{3}$$

$$\Delta G_{prop} = -k_B T \ln \frac{w}{(1+v)} \tag{4}$$

The cluster analysis is performed in the same way as in ref 8. The C_α-rmsd is calculated for each pair of structures after optimal superposition. The number of neighbors is then computed for each structure using a C_α-rmsd cutoff of 1.5 Å. The conformation with the highest number of neighbors is considered as the center of the first cluster. All the neighbors of this conformation are removed from the ensemble of conformations. The center of the second cluster is then determined in the same way as for the first cluster, and this procedure is repeated until each structure is assigned to a cluster. As pointed out by Daura et al., this way of clustering emphasizes the most populated cluster and can result in many clusters with a single member.[8]

To assess the robustness of the clustering, tests with a C_α-rmsd cutoff of 1.3 and 1.7 Å were performed. All the conclusions on thermodynamics and kinetics, which were drawn from the cluster analysis (see sections 3.1.3 and 3.2.4), are the same for the three cutoffs tested.

TABLE 1: Simulations Performed

	Ace-(AAQAA)$_3$-NHCH$_3$			Ace-V$_5$^{D}PGV$_5$-NH$_2$		
	temp range (K)	no. of simulns	length (ns)	temp range (K)	no. of simulns	length (ns)
N[a]	270−420	6	30	300−450	6	90
R	270−420	50 × 6	0.15−100	330−510[b]	50 × 7	0.03−100
U	330−420	20 × 4	0.03−18	360−510[c]	20 × 6	0.03−47

[a] Simulation type (see Methods section for the explanation). [b] The high temperature values were chosen to show the non-Arrhenius behavior of the folding rate. [c] Temperature values lower than 360 K were not used because they would be too CPU-time expensive.

3. Results

3.1. Equilibrium simulations (N). *3.1.1. Overall behavior.* The residue helicity for peptide A was averaged over the whole N270 simulation. A residue is considered helical if it belongs to a segment of at least three residues whose dihedral angles (φ, ψ) are within 30° from the ideal values (−57°, −47°).[17] The agreement with the corresponding data obtained by NMR chemical shift measurements at 274 K[16] is good. For the percentage of residue helicity, a correlation coefficient of $r = 0.92$ is found between the experimental values (Table 3 of ref 3) and those obtained from the equilibrium simulations (Figure 2). Both simulation and NMR data indicate that peptide A is more helical at the N-terminus than at the C-terminus. Despite the good correlation, the helicity values are smaller in the experiment than in the simulation (44% on average versus 75%), but a direct comparison is not obvious because the definition of the residue helicity is somewhat arbitrary. Yet, an overesti-

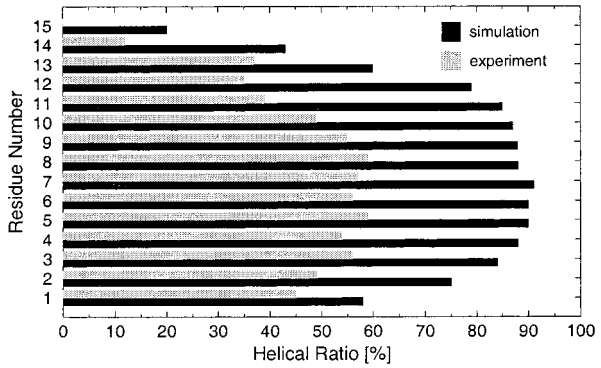

Figure 2. Simulated and experimental residue helicity of Ace-$(AAQAA)_3$-NHMe. The helical content was averaged over the whole simulation at 270 K. A residue is considered helical if it belongs to a segment of at least three residues whose dihedral angles (φ, ψ) are within 30° from the ideal values (−57°, −47°).

mation of stability due to the force field and solvation model cannot be ruled out.

To check the convergence of the N simulations, the helical ratio per residue was computed by considering different time intervals of the simulation, i.e., 0−10, 10−20, and 20−30 ns. The variance in the residue helicity values was 2.4% on average, which indicates that thermodynamic equilibrium has been reached. This is confirmed by Figure 3a,b,e, which shows the evolution of the C_α-rmsd from the folded conformation in the N270 and N360 simulations of peptide A. Many fast folding/unfolding transitions are observed. The residue helicity analysis shows that these structures are characterized by frayed ends.

The evolution of the C_α-rmsd in the N360 simulation of peptide B is depicted in Figure 3d,f. In the first 30 ns the β-hairpin unfolded transiently about 20 times for nearly 80 ps on average and then completely unfolded for the next 20 ns. The peptide refolded and unfolded several times in the last 40 ns. Overall, peptide B was folded nearly half of the time in the N360 simulation, which implies that the melting temperature is close to 360 K (Figure 4). Only type I′ and II′ β-turns were observed in the N simulations of peptide B. This is consistent with the experimental finding that types I′ and II′ are predominant in two-residue-loop β-hairpins.[40] The mean values of the φ and ψ angles of the two residues involved in the turn were computed at 300 K. For type I′ one obtains (φ_1, ψ_1) = (58.5, 24.7)° ± (10.5, 18.4)° and (φ_2, ψ_2) = (108.4, −30.4)° ± (28.2, 19.5)°, while for type II′ (φ_1, ψ_1) = (64.4, −92.6)° ± (9.1, 16.6)° and (φ_2, ψ_2) = (−121.7, −3.3)° ± (35.7, 24.2)°. It is worth noting that the mean values can differ by as much as 40° from the ideal values given by Wilmot and Thornton,[37] perhaps because of the D-Pro at the turn. At 300 K, type I′ was found to be as populated as type II′. It might be a peculiarity of D-Pro-Gly to be able to accommodate both types of turn.

3.1.2. Energetics. The population of the folded state in the equilibrium simulations N270−N450 is shown in Figure 4. Both peptides show a continuous loss of structure over the temperature range studied, instead of a sharp transition. This indicates that their folding is weakly cooperative or noncooperative in agreement with experimental data on a 21-residue alanine-based peptide.[41] Furthermore, it was shown for apomyoglobin that the formation of tertiary structure proceeds through a cooperative transition, whereas helix formation does not.[42] A gradual decrease of the β-hairpin population with temperature has been observed experimentally in the thermal unfolding of the C-terminal fragment (residues 41−56) of protein G B1[10,43] and in simulations of the synthetic β-hairpin BH8.[13] Abkevich et al. have reported that on a cubic lattice the folding transition of a model 36-mer, whose native state consists mostly of local contacts, is noncooperative.[44] For peptides A and B the melting temperature in the force field is close to 320 and 370 K, respectively.

Equations 3 and 4 and the N simulations of peptide A were used to calculate the free energy of helix nucleation and propagation. At 270 K, $\Delta G_{nucl} = 1.34$ kcal/mol and $\Delta G_{prop} = -0.16$ kcal/mol. Both free energy values increase with temperature, and at 420 K one has $\Delta G_{nucl} = 4.63$ kcal/mol and $\Delta G_{prop} = 1.83$ kcal/mol. The experimental determination of the Lifson−Roig parameters in alanine-based peptides yielded $\Delta G_{nucl} = 3.16$ kcal/mol and $\Delta G_{prop} = -0.26$ kcal/mol at 273 K.[45] As our force field underestimates the free energy of nucleation, the R simulations should give a lower bound for the mean folding time of peptide A.

The effective energy (intraprotein energy plus mean solvation energy) and the free energy (computed using eq 2) as a function of the fraction of folded state contacts Q are shown in Figure 5 averaged over the N runs at different temperatures. The contacts in the folded state of peptide A consist of 13 hydrogen bonds between residue i and $i + 4$, while the 14 contacts used to define the folded conformation of peptide B are listed in Table 2. In the temperature range $T = 270-330$ K for peptide A and $T = 330-420$ K for peptide B, there is a rather flat behavior of the average effective energy for $Q < 0.5$. At high temperature or close to the folded state the effective energy decreases smoothly toward the energy value of the native state.

To better investigate the origin of the relatively large negative slope, at high temperatures, of the effective energy as a function of Q, finite-difference Poisson−Boltzmann (PB) calculations were performed with the program UHBD[46,47] on nearly 1000 conformations of peptide A taken from the R420 runs, whose effective energy as a function of Q is essentially identical to the one of the N420 run. A correlation of $r = 0.92$ (slope $s = 1.00$) was found between the effective energies obtained by the SAS (eq 1) and the PB solvation models, which is remarkable because the PB model accounts only for the electrostatic contribution of solvation. Interestingly, for the same set of conformations the correlation between the energies in vacuo and the effective energies with electrostatic solvation calculated by PB is 0.86 ($s = 0.46$), while the average PB and SAS solvation energies correlate with $r = 0.92$ and $s = 0.80$. To assess the importance of solvation, five simulations at 300 K and five at 420 K were started from random conformations of peptide A with the CHARMM PARAM19 vacuum energy function. None of them reached the folded state in 50 ns, which indicates that the average effective energy in vacuo is not funnel-like. This points out that the funnel-like shape of the effective energy surface stems from a compensation between the intrapeptide energy and the free energy of solvation, whose electrostatic contributions are known to be anticorrelated.[48,49] To check whether the folded state was not reached because of a high energy, two vacuum simulations at 300 and 420 K were initiated from the folded state of peptide A and run for 50 ns. The average C_α-rmsd from the folded structure was less than 1 Å, and the effective energies were more favorable than the ones sampled during the runs started from random conformations by on average 20 and 15 kcal/mol at 300 and 420 K, respectively.

The overall similarity of the free energy profiles across the temperature range investigated indicates that equilibrium has been reached (Figure 5c,d). The free energy surface of peptide A has a single minimum whose position depends on the temperature; i.e., states with higher Q values (more helical) are predominantly populated at lower temperatures. Figure 5c shows

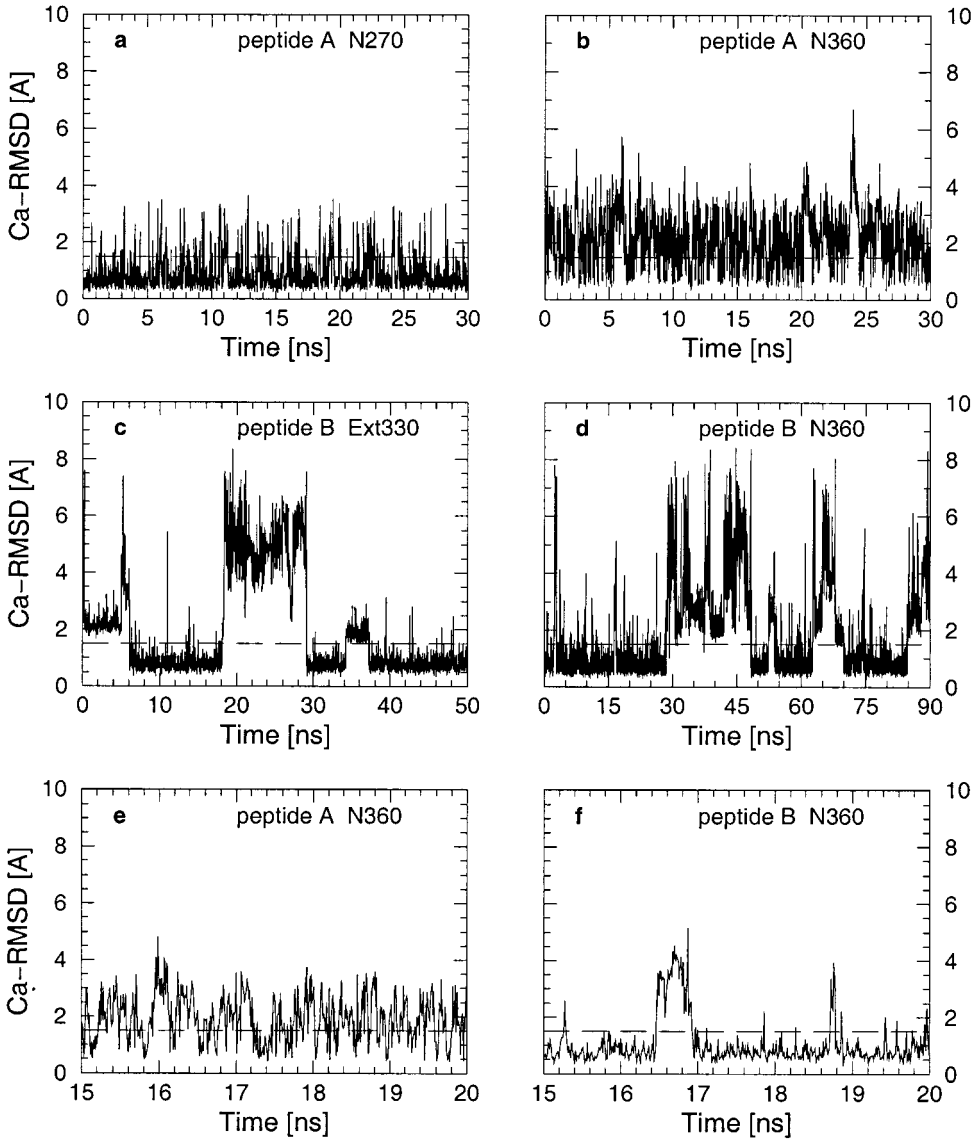

Figure 3. Evolution of the rmsd (Å) from the folded state for the C_α atoms as a function of time. (a) N270, (b) and (e) N360 simulations of Ace-$(AAQAA)_3$-$NHCH_3$ started from the α-helical structure. (c) Simulation at 330 K of Ace-$V_5{}^DPGV_5$-NH_2 started from the extended conformation, (d) and (f) N360 simulation of Ace-$V_5{}^DPGV_5$-NH_2 started from the β-hairpin structure. Conformations with a rmsd smaller than 1.5 Å (broken line) correspond to either the α-helical fold (a, b, e) or the β-hairpin fold (c, d, f).

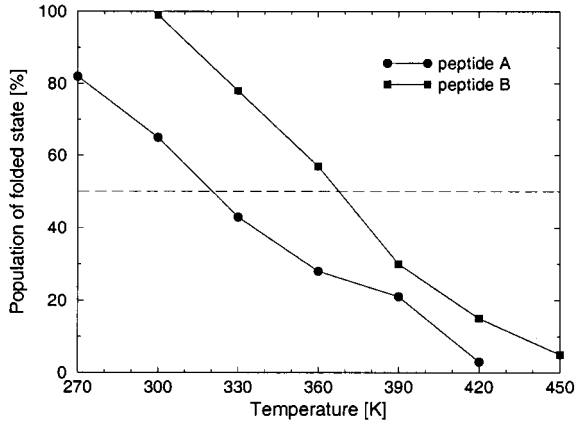

Figure 4. Population of the folded state during the N simulations of Ace-$(AAQAA)_3$-$NHCH_3$ (circles) and Ace-$V_5{}^DPGV_5$-NH_2 (squares) as a function of temperature. A structure is considered folded if its C_α-rmsd from the folded state is smaller than 1.5 Å.

that the helix−coil transition is not a first order (two-state) but a higher order phase transition. In this context, it has been reported that the enthalpies of folding of a 50-residue alanine-based peptide obtained by calorimetric measurement and using the two-state assumption differ significantly.[50] A much better agreement between theoretical and experimental enthalpies of the helix−coil transition was obtained using statistical mechanics models of α-helix formation.[39,51]

For peptide B two different behaviors are observed. Above the melting temperature (T = 370 K), the free energy has a pronounced minimum at $Q = 0$ and is nearly flat for values of Q between 0.2 and 0.7. At 330 and 360 K there are two minima separated by a barrier which locates the transition state at $Q \approx 0.3$. This agrees well with a statistical model for β-hairpin formation.[43] The contacts formed in the transition state lie mainly close to the turn. At 330 K, the hydrogen bond at the turn (i.e., contact 1) is present in 59% of the conformations with four contacts (i.e., $Q \approx 0.3$), and contacts 2 and 3 are present in the $Q \approx 0.3$ structures with a frequency of 78% and 64%, respectively. At 360 K, these percentages are 69%, 78%, and 54% for contacts 1−3. As a basis of comparison, among the contacts which involve residues separated by at least four amino acids (i.e., contacts 4−14), the most frequent one in the

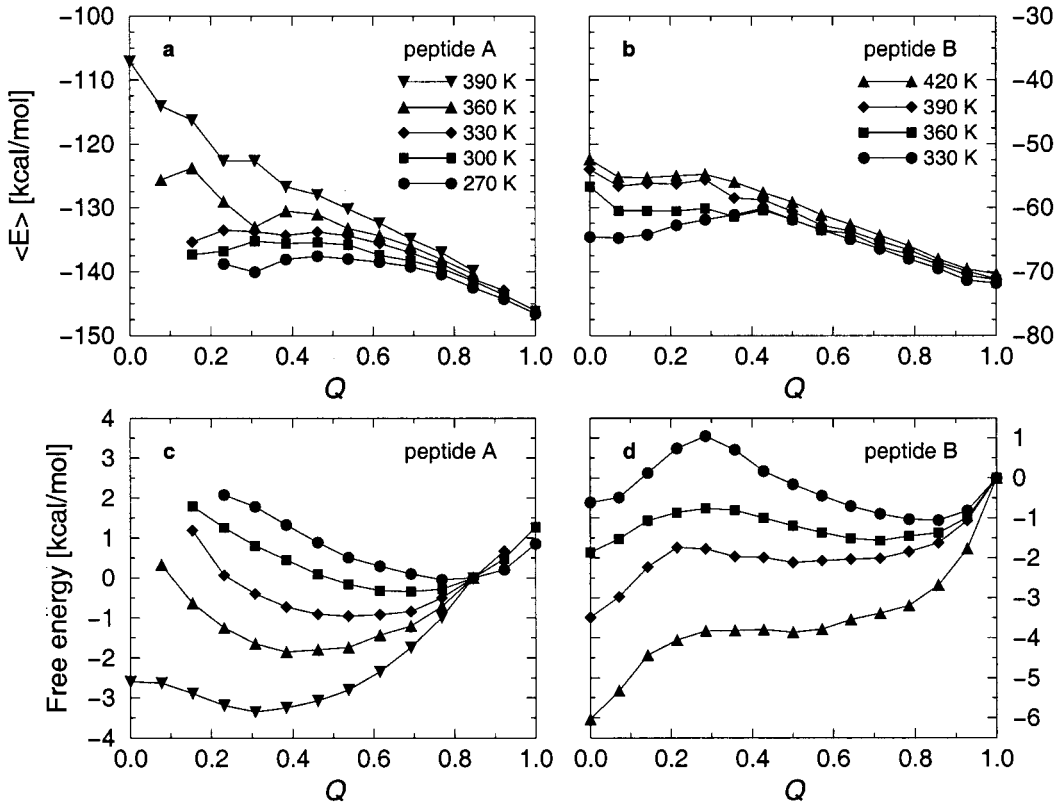

Figure 5. Average effective energy $\langle E \rangle$, and free energy of Ace-$(AAQAA)_3$-$NHCH_3$ (a, c) and Ace-$V_5{}^DPGV_5$-NH_2 (b, d) as a function of the fraction of folded state contacts (Q) present during the N runs at different temperatures. One thousand and three thousand conformations were selected from the N simulations of peptide A and B, respectively, at each temperature to compute $\langle E \rangle$. They were submitted to a 10 ps MD run at 300 K, followed by 300 steps of energy minimization, before evaluating the values of $\langle E \rangle$ and Q. The standard deviations of $\langle E \rangle$ averaged over all temperatures and all Q values are ± 4.5 kcal/mol for peptide A and ± 3.9 kcal/mol for peptide B. The average effective energy and the free energy at 300 K of peptide B, as well as a few data points for peptide A, are not shown because of insufficient statistics. The free energy was arbitrarily set to zero at $Q = 0.85 = 11/13$ for peptide A (due to the insufficient statistics at $Q = 1$ for N330, N360, and N390) and at $Q = 1$ for peptide B.

TABLE 2: List of Contacts in the β-Haripin Conformation of Ace-$V_5{}^DPGV_5$-NH_2

contact no.[a]	atom 1	atom 2	distance[b] (Å)
1	5O	8H	2.00
2	$5C_\beta$	$8C_\beta$	5.09
3	5H	8O	2.06
4	$4C_{\gamma 1}$	$9C_{\gamma 2}$	4.35
5	$4C_{\gamma 2}$	$9C_{\gamma 1}$	4.36
6	3O	10H	2.09
7	$3C_\beta$	$10_{C\beta}$	4.63
8	3H	10O	2.01
9	$2C_{\gamma 1}$	$11C_{\gamma 2}$	4.51
10	$2C_{\gamma 2}$	$11C_{\gamma 1}$	4.31
11	1O	12H	1.95
12	$1C_\beta$	$12C_\beta$	4.06
13	1H	12O	1.99
14	$1CH_3$[e]	12NT[d]	4.78

[a] Contacts are sorted from local to distant in sequence. [b] Distance in the folded state. [c] Carbon atom of CH_3 at N-terminal Ace. [d] Nitrogen atom of NH_2 at C-terminus.

$Q \approx 0.3$ conformations is contact 4 (50% in N330) and contact 5 (34% in N360). This shows, at least for the sequence of peptide B, that the first event in β-hairpin formation is the acquisition of the contacts close to the turn.

3.1.3. Cluster Analysis. Three thousand snapshots were selected from each of the N simulations of peptide A (every 10 ps) and peptide B (every 30 ps) to perform a cluster analysis.

Peptide A. A total of 24 clusters (15 of which with more than one member) at 270 K and 2155 clusters (170 with more than one member) at 420 K were found. The center of the most populated cluster corresponds to the α-helical fold at all temperatures (C_α-rmsd from the folded state less than 1.1 Å). However, the first cluster incorporates 89% of all the conformations at 270 K and only 4% at 420 K. A nonnegligible presence of π-helical hydrogen bonds is observed in the clusters with a minimum of 20 structures, especially at the C-terminus of the peptide. This result is surprising because π-helices are very rare in proteins. The central conformation of the second most populated cluster is basically the same over the temperature range investigated, the largest C_α-rmsd being 1.4 Å between 300 and 360 K. The center of the second cluster at 270 K has six π-helical hydrogen bonds. The amide hydrogen donors are Ala10, Ala11, Ala12, Gln13, Ala15, and the C-terminal $NHCH_3$. The present simulation results are consistent with a study by Shirley and Brooks,[18] who have performed MD simulations at 298 K of peptide A with explicit water molecules and periodic boundary conditions. They concluded that the π-helix is mainly stabilized by interactions between the side chains of Gln8 and Gln13. In the central conformation of the second cluster at 270 K, the distances between the oxygen of the side chain of Gln8 and the amide hydrogens of Gln13 are 4.8 and 5.0 Å, whereas these distances are 12.8 and 11.9 Å in the α-helical structure. Distances of about 5 Å were also observed between the oxygen of Gln13 and the amide hydrogens of Gln8, but less frequently. Very few of these distances were smaller than 2.6 Å and only for less than 0.5% of the time. Hence, the side chains of Gln8 and Gln13 do not form stable hydrogen bonds in the N270−N420 simulations, in agreement with the explicit water simulation results of Shirley and Brooks,[18] who observed an interaction between the amide hydrogen of Gln8 and the oxygen of Gln13 through a bridging water molecule. The percentage of π-helicity

calculated over the whole simulation at 300 K was around 7%; as a basis of comparison Shirley and Brooks obtained 10%.

It has been proposed in the past that 3_{10}-helices may play the role of intermediates in the folding and unfolding of α-helices.[52−54] 3_{10}-Helical hydrogen bonds were present 6.5% of the time in the simulation at 300 K with the highest percentage for Gln3 and Ala4. Moreover, 3_{10}-type hydrogen bonds were involved in the breaking and reformation of α-helical hydrogen bonds. On average, a 3_{10}-helical hydrogen bond was observed nearly 20% of the time just after (before) an α-helical hydrogen bond broke (re-formed). Hence, these results agree with previous simulations reviewed in [5]. *i,i*-3 hydrogen bonds were more present at the termini than in the middle of the peptide. None of the representatives of the clusters with a minimum of 20 member structures has more than 30% of 3_{10}-helical hydrogen bonds, which indicates that the 3_{10}-helix as a whole is not a minimum on the free energy surface in the present model. This agrees with previous work by Young and Brooks, who performed free energy calculations of Ace-$(Ala)_n$-$NHCH_3$, $n = (4,5,10,15)$, with explicit water molecules.[55] For helices of polyalanine Tirado-Rives et al. have reported that the free energy of the 3_{10} state is higher than the one of the α state by about 1.0 kcal/mol per residue.[56]

Peptide B. A total of 4 clusters (1 of which with more than one member) at 300 K and 1788 clusters (358 with more than one member) at 450 K were obtained. The central conformation of the most populated cluster corresponds to the β-hairpin at all temperatures (C_α-rmsd from the folded state less than 1.1 Å). The first cluster incorporates 99% of all the structures at 300 K and 7% at 450 K. This is an a posteriori justification for using a temperature range of 300−450 K for the N runs of peptide B instead of 270−420 K, which was used for peptide A. There are some correspondences between clusters obtained at different temperatures. For example, the representative of the third most populated cluster at 330 K corresponds to the center of the fourth cluster at 360 K, the second cluster at 390 K, and the eighth cluster at 420 K. Partially formed β-hairpins as well as out-of-register β-hairpins were often observed in the most populated clusters. At 330 K, for instance, the representative of cluster 2 is characterized by a type II β-turn at the V^DP site, the largest deviation of the φ and ψ angles from the ideal values being 57° for ψ_2. Furthermore, the center of cluster 4 is close to the center of cluster 2 (C_α-rmsd of 1.6 Å), both having a V^DP type II turn. At the same temperature the contacts 2−7 are present in the representative of the third cluster.

3.2. R and U Simulations. *3.2.1. Non-Arrhenius Behavior of the Folding Rate*. All 650 trajectories started from random conformations (R) reached the folded state in less than 100 ns, except one for peptide A at 270 K and three for peptide B at 330 K. A somewhat arbitrary folding time of 100 ns was assigned in these four cases. Table 3 lists the mean folding times while a plot of the logarithm of the folding and unfolding rates as a function of $1/T$ is shown in Figure 6. The folding rate shows Arrhenius-like behavior at low temperature ($T < 360$ K ≈ 1.1 T_m for peptide A, $T < 420$ K ≈ 1.1 T_m for peptide B) and a non-Arrhenius behavior at high temperature. The folding rate starts to decrease only above the melting temperature in our simulations, whereas it occurs at lower temperatures experimentally. This suggests that the entropic penalty is dominant only at very high temperatures, which maybe due to the small size of the two systems. Nevertheless, it cannot be ruled out that the approximations inherent to the force field and solvation model are at the origin of the very high temperature value (360 and 420 K for peptides A and B, respectively) for the maximum of the folding rate.

TABLE 3: Mean Folding Time of Ace-$(AAQAA)_3$-$NHCH_3$ and Ace-$V_5^DPGV_5$-NH_2 at Different Temperatures

temperature	Ace-$(AAQAA)_3$-$NHCH_3$			Ace-$V_5^DPGV_5$-NH_2		
(K)	τ_f [a] (ns)	τ_f^{fit} [b] (ns)	r [c]	τ_f [a] (ns)	τ_f^{fit} [b] (ns)	r [c]
270	10.07	6.08	0.97			
300	3.41	2.28	0.98			
330	1.45	1.18	0.99	21.05	14.89	0.97
360	0.95	1.12	0.94	5.96	5.35	0.98
390	1.42	1.67	0.96	2.08	2.02	1.00
420	3.87	3.70	0.99	1.53	1.59	0.99
450				1.59	1.54	0.99
480				2.15	2.32	0.99
510				4.34	3.19	0.99

[a] The mean folding time of the R simulations is calculated by averaging the simulation times for a given temperature. [b] Exponential fit ($P_u(t) = \exp(-t/\tau_f^{fit})$) to the evolution of the unfolded population as a function of time t. [c] Correlation between the exponential fit and the evolution of the unfolded population.

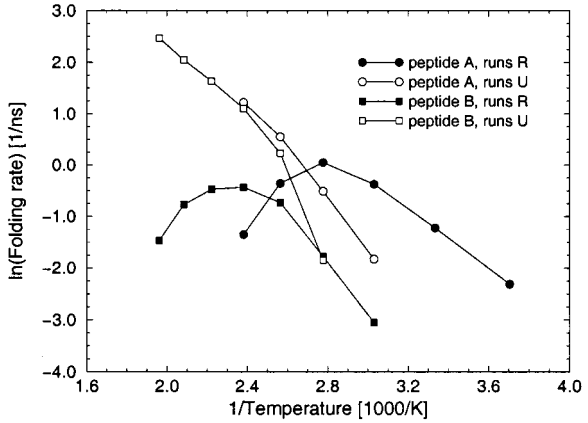

Figure 6. Arrhenius plot of the folding (filled symbols, R simulations) and unfolding (empty symbols, U simulations) rates of Ace-$(AAQAA)_3$-$NHCH_3$ (circles) and Ace-$V_5^DPGV_5$-NH_2 (squares).

The Arrhenius plot allows the determination of the enthalpic ($\Delta H_f^‡$) and entropic contribution ($-T\Delta S_f^‡$) to the free energy of activation for folding.[57] Pairs of consecutive data points were used to determine $\Delta H_f^‡$ and $-T\Delta S_f^‡$ in the middle of the segment joining two consecutive points. For peptide A one obtains $\Delta H_f^‡ = 5.8$ kcal/mol and $-T\Delta S_f^‡ = -4.9$ kcal/mol at 285 K, and $\Delta H_f^‡ = -10.9$ kcal/mol and $-T\Delta S_f^‡ = 11.5$ kcal/mol at 405 K. For peptide B these values are $\Delta H_f^‡ = 9.9$ kcal/mol and $-T\Delta S_f^‡ = -8.3$ kcal/mol at 345 K and $\Delta H_f^‡ = -11.4$ kcal/mol and $-T\Delta S_f^‡ = 12.4$ kcal/mol at 495 K. Hence, the free energy of activation is 0.9 and 0.6 kcal/mol for peptide A at 285 and 405 K, respectively. For peptide B it is 1.7 and 1.0 kcal/mol at 345 and 495 K, respectively. At low temperatures the activation enthalpy is positive, which implies that the barrier for folding is mainly enthalpic. As the temperature increases, the activation energy becomes negative, but as the activation entropy increases, the free energy of activation is still positive. The small values of the free energies of activation are probably related to the small size of the two peptides.

The unfolding rate shows an Arrhenius behavior in the temperature range investigated, which implies that the energy barrier for escaping from the folded state is always enthalpic. For peptide A a linear regression with the four data points of the unfolding rate yields $\Delta H_u^‡ = 9.4$ kcal/mol with a correlation coefficient r of 0.997. For peptide B a linear regression with the six data points yields $\Delta H_u^‡ = 9.9$ kcal/mol with $r = 0.965$. The curves of the folding and unfolding rates cross at 370 K

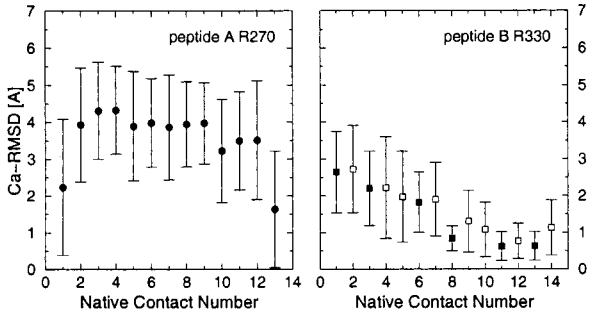

Figure 7. Average value of the C_α-rmsd from the folded state at the last disappearance of the α-helical contacts of Ace-$(AAQAA)_3$-$NHCH_3$ (left panel) and *β*-hairpin contacts of Ace-$V_5^DPGV_5$-NH_2 (right panel) in the R simulations. For Ace-$(AAQAA)_3$-$NHCH_3$ the α-helical contacts correspond to the 13 hydrogen bonds present in the α-helix and are given from the N-terminus to the C-terminus. For Ace-$V_5^DPGV_5$-NH_2 the numbering of the *β*-hairpin contacts corresponds to the one given in Table 2, and filled and empty square symbols represent backbone hydrogen bonds and van der Waals contacts, respectively. The last disappearance of a contact is defined as the last interval of at least 50 ps during which the contact is not present. A contact is said to be present if the distance between the two atoms defining the contact is less than that in the folded state times 1.3. Bars = 2 standard deviations.

for peptide A and 360 K for peptide B, while the melting temperature, at which the two rates should cross, are 320 K for peptide A and 370 K for peptide B. The agreement is good for peptide B, while the discrepancy for peptide A is probably due to the criterion chosen for stopping the U simulations.

3.2.2. Evolution of the Unfolded Population. The fraction of unfolded conformations at time *t*, $P_u(t)$, can be derived from the percentage of trajectories that have not reached the folded state at time *t*. Exponential curves were fitted to $P_u(t)$, and the characteristic times τ_f^{fit} are listed in Table 3. No significant deviation from the simple exponential law is observed. For both peptides the largest discrepancy between τ_f (the mean folding time calculated by averaging the R simulation times) and τ_f^{fit} (the characteristic time from the exponential fit) occurs at the lowest temperature. A monoexponential fit is preferred because the statistics is not sufficient to warrant a multiexponential description of the decay of the unfolded population without the risk of overfitting. Removing the data points with insufficient statistics, i.e., the three points for peptide A and the three for peptide B with $P_u(t) < 0.05$, yields a τ_f of 6.30 ns instead of 10.07 ns for peptide A, and a τ_f of 16.01 ns instead of 21.05 ns for peptide B, whereas the values of τ_f^{fit} are unchanged within 0.01 ns. Therefore the "noise" at low $P_u(t)$ explains most of the deviation from the exponential law. The single-exponential decay of the unfolded population is expected for peptide B at 330 and 360 K, due to the shape of the free energy landscape (two minima). In the case that the kinetics do not follow a two-state model, the single-exponential decrease of the unfolded population indicates that there is a fast interconversion between the different subensembles of unfolded states (type 0A scenario for downhill folding in the nomenclature of Bryngelson et al.[58]).

3.2.3. Sequence of Events. The evolution of the contacts as a function of the C_α-rmsd from the folded state is depicted in Figure 7. The results are shown at 270 K for peptide A and at 330 K for peptide B, i.e., at the lowest temperature of the R simulations. A similar behavior is found over the whole temperature range and for the unfolding simulations U. That the sequence of events does not depend either on the temperature or on the direction of the process (folding vs unfolding) was also found in MD simulations of CI2 at 300, 375, and 475 K.[59,60] The large standard deviations indicate that multiple folding routes are possible for both peptides and more so for the α-helix. No preferred pathway emerges from the simulation results of peptide A if one excludes the late formation of the hydrogen bonds at the termini. This might be related to their intrinsic instability. Peptide B initiates folding mainly at the *β*-turn and then progressively zips up. No major differences are seen between hydrogen bonds and interstrand side-chain contacts involving the same pair of residues, suggesting that they are concomitant in *β*-hairpin formation. These results have been found also with the help of a statistical model developed to describe the thermodynamics and kinetics of a *β*-hairpin (with turn at residues Ala48-Thr49) extracted from protein G B1.[10,43] However, two recent simulation studies have pointed out that the first folding event of the *β*-hairpin from protein G B1 is the formation of contacts involved in the hydrophobic cluster made up of the residues W43, Y45, F52, and V54.[14,15]

TABLE 4: Clusters with a Minimum of 20 Member Structures from R Simulations

	ACE-$(AAQAA)_3$-$NHCH_3$ 270 K			Ace-$V_5^DPGV_5$-NH_2 330 K		
cluster	rmsd[a]	members[b]	*N*[c]	rmsd[a]	members[b]	*N*[c]
1	7.61[d]	271	0	1.92[e]	877	0
2	8.27	270	0	3.21	727	0
3	8.05	176	0	2.02	139	0
4	8.00	88	0	2.66	79	3
5	7.59	63	0	2.16	69	1
6	5.19	61	3	3.47	62	1
7	4.85	59	5	3.00	31	1
8	5.91	43	1	2.87	28	0
9	6.30	42	0	5.83	26	0
10	8.32	41	0	4.02	25	0
11	7.69	41	0	5.87	22	1
12	3.57	40	3	2.40	20	4
13	5.41	33	1			
14	5.60	31	2			
15	6.41	24	0			
16	8.07	22	0			
17	6.97	21	2			
18	4.94	20	3			
19	3.81	20	4			

[a] C_α-rmsd (Å) between the central structure of the cluster and the folded state. [b] Number of members in the cluster. [c] Number of contacts in common with the folded structure. Ace-$(AAQAA)_3$-$NHCH_3$ and Ace-$V_5^DPGV_5$-NH_2 have 13, respectively, 14 contacts, in their folded conformation. At each temperature value the cluster analysis was performed on an ensemble of 3000 conformations chosen among the 50 R trajectories such that the time interval between consecutive snapshots is nearly the same. [d] Conformation shown in Figure 8 (top panel). [e] Conformation shown in Figure 8 (bottom panel).

3.2.4. Cluster Analysis. Peptide A. A cluster analysis on the R trajectories reveals the presence of 19 clusters with a minimum of 20 member structures at 270 K (Table 4), 13 at 300 and 330 K, 10 at 360 K, 6 at 390 K, and 2 at 420 K. They incorporate 46%, 38%, 24%, 15%, 10%, and 2%, respectively, of the 3000 conformations analyzed for each temperature value. There are a total of 880 clusters (242 of which with more than one member) at 270 K and 2681 clusters (108 with more than one member) at 420 K. At 270 K, the fraction of α-helical contacts in the most populated clusters (the ones having at least 20 conformations) is on average close to 10%. This indicates that nucleation is the rate-determining step in helix formation at low temperature, because there are few clusters with a significant percentage of helicity. At 360 K, the structures in the most populated clusters have a *Q* value of 0.35 on average. The higher temperature facilitates nucleation and therefore results in faster folding. Ten, respectively 7, of the 13 α-helical contacts are formed in the two most populated clusters at 420 K. At this temperature, the accessible space becomes so large that an

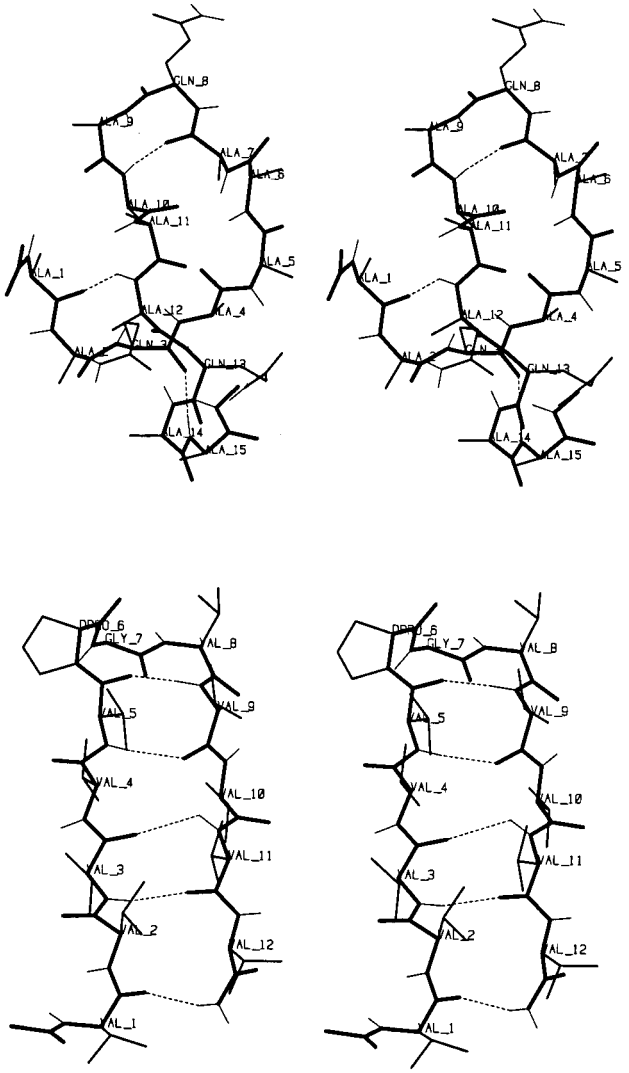

Figure 8. Stereo picture (relaxed eyes) of the center of the most populated cluster of the R270 simulations of Ace-$(AAQAA)_3$-$NHCH_3$ (top panel) and the R330 simulations of Ace-$V_5{}^DPGV_5$-NH_2 (bottom panel). The backbone atoms and the carbonyl oxygens are shown in bold, the side chains in medium lines, and the hydrogen atoms in thin lines. Dashed lines represent hydrogen bonds.

entropy bottleneck slows down the folding rate. The mean nucleation time (τ_{nucl}) can be estimated from the R simulations by averaging the time needed to form the first i, $i + 4$ hydrogen bond over the 50 (49 at 270 K) trajectories. Hence, the mean propagation time (τ_{prop}) is given by $\tau_{prop} = (\tau_f - \tau_{nucl})/(N - 1)$, where τ_f is the mean folding time and N the number of hydrogen bonds in the folded state. With these definitions one obtains $\tau_{nucl} = 2.1$ ns and $\tau_{prop} = 0.7$ ns at 270 K, $\tau_{nucl} = 0.21$ ns and $\tau_{prop} = 0.06$ ns at 360 K, and $\tau_{nucl} = 0.18$ ns and $\tau_{prop} = 0.31$ ns at 420 K. Propagation is faster than nucleation at low and medium temperatures, but not at high because of the helix instability which penalizes propagation and the larger accessible space which facilitates nucleation. The highest folding rate (at 360 K, Figure 6) originates from fast nucleation and propagation steps.

Partially formed β-hairpins are found frequently in the clusters of the R simulations of peptide A, mainly below 360 K. For instance, the center of clusters 1 (shown in Figure 8a), 2, 4, and 5 at 270 K have a type II′ β-turn at residues Gln8-Ala9 with four, six, seven, and six backbone hydrogen bonds, respectively. At the turn residues the largest deviation of the φ and ψ angles from the ideal values is 60° for ψ_1 in cluster 4. Furthermore, the center of cluster 3 has a type I β-turn at the Gln8-Ala9 site with six main-chain hydrogen bonds. None of these "β-hairpins" were found in the clusters with a minimum of 20 structures in the N simulations, indicating that they are misfolded structures rather than local minima on the free energy surface (see also ref 13). To verify this conclusion, an additional 50 ns MD simulation at 270 K was carried out from the center of cluster 1 of the R270 trajectories. Its β-hairpin with type II′ turn conformation broke apart at about 10 ns and never re-formed. The α-helical conformation was reached after nearly 15 ns, and an equilibrium behavior similar to the one shown in Figure 3a was then observed until the end of the simulation.

Peptide B. Twelve clusters with a minimum of 20 member structures were found at 330 K (Table 4), 18 at 360 K, 15 at 390 K, 10 at 420 K, 4 at 450 K, and 1 at 480 and 510 K. They incorporate 70%, 50%, 24%, 11%, 5%, 1.8%, and 1.1% of the conformations, respectively. There are a total of 478 clusters (137 of which with more than one member) at 330 K and 2160 clusters (390 with more than one member) at 510 K. The center of the most populated cluster at 330 K has a 1.9 Å C_α-rmsd from the β-hairpin and is shown in Figure 8b. It has a three-residue loop (^DPGV) and five interstrand hydrogen bonds. The center of cluster 2 has an out-of-register type II turn at residues Val5-^{D}Pro6 with a bulge and three interstrand hydrogen bonds. Cluster 1 at 330 K corresponds to cluster 2 at 360 and 390 K, while cluster 2 at 330 K corresponds to cluster 1 at 360 and 390 K. However, clusters 1 and 2 are less populated at higher temperature values. Clusters 1 and 2 together incorporate 53%, 29%, and 10% of the conformations at 330, 360, and 390 K, respectively. In this temperature range, the folding time is therefore dominated by the time needed to get out from these two ensembles of structures. Above 420 K, the high activation entropy results in a slowing down of the folding rate as for peptide A at $T > 360$ K. It is interesting to note that the center of cluster 1 at 330 K (Table 4 and Figure 8) corresponds to the center of cluster 2 of the N trajectories at 360 and 450 K, and to the center of cluster 3 at 390 and 420 K (C_α-rmsd less than 1 Å). No correspondence was found between the clusters 1 and 2 of the R330 simulations and the clusters of the N run at either 300 or 330 K. To verify the stability of the three-residue loop at the ^DPGV site (R330 cluster 1) and the type II β-turn at V^{D}P (R330 cluster 2), two additional 50 ns MD simulations at 330 K were performed starting from the centers of clusters 1 and 2 of the R330 trajectories. In both cases the initial conformation broke apart and re-formed several times and was present 63% and 87% of the time for the ^DPGV loop and the type II turn at V^{D}P, respectively. This indicates that they are thermodynamically stable. It also implies that relevant conformations were not sampled in the N trajectory at either 300 or 330 K. As these two structures have none of the folded state contacts, this is not supposed to affect the energy landscape for $Q > 0$.

4. Discussion

To quantitatively investigate the thermodynamics and kinetics of folding, MD simulations of two model peptides, Ace-$(AAQAA)_3$-$NHCH_3$ (α-helical stable structure) and Ace-$V_5{}^DPGV_5$-NH_2 (β-hairpin), were performed using an implicit solvation model at different values of the temperature. Different starting conformations (folded and random) were used to obtain a statistically significant sampling of conformational space at each temperature value. The present discussion focuses on the main findings concerning the α-helix and β-hairpin, as well as a number of aspects that have emerged from the simulations and are of general interest for the protein folding problem.

4.1. α-Helix. Both experimental (see ref 61 for a review) and theoretical[62,63] studies have proposed that helix formation should occur on the submicrosecond time scale. On the other hand, Clarke et al. have reported recently that an alanine-based peptide folds with a rate constant of 15 s^{-1} at 273 K.[64] The use of an implicit solvation model accelerates conformational transitions in two ways. First, since it provides a mean solvation force, the solute−solvent potential energy is smoothed and therefore transitions are facilitated. Second, the friction exerted by the water molecules is missing. As a result, the folding rates calculated in this work should be seen as an upper bound. Nevertheless, the present results are consistent with the submicrosecond time scale, because it seems quite unlikely that the absence of explicit solvent could accelerate folding by nearly 7 orders of magnitude.

Under folding conditions helical nucleation is slower than propagation. Furthermore, nucleation can happen everywhere (apart from termini) and multiple folding pathways are possible in accord with recent reports on kinetic heterogeneity of protein folding.[65,66]

4.2. *β*-Hairpin. Little is known about the time scale for *β*-hairpin formation. A folding time of 6 *μ*s was measured at 300 K by temperature-jump experiments on a *β*-hairpin extracted from protein G B1.[10] The folding time of peptide B at 300 K is estimated to be 95.6 ns by a linear regression from the folding times at 330, 360, and 390 K (Arrhenius behavior for $T \leq 390$ K). It follows that peptide B folds nearly 30 times slower than peptide A at 300 K. At their respective melting temperatures of 320 and 370 K (Figure 4), peptide A folds about 2 times faster than peptide B (Table 3).

β-Hairpin formation can be favored by hydrophobic collapse, hydrogen bonding, and/or secondary structure propensities mainly related to the turn sequence.[3,4] It is likely that the relative importance of these factors depends on the amino acid sequence. The simulation results indicate that the D-Pro residue strongly favors a turn conformation which influences the sequence of events; i.e., folding is initiated by formation of the contacts close to the turn. Further, the turn sequence D-Pro-Gly equally populates the I′ and II′ types.

At mild temperatures, the free energy landscape of the *β*-hairpin shows a two-state behavior (with minima for folded and unfolded states), an important characteristic of the folding of small proteins. Therefore, future studies on *β*-hairpins and *β*-sheets will perhaps shed light on the protein folding problem. The *β*-hairpin simulations have revealed the presence of stabilizing nonnative interactions which result in local minima in the free energy landscape. This was not predicted by a statistical mechanical model for *β*-hairpin formation, because only native interactions were considered.[43] In the spirit of ref 43, one could use the MD approach to investigate the effects of mutations on the thermodynamics and kinetics of *β*-hairpin formation. This would clarify for instance which factors determine the position of the transition state which seems to be dependent on the amino acid sequence.

4.3. Implications for Protein Folding. A number of insights derived from the quantitative results of the present study are of interest because they may play a role in protein folding, in general. At mild temperature conditions, the profile of the average effective energy is rather flat along the first half of the folding process ($Q < 0.5$) and cannot explain fast folding. The flat energy profile is in agreement with MD simulations of folding and unfolding of CI2.[59,60]

An important feature of the folding of both peptides is the negative activation enthalpy at high temperatures. The rate constant for folding initially increases with temperature, goes through a maximum at about T_m, and then decreases. The non-Arrhenius behavior of the folding rate, demonstrated here with an atomistic model for the first time to the best of our knowledge, is in accord with experimental data on two mainly alanine α-helical peptides,[64,67] a *β*-hairpin,[10] CI2 and barnase,[68] lysozyme,[69] and lattice simulation results.[70−72] It has been proposed that the non-Arrhenius profile of the folding rate originates from the temperature dependence of the hydrophobic interaction.[73,74] Our results show that a non-Arrhenius behavior can arise at high values of the temperature in a model where all the interactions are temperature independent. This has been found also in lattice simulations.[70,71] The curvature of the folding rate at high temperature may be a property of a reaction dominated by enthalpy at low temperatures and entropy at high temperatures.[72] The non-Arrhenius behavior for a system where the interactions do not depend on the temperature might be a simple consequence of the temperature dependence of the accessible configuration space. At low temperatures, an increase in temperature makes it easier to get over the energy barriers, which are rate limiting. However, at very high temperatures, a larger portion of the configuration space becomes accessible, which results in a slowing down of the folding process.

According to the present simulation results, the sequence of events for unfolding is the inverse of the one for folding and does not depend on the temperature. These results are probably due to the small size of peptides, which makes their folding reaction less complex than the one of proteins. Nevertheless, targeted MD trajectories of CI2[59] and lattice simulation analysis of a 125-residue protein model[75] indicate that the sequences of events for folding and unfolding are similar for proteins that lack off-pathway intermediates.

Acknowledgment. We thank Dr. E. Paci and Prof. M. Karplus for interesting discussions. Ph.F. is a Fellow of the Roche Research Foundation. This work was supported in part by the Swiss National Science Foundation (Grant 31-53604.98 to A.C.).

References and Notes

(1) Chakrabartty, A.; Baldwin, R. L. *Adv. Protein Chem.* **1995**, *46*, 141−176.
(2) Munõz, V.; Serrano, L. *Curr. Opin. Biotechnol.* **1995**, *6*, 382−386.
(3) Blanco, F.; Ramirez-Alvarado, M.; Serrano, L. *Curr. Opin. Struct. Biol.* **1998**, *8*, 107−111.
(4) Gellman, S. H. *Curr. Opin. Chem. Biol.* **1998**, *2*, 717−725.
(5) Caflisch, A.; Karplus, M. Molecular dynamics studies of protein and peptide folding and unfolding. In *The Protein Folding Problem and Tertiary Structure Prediction*; Merz, K. M., Jr., LeGrand, S. M., Eds.; Birkhäuser: Boston, 1994; pp 193−230.
(6) Wu, X.; Wang, S. *J. Phys. Chem. B* **1998**, *102*, 7238−7250.
(7) Elber, R.; Meller, J.; Olender, R. *J. Phys. Chem. B* **1999**, *103*, 899−911.
(8) Daura, X.; van Gunsteren, W. F.; Mark, A. E. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 269−280.
(9) Galzitskaya, O.; Caflisch, A. *J. Mol. Graphics Modell.* **1999**, *17*, 19−27.
(10) Munõz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196−199.
(11) Prevost, M.; Ortmans, I. *Proteins: Struct., Funct., Genet.* **1997**, *29*, 212−227.
(12) Sung, S. *Biophys. J.* **1999**, *76*, 164−175.
(13) Schaefer, M.; Bartels, C.; Karplus, M. *J. Mol. Biol.* **1998**, *284*, 835−848.
(14) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068−9073.
(15) Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9062−9067.
(16) Shalongo, W.; Laxmichand, D.; Stellwagen, E. *J. Am. Chem. Soc.* **1994**, *116*, 8288−8293.

(17) Sung, S.; Wu, X. *Proteins: Structure, Function and Genetics* **1996**, *25*, 202−214.
(18) Shirley, W. A.; Brooks, III, C. L. *Proteins: Struct., Funct., Genet*. **1997**, *28*, 59−71.
(19) Karle, I. L.; Awasthi, S. K.; Balaram, P. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8189−8193.
(20) Haque, T. S.; Little, J. C.; Gellman, S. H. *J. Am. Chem. Soc*. **1996**, *118*, 6975−6985.
(21) Haque, T. S.; Gellman, S. H. *J. Am. Chem. Soc*. **1997**, *119*, 2303−2304.
(22) Stanger, H. E.; Gellman, S. H. *J. Am. Chem. Soc*. **1998**, *120*, 4236−4237.
(23) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem*. **1983**, *4*, 187−217.
(24) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys*. **1996**, *105*, 1902−1921.
(25) Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928−1931.
(26) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct., Genetics* **1999**, *35*, 133−152.
(27) Mehler, E. L.; Eichele, G. *Biochemistry* **1984**, *23*, 3887−3891.
(28) Ramstein, J.; Lavery, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7231−7235.
(29) Schaefer, M.; Bartels, C.; Karplus, M. *Theor. Chem. Acc*. **1999**, *101*, 194−204.
(30) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199−203.
(31) Hasel, W.; Hendrickson, T. F.; Still, W. C. *Tetrahedron Comput. Methodol*. **1988**, *1*, 103−116.
(32) Fraternali, F.; van Gunsteren, W. F. *J. Mol. Biol*. **1996**, *256*, 939−948.
(33) De Alba, E.; Santoro, J.; Rico, M.; Jiménez, M. A. *Protein Sci*. **1999**, *8*, 854−865.
(34) Schenck, H. L.; Gellman, S. H. *J. Am. Chem. Soc*. **1998**, *120*, 4869−4870.
(35) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys*. **1984**, *81*, 3684−3690.
(36) van Holde, K. E.; Johnson, W. C.; Ho, P. S. *Principles of physical biochemistry*; Prentice Hall: New York, 1998; p 157.
(37) Wilmot, C. M.; Thornton, J. M. *J. Mol. Biol*. **1988**, *203*, 221−232.
(38) McQuarrie, D. A. *Statistical mechanics*; Harper and Row: New York, 1976.
(39) Lifson, S.; Roig, A. *J. Chem. Phys*. **1961**, *34*, 1963−1974.
(40) Sibanda, B. L.; Thornton, J. M. *Nature* **1985**, *316*, 170−174.
(41) Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. *Biochemistry* **1996**, *35*, 691−697.
(42) Gilmanshin, R.; Williams, S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3709−3713.
(43) Munõz, V.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5872−5879.
(44) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *J. Mol. Biol*. **1995**, *252*, 460−471.
(45) Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. *Protein Sci*. **1994**, *3*, 843−852.
(46) Davis, M. E.; McCammon, J. A. *J. Comput. Chem*. **1989**, *10*, 386−391.
(47) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun*. **1991**, *62*, 187−197.
(48) Scarsi, M.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. A* **1997**, *101*, 8098−8106.
(49) Scarsi, M.; Caflisch, A. *J. Comput. Chem*. **1999**, *14*, 1533−1536.
(50) Scholtz, J. M.; Marqusee, S.; Baldwin, R. L.; York, E. J.; Stewart, J. M.; Santoro, M.; Bolen, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2854−2858.
(51) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys*. **1959**, *31*, 526−535.
(52) Soman, K. V.; Karimi, A.; Case, D. A. *Biopolymers* **1991**, *31*, 1351−1361.
(53) Tobias, D. J.; Brooks, III, C. L. *Biochemistry* **1991**, *30*, 6059−6070.
(54) Tirado-Rives, J.; Jorgensen, W. L. *Biochemistry* **1991**, *30*, 3864−3871.
(55) Young, W. S.; Brooks, III, C. L. *J. Mol. Biol*. **1996**, *259*, 560−572.
(56) Tirado-Rives, J.; Maxwell, D. S.; Jorgensen, W. L. *J. Am. Chem. Soc*. **1993**, *115*, 11590−11593.
(57) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *J. Chem. Phys*. **1994**, *101*, 6052−6062.
(58) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct., Funct., Genet*. **1995**, *21*, 167−195.
(59) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins: Struct., Funct., Genet*. **2000**, *39*, 252−260.
(60) Ferrara, P.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem*. **2000**, in press.
(61) Eaton, W. A.; Munoz, V.; Thompson, P. A.; Chan, C. K.; Hofrichter, J. *Curr. Opin. Struct. Biol*. **1997**, *7*, 10−14.
(62) Brooks, III, C. L. *J. Phys. Chem*. **1996**, *100*, 2546−2549.
(63) Takano, M.; Yamato, T.; Higo, J.; Suyama, A.; Nagayama, K. *J. Am. Chem. Soc*. **1999**, *121*, 605−612.
(64) Clarke, D. T.; Doig, A. J.; Stapley, B. J.; Jones, G. R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 7232−7237.
(65) Goldbeck, R. A.; Thomas, Y. G.; Chen, E.; Esquerra, R. M.; Kliger, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2782−2787.
(66) Sabelko, J.; Ervin, J.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6031−6036.
(67) Lednev, I. K.; Karnoup, A. S.; Sparrow, M. C.; Asher, S. A. *J. Am. Chem. Soc*. **1999**, *121*, 8074−8086.
(68) Oliveberg, M.; Tan, Y. J.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8926−8929.
(69) Segawa, S. I.; Sugihara, M. *Biopolymers* **1984**, *23*, 2473−2488.
(70) Karplus, M.; Caflisch, A.; Šali, A.; Shakhnovich, E. *Mol. Eng*. **1995**, *5*, 55−70.
(71) Karplus, M. *Folding Des*. **1997**, *2*, S69−S75.
(72) Dobson, C. M.; Šali, A.; Karplus, M. *Angew. Chem., Int. Ed*. **1998**, *37*, 869−893.
(73) Scalley, M. L.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10636−10640.
(74) Chan, H. S.; Dill, K. A. *Proteins: Struct., Funct., Genet*. **1998**, *30*, 2−33.
(75) Dinner, A. R.; Karplus, M. *J. Mol. Biol*. **1999**, *292*, 403−419.