Characterization of the Stereochemical Selectivity of β-Hairpin Formation by Molecular Dynamics Simulations

Patricia Soto* and Ronen Zangi

Groningen Biomolecular Sciences and Biotechnology Institute (GBB), Department of Biophysical Chemistry, University of Groningen, Groningen, The Netherlands

Received: September 20, 2004; In Final Form: November 9, 2004

The stability of secondary structure motifs found in proteins is influenced by the choice of the configuration of the chiral centers present in the amino acid residues (i.e., D vs L). Experimental studies showed that the structural properties of the tetrapeptide $^{L}V^{L}P^{L}A^{L}L$ (all-L) are drastically altered upon mutating the L-proline and the L-alanine by their D-enantiomers [*J. Am. Chem. Soc.* **1996**, *118*, 6975]. The all-L diastereomer is unstructured, experiencing little or no β-hairpin formation, while the $^{L}V^{D}P^{D}A^{L}L$ peptide exhibits a substantial population of β-hairpin conformation. In this study, we perform molecular dynamics simulations to investigate the folding propensity of these two model peptides. The results confirm the experimental findings, namely, that the presence of D-amino acids in the loop region strongly induces β-hairpin formation (a population increase from about 1.5% to 50% is observed). The major factor determining the different behavior is found to be the large difference in energy between the two diastereomers, approximately 22 kJ/mol, when they adopt a β-hairpin structure. The higher energy observed for the all-L peptide is a consequece of none-ideal hydrogen bond formation and of steric repulsions. The results suggest that selective incorporation of D-amino acids in proteins can be used to enhance certain secondary structure elements. The kinetic behavior of the folding process observed in the simulations is also investigated. We find that the decay rate of the folded structure fits to a biexponential function, suggesting that the folding/unfolding process of a β-hairpin is governed by two different mechansims.

Introduction

Understanding the tendency of the different amino acids to induce the formation of secondary structures is crucial for the prediction and engineering of protein structure. Most proteins are compact with a globular shape. This is facilitated by the numerous occurrence of β-turn secondary motifs allowing the reversal of the direction of the polypeptide chain. In this secondary structure, the ϕ, ψ conformational space of residues 2 and 3 (the central residues) is restricted. On the basis of the values these torsional angles can have, the different conformations observed for β-turns are classified into different types. The majority of the β-turns in folded proteins are of type I ($\phi_2 = -60°$, $\psi_2 = -30°$, $\phi_3 = -90°$, $\psi_3 = 0°$) and type II ($\phi_2 = -60°$, $\psi_2 = +120°$, $\phi_3 = +90°$, $\psi_3 = 0°$) conformations which are also known as "common" turns. In addition to their importance in allowing the polypeptide chain to attain a globular shape, β-turns play also an important role in the formation of antiparallel β-sheet structures that are linked by hydrogen bonds. The arrangement of the polypeptide chain into strand-loop-strand structure is often observed in proteins and referred to as a β-hairpin. The shortest common loop is two residues long (residues 2 and 3 of a β-turn). However, in contrast to the conformations of the "common" turns, the vast majority of two-residue β-hairpin loops exhibit different conformations. They are characterized by ϕ and ψ torsional angle values that are opposite in sign compare to the corresponding values found in the "common" turns. These β-turns, also known as "mirror image" turns, are named type I′ and type II′. The "mirror image" conformations were proposed by Sibanda and Thornton[1] to be adopted by two-residue β-hairpin loops (in contrast to the general preference for type I and type II turns) due to their geometrical compatibility with the twist formed by the two strands of antiparallel β-sheet. They are predicted to be energetically less favorable because of steric hindrance.[2] However, they exist in β-hairpin loops due to the energy gain associated with the formation of hydrogen bonds between the two β-strands.

The observation that a peptide chain from all-L amino acids residues exhibits preference for "common turns" raised the intriguing possibility of enhacing the preference toward "mirror image" turns by replacing the l-amino acids in the loop region with their D-enantiomers and thereby promoting the formation of β-hairpins. Gellman and co-workers[3] studied by IR and NMR spectroscopies the folding propensity of diastereomeric "minimal" β-hairpin models. "Minimal" β-hairpins are formed by a four residue segments. They are defined by the presence of backbone hydrogen bonds between residues 1 and 4 (thus, forming two intramolecular hydrogen bonds). The advantage in comparing diastereomeric peptides (obtained by replacing some of the L-amino acids with their D enantiomers) is that the options of intramolecular hydrogen bonds formation are the same, however, the backbone conformational strain is different. Short peptides often adopt a random coil configuration in solution because the potential energy gain is insufficient to overcome the entropy loss upon folding. Thus, stable folded structures are observed if favorite intramolecular hydrogen bondings can form. This was enhanced by conducting the experimental studies in a relatively nonpolar solvent which does not provide hydrogen bonding competition. In addition, the

* Corresponding author. Current address: Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106. Telephone: +1 (805) 8932767. E-mail: psoto@chem.ucsb.edu.

10.1021/jp045729z CCC: $30.25
Published on Web 01/20/2005

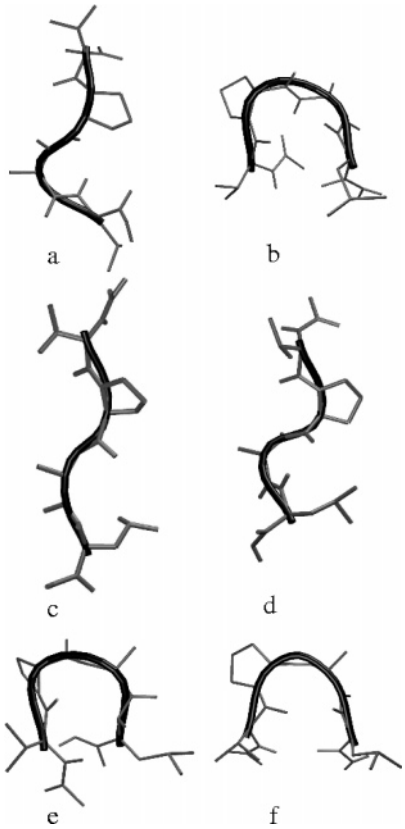

Figure 1. (a) Initial structures of trajectories D1 to D10 and (b) central structure of the most populated cluster of the combined trajectory of $^{L}V^{D}P^{D}A^{L}L$. (c) Initial structure of 100 ns trajectory and (d) central structure of the most populated cluster of $^{L}V^{L}P^{L}A^{L}L$. (e) Initial structure of the trajectory started from a turn geometry with ω of D-Pro in cis configuration for $^{L}V^{D}P^{D}A^{L}L$ and (f) the structure after 60 ns of simulation of this trajectory. Figures generated by Raster3D.[7]

presence of L-proline at the second of the four β-turn residues is well-known to promote formation of type I and type II turns. For these reasons, the model end-capped tetrapeptides (among others), $^{L}V^{L}P^{L}A^{L}L$ and $^{L}V^{D}P^{D}A^{L}L$, were examined in solution of methylene chloride. They found that the all-L tetrapeptides experience little or no β-hairpin formation while the diastereomers are largely folded into β-hairpin conformation. The difference in the folding behavior was attributed to the difference between torsional strain, steric repulsion and conformational entropy of the two diastereomeric backbones. The same behavior was obtained with the diastereomeric pair, $^{L}V^{L}PG^{L}L$ and $^{L}V^{D}PG^{L}L$ as well as with analogous depsipeptides with lactic acid or glycolic acid residue at the third position. However, the diastereomeric VPAL peptides showed the most distinctive folding preferences. Therefore, we have chosen them as model peptides for this study (Figure 1) to investigate the factors that determine the dramatic difference in the folding behavior. More complex systems have also been analyzed, both with experimental NMR methods[4] and molecular simulations,[5] to study the influence of stereochemistry on β-hairpin formation. In those cases, not only the β-turn propensity but also the hydrophobicity of strand residues were determinant to dictate β-hairpin stability.

For this study the system chosen has two main advantages. First, it is sufficiently small to be simulated on a 10^2 ns time scale in explicit solvent. Second, the peptides show distinct conformational preferences and have been well characterized experimentally using various techniques. This is important as a recurring difficulty in assessing whether molecular dynamics (MD) simulations accurately reproduce the conformational behavior of peptides in solution is that the available experimental data is limited, and that experimental data at atomic resolution is normally only available for equilibrium distributions of conformations under specific conditions. In addition, the difference in free energy between microscopic states is normally small[6] (a few k_BT), placing high demands on the accuracy of the atomic interaction function (force field) used in the simulations.

We have simulated the tetrapeptides for more than a combined time of 1 μs in explicit solvent. This has enabled the back calculation from the trajectories of elements of the IR and NMR spectra. The agreement between the calculated and experimental spectra indicates that a statistically significant proportion of the conformational space accessible to the two peptides has been sampled providing therefore a solid ground for the characterization of the folding behavior of the peptides. In particular, we find that the all-L peptide in solution display hardly any β-hairpin conformation while the diastereomer (obtained by mutating the two central amino acids to their D-enantiomers) is found approximately 50% of its time in a β-hairpin conformation. The major factor determining the different behavior is found to be the large difference in energy between the two diastereomers, approximately 22 kJ/mol, when they adopt a β-hairpin structure. The higher energy observed for the all-L peptide is a consequece of none-ideal hydrogen bond formation and of steric repulsions. The kinetic behavior of the folding process observed in the simulations is also investigated. We find that the decay rate of the folded structure fits to a biexponential function suggesting that the folding/unfolding process of a β-hairpin is governed by two different mechanisms.

Methods

MD Simulations. Atomistic MD simulations in explicit solvent were performed to study the reversible folding of two diastereomeric peptides: $^{L}V^{D}P^{D}A^{L}L$ (Ace-Val-D-Pro-D-Ala-Leu-Nme_2) and $^{L}V^{L}P^{L}A^{L}L$ (Ace-Val-Pro-Ala-Leu-Nme_2). Experiments to study folding propensities of these peptides were performed in methylene chloride. An aprotic solvent was chosen to prevent hydrogen bonding competition with the intramolecular bonding network. In the simulations chloroform was used as the solvent. This was because a well tested model for chloroform was available in the force field. As the physical properties of chloroform and methylene chloride are similar it is expected that the folding behavior of the peptide in the two solvents will be highly comparable.

The peptides were simulated in a (periodic) dodecahedron box with the minimum distance between the solute and the wall of the box being 0.9 nm when the peptide was in an extended conformation. The total charge on the peptide was zero. The GROMOS96 43A1 force field[8,9] was used for the peptide and the solvent. The temperature was maintained close to the intended value of $T = 300$ K by weak coupling to an external temperature bath[10] with a coupling constant of 0.1 ps. The LINCS[11] algorithm was used to constrain bond lengths within the peptide and the solvent. The time step used to integrate Newton's equations of motion was 2 fs. A twin range cutoff of 0.8/1.4 nm was used for nonbonded interactions. According to this method, the interactions within the shorter range cutoff are updated every step, the interactions between the shorter and the longer range cutoff are updated every five steps, and the interactions beyond the longer range cuttoff are ignored. All simulations and analysis were performed using the GROMACS package version 3.1.[12,13] After the solvation of each peptide in chloroform and energy minimization the systems were each

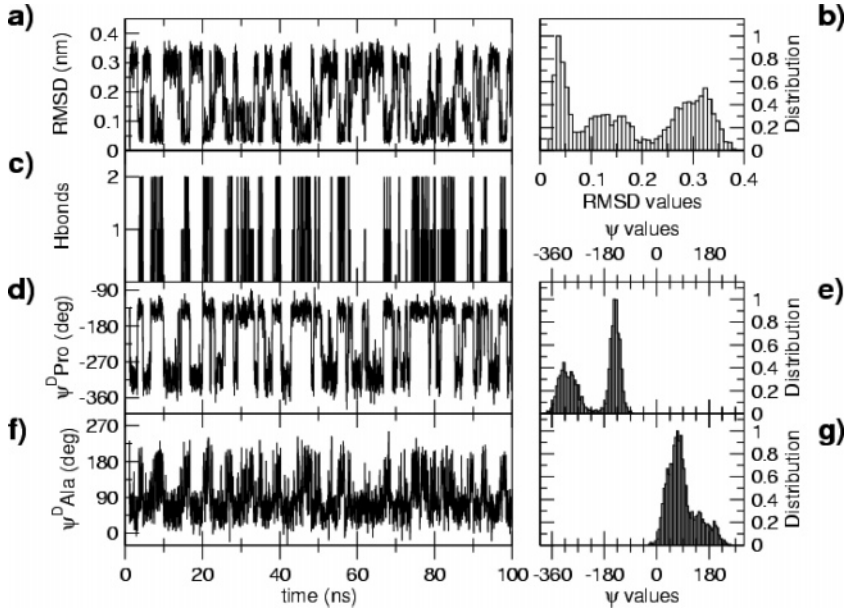

Figure 2. RMSD of D1 (peptide $^{L}V^{D}P^{D}A^{L}L$, initial structure in an extended conformation) with respect to the central structure of the most populated cluster (a) and the respective distribution of RMSD values (b). Existence map of hydrogen bonds connecting the backbone of residues i, i + 3 for D1 (c) hbond1 corresponds to V1 N V1 H−L4 O; hbond2 corresponds to L4 O L4 H−V1 O. Time series and ψ dihedral distribution of D-Pro (d and e) and D-Ala (f and g).

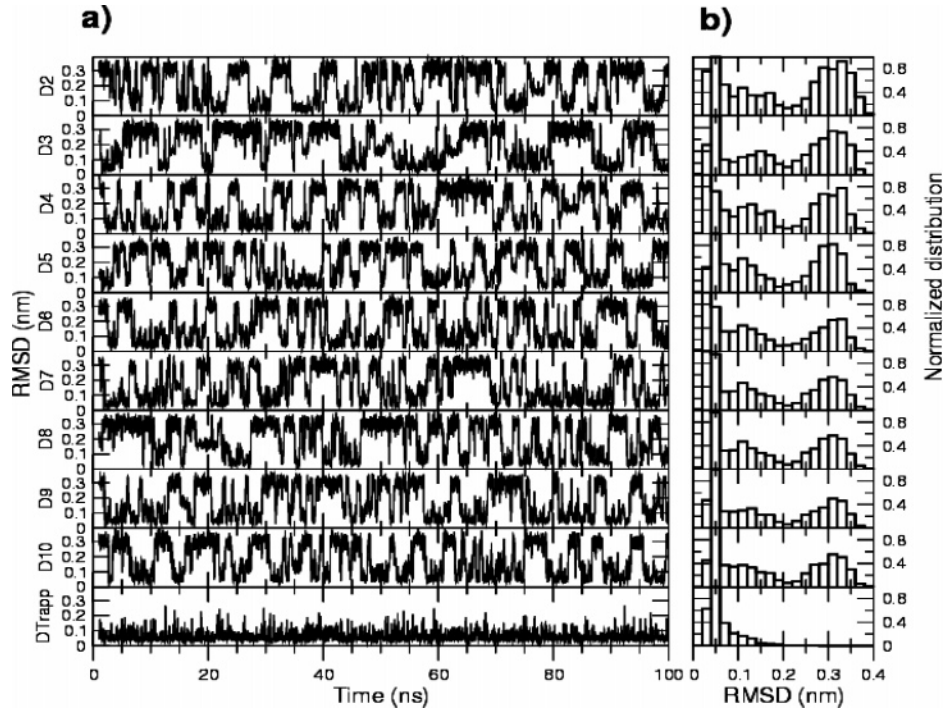

Figure 3. RMSD of D2 to D10 (initial structure in an extended conformation, each trajectory with different set of initial velocities) and of D-trapp (initial structure in a turn conformation with proline as cis rotamer). In each case, RMSD calculated with respect to the central member of the respective most populated cluster (a) and the distribution of RMSD values (b).

equilibrated for 0.2 ns. The resuling systems were used as the starting configurations for further work. For $^{L}V^{D}P^{D}A^{L}L$, a total of ten 100 ns runs, started from an extended conformation but each with a different set of initial velocities, were performed, giving 1 μs in total. In addition a 100 ns simulation was performed starting from an ideal turn. For $^{L}V^{L}P^{L}A^{L}L$, two runs were performed, one 100 ns starting from an ideal turn and one 100 ns starting from an extended conformation. For analysis, the first nanosecond of each trajectory was disregarded.

Analysis of the Trajectories. Structural Properties. The positional root-mean-square deviation (RMSD) of atoms was calculated after fitting the nth structure (R_n) to the reference structure (R_{ref}) subsequently calculating the RMSD via the equation

$$\mathrm{RMSD}(R_n,R_{ref}) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\mathbf{R}_{n_k} - \mathbf{R}_{\mathrm{ref}_k})^2}$$

where $\mathbf{R}_{n_k}$ and $\mathbf{R}_{\mathrm{ref}_k}$ represent the Cartesian vector of the kth atom (atom k = N, H, Cα, C, O for all residues) of structures n and ref, respectively.

Cluster Analysis. A series of nonoverlapping clusters of structures were obtained by calculating the RMSD between all

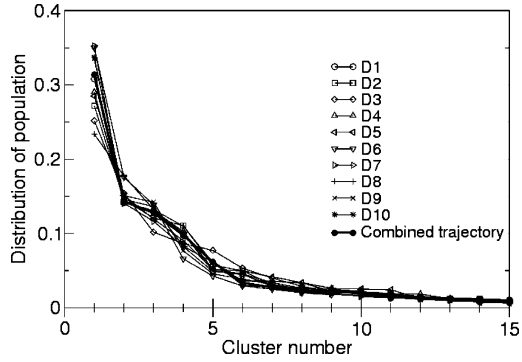

Figure 4. Distribution of population of the 10 trajectories D1 to D10 and of the combined trajectory.

pairs of structures. Then, the structure with the largest number of neighbors that satisfy the condition RSMD < 0.08 nm (considered the central structure of the cluster) was taken together with the neighbors to form the (first) cluster and eliminated from the pool of structures. This process was repeated until the pool of structures was empty.[14] For trajectories < 100 ns, structures were sampled every 0.04 ns. For the combined trajectory (∼1 μs), structures were sampled every 0.08 ns.

Hydrogen Bonds. Hydrogen bonds were calculated between residues i and $i + 3$. A β-turn was considered to exist if a hydrogen bond was found between Val NH−Leu CO or Leu NH−Val CO. A hydrogen bond is defined to exist if the donor−acceptor distance was less than 0.25 nm and the hydrogen−donor−acceptor angle less than 60°.

IR Spectra. For each trajectory, central structures of the clusters that together cover 90% of the total population sampled were used for the calculation of the IR spectrum. The semiempirical PM3 Hamiltonian[15] within MOPAC 2000[16] was used to calculate the contribution of each structure. Each structure was first geometry optimized. A normal mode calculation was then performed by diagonalization of the mass-weighted Hessian matrix **H**:

$$H_{ij}^{m} = \frac{1}{\sqrt{m_i m_j}} \frac{\partial^2 E}{\partial x_i \partial x_j}$$

to obtain the vibrational frequencies and corresponding normal coordinates. The transition dipole was then calculated for the fundamental of each normal mode:

$$|T_{oi}|^2 = |\langle 0|\mathbf{D}|i\rangle|^2$$

with T_{oi} the magnitude of the transition dipole moment, $|0\rangle$ the ground state, $|i\rangle$ the ith mode, and **D** the dipole operator. The absorbance was taken as proportional to the magnitude of the transition dipole and the contribution of each structure to the spectrum was weighted according to the population of the cluster it represents. The curve reported in Figure 8 corresponds to a least-squares fit of the data to a polynomial of the 10th degree. To check the contribution to the spectrum of the representative structure of each cluster, an additional test was performed for structures randomly selected that belong to the same cluster. It was found that frequency and intensity values for the alternative structures converge to the same values.

NMR Cross Relaxation Rates. NOE intensities as a function of mixing time were calculated for specific proton pairs using the method of Peter et al.[19] in which the internal dynamics of the peptide is included in the algorithm by computing the spectral density functions for all proton pairs from the time

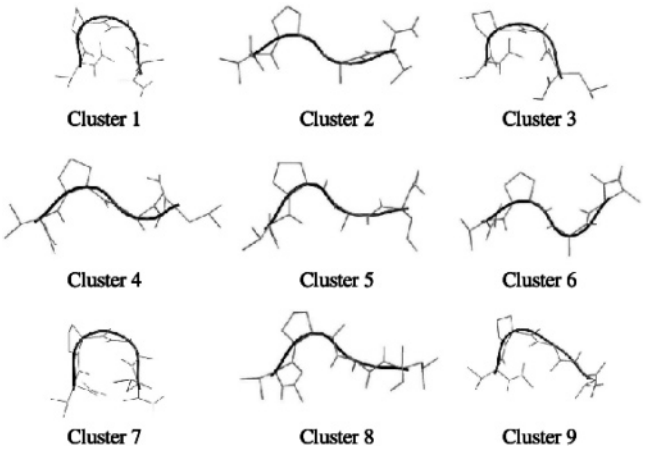

Figure 5. Central structures of the top nine most populated clusters after applying the cluster procedure to the combined trajectory of $^{L}V^{D}P^{D}A^{L}L$.

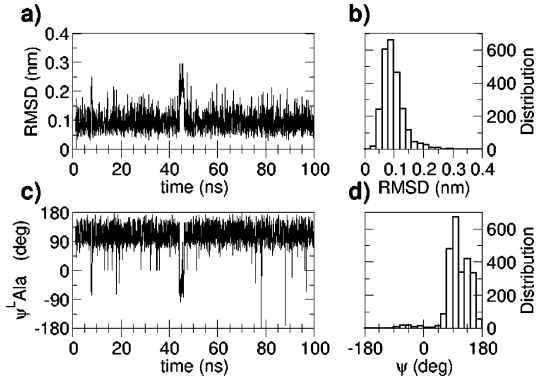

Figure 6. RMSD of 100 ns trajectory of $^{L}V^{L}P^{L}A^{L}L$ (initial structure in an extended conformation) respect to central structure of the most populated cluster (a) and the respective distribution of RMSD values (b). Time series of ψ backbone dihedral angle of L-Ala (c) and the corresponding distribution of ψ values (d).

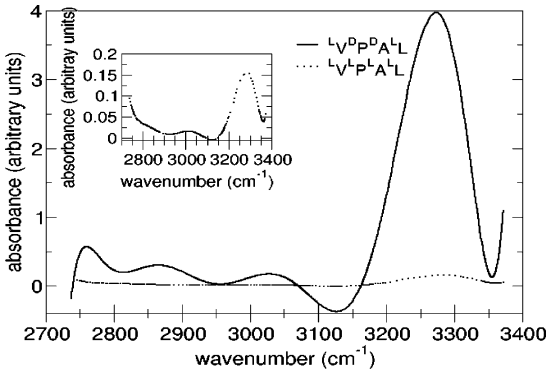

Figure 7. Calculated IR spectra of $^{L}V^{D}P^{D}A^{L}L$ and $^{L}V^{L}P^{L}A^{L}L$. Inset: Expansion of the $^{L}V^{L}P^{L}A^{L}L$ spectra.

correlation functions of the interproton distance vectors and then solving the system of coupled differential equations for the NOESY intensities by diagonalizing the relaxation matrix **R** (see ref 19):

$$\frac{d\mathbf{M}(t)}{dt} + \mathbf{R}\mathbf{M}(t) = 0$$

with $M_{ii} = (I_z(t) - I_o)_i$, for $i = j$, the deviation of the z component expectation value of the spin operator of spin i, and $M_{ij} = 0$, for $i \neq j$. $R_{ii} = \rho_i$, the total longitudinal relaxation rate of spin i including relaxation due to other relaxation mecha-

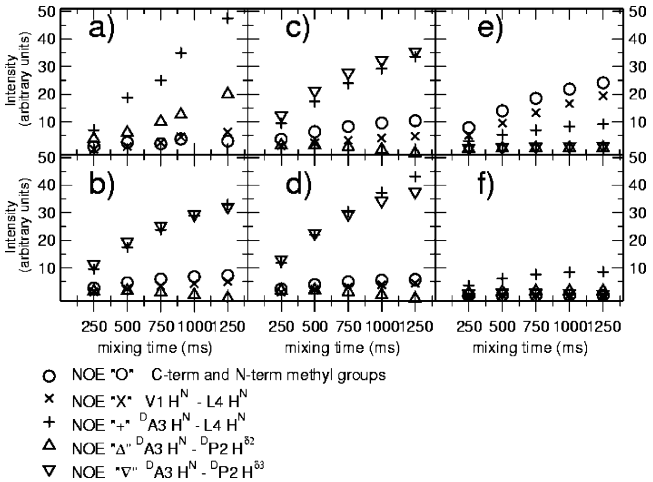

Figure 8. (a) NOE cross-peak intensities at different mixing times taken from Figure 11 of Haque et al.[2] (b) NOE intensities as a function of mixing time averaged over the 10 ^{L}V^{D}P^{D}A^{L}L trajectories D1 to D10. Results from the two trajectories that show the greatest deviations from the average, D1 (c) and D6 (d). NOE intensities as a function of mixing time for D-trapp (e) and those for the 100 ns trajectory of ^{L}V^{L}P^{L}A^{L}L (f). NOEs are indicated as follows: V1 H^{N}−L4 H^{N} "×", C-terminal and N-terminal methyl groups "○", ^{D}A3H^{N}−L4H^{N} "+", ^{D}A3H^{N}−^{D}P2 H$^{\delta 2}$ "Δ", and ^{D}A3H^{N}−^{D}P2 H$^{\delta 3}$ "∇".

nisms, and $R_{ij} = \sigma_{ij}$, the cross relaxation rate. $\mathbf{M}(t_m)$ is the matrix with the NOE intensities after a mixing time t_m.

Results and Discussion

^{L}V^{D}P^{D}A^{L}L. Figure 2a shows the root-mean-square positional deviation (RMSD) of the main chain atoms (N, H, C_α, C, O) as a function of simulation time for the trajectory D1 (of ^{L}V^{D}P^{D}A^{L}L peptide). As there is no experimental 3-dimensional structure of this tetrapeptide, the RMSD was calculated with respect to the central structure of the most populated cluster (see Figure 1b). As can be seen in Figure 2a there is a rapid interconversion between large (RMSD > 0.25 nm) and small (RMSD < 0.1 nm) values of the RMSD.

The two-state behavior of the system is also shown in the distribution of the RMSD values plotted in Figure 2b.

To test how reliably a 100 ns simulation samples the configurational space accessible to this (very) short peptide, nine more simulations were initiated from the same starting configuration but with different initial velocities. Figure 3 shows the RMSD as a function of time for each of these trajectories with respect to the central member of the dominante cluster in each trajectory. Although the trajectories are clearly distinguishable from each other the statistical averages of the thermodynamical properties are the same. In addition, the central members of the dominante cluster in each trajectory are essentially identical all having an RMSD within 0.03 nm. The population of the dominant cluster varies from 35% for trajectory D7 to 23% for D5.

To characterize the conformations sampled, elements of secondary structure present in trajectory D1 were examined. A β-hairpin was considered to exist if there was a hydrogen bond connecting the backbones of residues Val and Leu (i to $i + 3$). The existence of a hydrogen bond was determined based on geometric criteria (see Methods). The presence of the two possible hydrogen bonds between Val and Leu as a function of time for trajectory D1 is plotted in Figure 2c. As can be seen there is a strong correlation between low RMSD values and the presence of a β-hairpin.

The behavior of the backbone dihedral angles was also analyzed. A plot of the ψ dihedral angle of D-Pro and D-Ala (Figure 2, parts d−g) supports the idea of transitions between states. The transitions of the ψ torsional angle of D-Pro and D-Ala seems to be correlated to the hydrogen bond formation. The distribution of the ω angle of D-Pro is centered around −180°, indicating a trans configuration for this residue. The general correlation between the presence of hydrogen bonds between Val and Leu, and ψ dihedral of D-Pro and D-Ala is present in all 10 trajectories (data not shown) as well as the distribution of ω values (data not shown). All this evidence let us propose that the conformational transitions illustrated in Figure 2 and Figure 3 correspond essentially to an interconversion between extended and hairpinlike conformations.

To generate a 1 μs trajectory the 10 100 ns trajectories were combined and the cluster procedure reapplied. The distribution of the population of clusters in each trajectory and in the combined trajectory is shown in Figure 4. The central members of the most populated clusters from the combined trajectory is shown in Figure 5. As can be seen the dominant cluster (β-hairpin-like) comprises approximately 30% of the ensemble. The other significantly populated conformers are mainly (partially) extended structures. No conformer beyond the fifth cluster comprises more than 5% of the ensemble in any trajectory. The top five clusters make up approximately 75% of the ensemble, and the top 14 clusters make up 90% of the ensemble.

From the above, it is clear that the trajectories all sample similar regions of the configurational space when initiated from an extended structure. An additional simulation was, however, performed under the same conditions starting from a turn geometry with ω of D-Pro in cis configuration, i.e., angle around C(D-Pro)−N(D-Ala) equal to zero (see Figure 1e) for 100 ns. This simulation was performed in order to provide a reference for the calculation of various experimental properties as discussed later. As is indicated in the lowest plot in Figure 3a and Figure 1f, the system remained trapped in the turn geometry for 100 ns and no transitions to significantly different conformational states were observed. Comparing the ψ dihedral angle distribution of D-Pro between the 10 nonbiased trajectories and the trapped one (starting with a cis configuration for D-Pro) indicates that the latter samples a region of the Ramachandran plot that is never visited by the other trajectories. The cis configuration for D-Pro cannot be considered a state of global minimum because the results from the simulations indicate that its intrapeptide energy is higher by approximately 15 kJ/mol than the β-hairpin conformation with a trans configuration for D-Pro (assuming that the difference of the interactions with the solvent and the difference of the configurational entropies are small). Therefore, this trapped conformation corresponds to a state sorrounded by high free energy barriers. Since the cis configuration break up for the all-L peptide after 1 ns (results not shown) this topology of the free energy surface is clearly due to the presence of the D-amino acids residues that constrain the dynamics from the turn geometry to the other states in the system.

^{L}V^{L}P^{L}A^{L}L. Figure 6a shows the RMSD as a function of time for the 100 ns simulation of ^{L}V^{L}P^{L}A^{L}L. Again the reference structure is the central structure of the most populated cluster (see Figure 1d). Unlike what was observed for ^{L}V^{D}P^{D}A^{L}L, ^{L}V^{L}P^{L}A^{L}L shows no obvious preference for a single well-defined conformation. Note that the relatively small magnitude of the fluctuations observed in the RMSD plot in Figure 6a does not necessarily indicate that the peptide adopts a single structure. In fact, backbone dihedral angles of the residues in the middle of the sequence adopt a wide range of values (−85° < ϕ(L-Pro) < −30°, 60° < ψ(L-Pro) < 180°, −165° < ϕ(L-Ala)

$< -135°$, $60° < \phi$(L-Ala) $< 170°$) that did not correlate to patterns that could be associated with elements of secondary structure. Only one event which could clearly be considered a structural transition was observed after approximately 44 ns. This involved a transition of the ψ angle of L-Ala (Figure 6c) and the transient formation of a hydrogen bond between Val and Leu suggesting the formation of a hairpin structure. The remarkable different behavior observed for the two diastereomers in solution raises questions in regard to the driving force that prevent the $^{L}V^{L}P^{L}A^{L}L$ peptide from adopting the β-hairpin conformation. From the analysis of the interactions inside the peptide we find that the energy of the $^{L}V^{L}P^{L}A^{L}L$ in a β-hairpin conformation is higher by 22 kJ/mol than the intrapeptide energy of the $^{L}V^{D}P^{D}A^{L}L$ in the same conformation. Although, in both cases the unfolded conformation is higher in energy relative to the folded state (differences of 10 and 36 kJ/mol were obtained for the all-L and $^{L}V^{D}P^{D}A^{L}L$, respectively), only in the $^{L}V^{D}P^{D}A^{L}L$ case, this energy gain can compensate for the loss in entropy of the folded state. However, since these are very short peptides, the magnitude of the thermal fluctuations relative to the heights of the barriers that separate different minima is large. This mean that minima that do not correspond to the most stable state can also be sampled according to the Boltzmann weight of their free energies. The larger interpeptide energy observed for the all-L peptide is a result of both higher Coulumb and higher Lennard-Jones interactions. Thus, this is probably due to none-ideal hydrogen bond geometries and of steric repulsions as proposed by Sibanda and Thornton.[17]

To test if the formation of a stable hairpin was possible and to determine whether the first simulation was representative of the conformational preferences of $^{L}V^{L}P^{L}A^{L}L$, a second simulation was initiated starting from an ideal turn geometry. In this simulation the turn conformation was lost immediately. As only one transition was observed in 200 ns and the lifetime of the hairpin conformation was short, any hairpinlike structures are unlikely to contribute significantly to the ensemble of states in solution. Therefore, the thermodynamic average obtained for this state is likely to suffer from insufficient sampling. Nevertheless, the results agree with the observations from the experiments indicating that the tetrapeptide $^{L}V^{L}P^{L}A^{L}L$ has a little or no propensity to fold into a β-hairpin while the $^{L}V^{D}P^{D}A^{L}L$ has a high propensity to adopt a β-hairpinlike fold.

Comparison to Spectroscopy Data. To validate the results of the simulations and to obtain insight into the meaning of spectroscopic data available aspects of the infrared (IR) and nuclear magnetic resonance (NMR) spectra which give insight into the conformational properties of the peptides were calculated based on the trajectories as described in the methods.

IR Absorbance. FT-IR spectra, in particular the N−H stretch band, have been used extensively to analyze intramolecular hydrogen bonding in peptides, and related molecules.[7,18] Experimentally, Haque et al.[2] reported significant differences in the region of the N−H stretch band of the IR spectra between the two peptides. For $^{L}V^{D}P^{D}A^{L}L$ there was an intense peak at 3324 cm^{-1} and a weaker peak at 3420 cm^{-1}. For $^{L}V^{L}P^{L}A^{L}L$ the inverse was found, an intense peak at 3420 cm^{-1} and a much weaker peak at 3323 cm^{-1}. The peaks at ∼3324 were considered to correspond to N−H groups involved in internal hydrogen bonds. The peaks at ∼3420 were considered to correspond to N−H groups weakly hydrogen bonded to solvent. The corresponding spectra calculated from the trajectories are given in Figure 7. The relative IR intensity (absorbance) in the lower frequency N−H stretch band between $^{L}V^{D}P^{D}A^{L}L$ and $^{L}V^{L}P^{L}A^{L}L$ is qualitatively reproduced. There is however a shift in the position of the maximum for $^{L}V^{D}P^{D}A^{L}L$ and $^{L}V^{L}P^{L}A^{L}L$ of around 50 wavenumbers with respect to the experimental value (experimental maximum at ∼3324 cm^{-1}, calculated maximum at ∼3280 cm^{-1}). This shift is likely to arise from conducting the calculations of the IR spectra in a vacuum while the experimental measurements were done in methylene chloride. The effect of the solvent was thus neglected in the calculation of the IR frequencies. Note that the calculations do not reproduce the absorbance due to a free N−H group. Other factors which influenced the frequency calculation include the semiempirical Hamiltonian used (PM3) and the neglect of vibrational anharmonicity.

Despite some inaccuracies in the calculation of the vibrational frequencies it is clear that the IR spectra correlate with the markedly different structural preferences of these peptides. Conformational analysis of the peptide indicate that the highest relative absorbance values arise from structures representing clusters that contain configurations in a β-turn fold. A second maximum at higher frequency is observed experimentally. This peak was considered by Haque et al.[2] to correspond to a fully solvated N−H hydrogen bonded to solvent. As the calculations of the IR spectra were performed in a vacuum no second peak in this region is expected.

NMR Cross Relaxation Rates. Experimentally, NOESY data has supported the deductions about the presence of folded β-hairpin conformations in the tetrapeptides.[2] The experimental build-up curves of the cross-peaks for $^{L}V^{D}P^{D}A^{L}L$ taken from Figure 11 of Haque et al.[2] are reproduced in Figure 8a. For convenience the symbols used to identify specific proton pairs are the same as those used previously by Haque et al.[2] The cross-peaks indicated in Figure 8 include two long-range NOEs: (i) V1 H^{N}−L4 H^{N} (×) and (ii) an NOE between the C- and N-terminal methyl groups (○). Although the long-range signals are relatively weak, the data indicate that these cross-peaks do not result from transferred NOE effects. Two short-range NOEs, (i) ^{D}A3 H^{N}−L4 H^{N} (+) and (ii) a peak assigned to ^{D}A3 H^{N}−^{D}P2 $H^{\delta 2}$ (Δ), are also reported; however, the NOEs between protons on sequentially adjacent residues can arise from conformations other than the β-turn.[3] For comparison the relative NOE intensity as a function of mixing time for the same pairs of protons have been calculated from the trajectories as described by Peter et al.[19] In addition the intensity of the NOE to the alternative δ-methylene proton ^{D}A3 H^{N}−^{D}P2 $H^{\delta 3}$ (∇) was also calculated. Note, for short peptides such as $^{L}V^{D}P^{D}A^{L}L$ which display high internal mobility and rapid transitions between conformational states and for which there is coupling between rotational and internal motions, it is inappropriate to relate NOE intensities (cross relaxation rates) directly to average interproton distances simply using $\langle r^{-6}\rangle^{-1/6}$ or $\langle r^{-3}\rangle^{-1/3}$ averaging, r being the interproton distance.[20] Instead, it is necessary to estimate NOE intensities from the spectral densities of the interproton vectors and consider the full relaxation matrix as described in the methods. This, however, is only possible when the overall tumbling and internal motion of the peptide has been extensively sampled.

Figure 8b shows the NOE intensity as a function of mixing time averaged over the 10 $^{L}V^{D}P^{D}A^{L}L$ trajectories started from an extended conformation. Parts c and d of Figure 9 show the results from the two trajectories that exhibit the greatest deviations from the average, D1 and D6, respectively. For identifying the presence of the β-hairpin in the case of $^{L}V^{D}P^{D}A^{L}L$ the two long-range NOEs corresponding to V1 H^{N}−L4 H^{N} (×) and between the N and C terminal methylene groups (○) were considered to be of particular importance. Although these long-

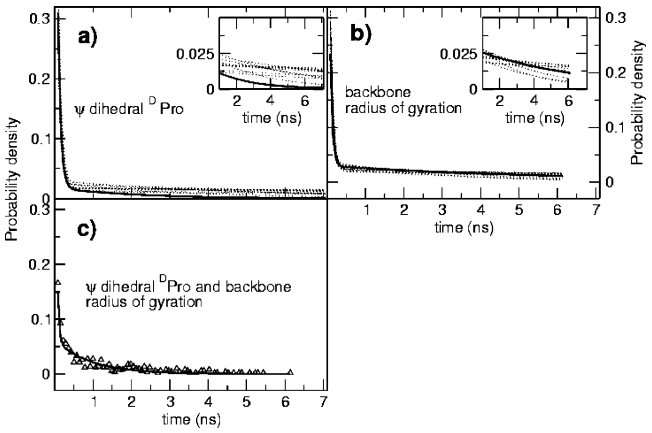

Figure 9. Decay curves (distribution of the residence time) of the folded state (i.e., how long the structure stays folded). The identification of the folded state was performed using two indicators: (a) a cutoff value of the ψ dihedral angle of D-Pro; (b) a cutoff value of the radius of gyrations. Results obtained for each of the individual trajectories are plotted with dotted lines while results obtained for the combined trajectory are plotted with thick solid lines. Panel c shows the decay curve using both indicators for the combined trajectory (fit shown with a solid line).

range interactions are relatively weak, the experimental data indicate that these are through space interactions and do not result from transferred NOE effects. Comparing Figure 8a to parts b−d of Figure 8, it can be seen that the intensities of these long range NOEs calculated from the trajectories closely match that observed experimentally. The primary discrepancy is that the intensity of the "○" cross-peak is slightly higher than observed experimentally. The variation between the trajectories is not significant. As can be seen the simulations also reproduce the relative intensity of the NOE ^{D}A3 H^{N}−L4 H^{N} (+) which experimentally is the most intense of the 4 cross-peaks.

Experimentally the intensity of the fourth NOE identified by Haque et al.[2] to be ^{D}A3 H^{N}−^{D}P2 H$^{\delta 2}$ (Δ) was between that of "+" and "○" or "×". In the simulations of ^{L}V^{D}P^{D}A^{L}L starting from an extended structure the intensity of the NOE ^{D}A3 H^{N}−^{D}P2 H$^{\delta 2}$ (Δ) was found to be very low becoming negative at long mixing times. However, the intensity of the NOE to the alternate methylene proton ^{D}A3 H^{N}−^{D}P2 H$^{\delta 3}$ (∇) was found to be comparable to "+". Overall there is a very good correspondence between the calculated NOE intensities as a function of mixing time and the experimental data of Haque et al.[2] The NOE intensities as a function of mixing time for the same set of proton pairs were also calculated for the 100 ns trajectory starting from a turn with ω of D-Pro in cis configuration (D-trapp) and are shown in Figure 8e. As discussed earlier, in this trajectory the peptide remained trapped in a turn conformation for 100 ns. It is evident from Figure 8e that the "Δ" and "∇" cross-peaks have almost zero intensity. The intensity of the "+" cross-peak is low compared to Figure 8a−d. The intensities of "○" and "×" are in contrast higher than that observed experimentally, or calculated from the simulations started from an extended structure. It is clear from these results that only the two long-range NOEs corresponding to V1 H^{N}−L4 H^{N} (×) and between the N and C terminal methylene groups (○) can be considered diagnostic of a β-hairpin and in fact the high relative intensity of the short range "Δ" and "+" cross-peaks observed experimentally is most likely indicative of a significant population on non β-hairpin conformations in solution. The fact that this particular trajectory samples a region of the conformational space not reached by the other 10 trajectories and that the NOE intensities exhibit a different trend as compared with the experimental ones or with those from the 10 trajectories agrees with the observation from experiment that the Proline exists mainly as the trans rotamer.

Figure 8f shows the NOE intensities as a function of mixing time for the same set of proton pairs calculated for the 100 ns trajectory for ^{L}V^{L}P^{L}A^{L}L. Experimentally, the long-range cross-peaks were not observed for ^{L}V^{L}P^{L}A^{L}L, while it did display an NOE cross-peak between the amide protons of the third and fourth residues, which is conventionally interpreted to indicate the presence of β-turn conformations. From the simulations, only the cross-peak of ^{D}A3 H^{N}−L4 H^{N} (+) shows any significant intensity. A low intensity peak was also reported experimentally, but this arises from other conformational states and by itself is not diagnostic of a hairpin.

Folding Rate. To describe the unfolding kinetics of peptide ^{L}V^{D}P^{D}A^{L}L, the distribution of the residence time (the decay curve) of the folded state is plotted in Figure 9. To calculate this distribution, two different indicators were used. Parts a and b of Figure 9 show the time distribution of the folded state defined according to the ψ dihedral of D-Pro and the radius of gyration (R_{g}yr) of the main chain atoms of the peptide for each trajectory and for the combined trajectory, respectively. Figure 9c displays the decay curve using both indicators for the combined trajectory. Neither the shape of the time distributions nor the trends in statistical behavior of the distributions are highly sensitive to the cutoffs chosen to define the folded conformation. The values of the cutoffs chosen correspond to the minimum of each indicator distribution function that divides the sampled space in two regions ($\psi_{cutoff} = -210°$, see Figure 2e; $R_{gyr_{cutoff}} = 0.35$ nm). By fitting the decay curves to an exponential form we find that they fit better to a biexponential function of the form $p(t) \sim \alpha e^{-k_{slow}t} + (1 - \alpha)e^{-k_{fast}t}$ than to a monoexponential function. This suggests that the folding process is governed by two different mechanisms. The values obtained for the rate constants indicate that the unfolding process of one of the mechanism is much faster than the other.

When the individual trajectories are fitted to a biexponential, the time constant of the fast decay is relatively well-defined. When using the ψ dihedral of D-Pro as the criterion for the folded state, the time constant of the fast decay for the individual trajectories is found to be in the range 1/83 ps $< k_{fast} <$ 1/58 ps. When using the radius of gyration criterion to determine the folded state, k_{fast} is calculated to be in the range 1/71 ps $< k_{fast} <$ 1/50 ps. When both indicators are applied to the combined trajectory, $k_{fast} \sim$ 1/37 ps. This fast decay may arise as consequence of fast fluctuations in the structure on a short time scale dictated primarily by the ψ dihedral transitions of D-Pro. Other theoretical calculations using coarse molecular dynamics on a dialanine peptide predict values for $k \sim {}^{1}/_{1000\ ps}$ for dihedral interconversions.[21]

The value of the slow rate constant of the folding/unfolding process, k_{slow}, obtained from each individual trajectory yield a broad range of values for k_{slow}. This suggests that the slow decay is subject to a more complex behavior and that the statistics from individual trajectories is not sufficient to properly sample this mechansim (see insets in Figure 9, parts a and b). Therefore, we use the results obtained from the combined trajectory. The value of the slow rate constant is, $k_{slow} \sim {}^{1}/_{2174\ ps}$ when the ψ dihedral of D-Pro is used as the criterion for the folded state, $k_{slow} \sim {}^{1}/_{1666\ ps}$ when the radius of gyration is used and $k_{slow} \sim {}^{1}/_{926ps}$ when both indicators are used. The results we obtained in this study in regard to the biexponential character of the folding rate are in line with previous observations obtained from stochastic dynamics simulations of alanine dipeptide with

implicit solvent. In those simulations, the dynamics (although on different time scales) is dominated by the barrier crossing between states defined by the rotation of backbone dihedral angle.[22] Our estimates for the rate constant are in a similar range of values of theoretical calculations of five-residue peptide loop-closure kinetics under diffusion-limited quenching.[23]

A rough estimate of the free energy difference, ΔG, between the unfolded and folded states (i.e., of the unfolding process) using the same cutoffs as for the previous calculations ($\Delta G = -RT \ln K_{eq}$, with $K_{eq} = \rho_{unfolded}/\rho_{folded}$) suggests its magnitude fall in the range of thermal fluctuations. For the combined trajectory when the folded/unfolded states are defined based on the ψ dihedral of D-Pro, $K_{eq} \sim 0.88$; based on the radius of gyration, $K_{eq} \sim 0.85$, and using both indicators, $K_{eq} \sim 1.0$. These values of the equilibrium constants yield a free energy difference between the unfolded and folded state in the range 0.0 and -0.4 kJ/mol. Note that the value close to 1 obtained for the equilibrium constant between the β-hairpin and coil conformations (for the $^{L}V^{D}P^{D}A^{L}L$) while the energy is more favorable for the β-hairpin state indicates that the conformational entropy of the unfolded state is larger. This is because in the folded (β-hairpin) state the conformational space is highly constrained while in the unfolded state (coil) the peptide samples much larger rotational states.

Conclusions

From independent MD simulations of the two tetrapeptide diastereomers in explicit solvent that cover more than 1 μs of combined time we are able to cast light on the factors that affect their folding propensity. As was found experimentally, we observe that the all-L peptide is primarily unstructured and hardly present as a β-hairpin conformer while the heterochiral isomer ($^{L}V^{D}P^{D}A^{L}L$) exists as an equilibrium between hairpin and extended conformations with approximately 50% of the time in each state. Thus, the incorporation of the two D-amino acids in the middle of the loop promote the formation of β-hairpin. The reason for the different behavior is a higher energy of the β-hairpin structure of the all-L peptide compare to the $^{L}V^{D}P^{D}A^{L}L$ stereoisomer due to hydrogen bond geometries that are not optimized and of none-bonded steric repulsions. The qualitative agreement achieved between simulations and experimental IR and NOE data gives confidence that the ensembles generated are highly representative of the peptide behavior in solution.

The observation that the trajectories that started from extended conformation only sampled configurations with D-Pro in a trans conformation indicates that the cis conformer has a higher free energy. Indeed, the cis conformer has a higher potential energy by about 14 kJ/mol than the trans isomer. Since both the cis and the trans folded states are expected to have the same entropy, the trans conformer has a lower free energy; thus, it is more stable.

The folding process of $^{L}V^{D}P^{D}A^{L}L$ is found to be best described by two rate constants, implying that it is goverened by at least two mechanisms. The first mechanism is fast, with a relaxation time of $\sim 4 \times 10^{1}$ ps. The second mechanism is slower, with a relaxation time of $\sim 1 \times 10^{3}$ ps. The wide distribution of the values obtained from the individual trajectories indicate the combined trajectories or long enough ones (~ 1 μs) are required to elucidate accurate kinetic properties for such molecules.

Acknowledgment. We thank Dr. S. Gellman and Dr. T. Haque for helpful discussions on the interpretation of NMR data, Dr. C. Peter for a helpful discussion on the calculation of NOE intensities, and Prof. Dr. A. E. Mark for general insightful discussions.

References and Notes

(1) Sibanda, B. L.; Thornton, J. M. *J. Mol. Biol.* **1993**, *229*, 428.

(2) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. *Biochim. Biophys. Acta* **1973**, *300*, 211.

(3) Haque, T. S.; Little, J. C.; Gellman, S. H. *J. Am. Chem. Soc.* **1996**, *118*, 6975.

(4) Santiveri, C. M.; Santoro, J.; Rico, M.; Jiménez, M. A. *Protein Sci.*. **2004**, *13*, 1134.

(5) Colombo, G.; De Mori, G. M. S.; Roccatano, D. *Protein Sci.* **2003**, *12*, 538.

(6) Santiveri, C.; Santoro, J.; Rico, M.; Jiménez, M. A. *J. Am. Chem. Soc.* **2002**, *124*, 14903.

(7) Merritt, E. A.; Bacon, D. J. *Meth. Enzymol.* **1997**, *277*, 505.

(8) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; vdf Hochschulverlag AG an der ETH Zürich and BIOMOS b.v.: Zürich, Switzerland, and Groningen, The Netherlands, 1996.

(9) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596.

(10) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(11) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

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

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

(14) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem. Int. Ed.* **1999**, *38*, 246.

(15) Stewart, J. J. P. *J. Comput. Chem*. **1989**, *10*, 209.

(16) Stewart, J. J. P. *MOPAC 2000 Manual*; Fujitsu Limited: Tokyo, Japan, 1999.

(17) Sibanda, B. L.; Thornton, J. M. *Nature (London)* **1985**, *316*, 170.

(18) Dado, G.; Gellman, S. H. *J. Am. Chem. Soc.* **1993**, *115*, 4228.

(19) Peter, C.; Daura, X.; van Gunsteren, W. F. *J. Biomol. NMR* **2001**, *20*, 297.

(20) Tropp, J. *J. Chem. Phys.* **1980**, *72*, 6035.

(21) Hummer, G.; Kevrekidis, I. G. *J. Chem. Phys.* **2003**, *118*, 10762.

(22) Smith, P. E.; Pettitt, B. M.; Karplus, M. *J. Phys. Chem.* **1993**, *97*, 6907.

(23) Yeh, I.; Hummer, G. *J. Am. Chem. Soc.* **2002**, *124*, 6563.