Computer Simulation Studies of Finite Temperature Conformational Equilibrium in Alanine-Based Peptides

S. Samuelson and G. J. Martyna*

Department of Chemistry, Indiana University, Bloomington, Indiana 47405-4001

Received: October 8, 1998; In Final Form: December 18, 1998

Conformational equilibria in two poly-alanine and one lysine-doped poly-alanine peptide, $[Ala_8]^+$, $[Ala_{16}]^+$, and $[Ac-(AAAAK)_3A-NH_2]^{3+}$, respectively, has been examined in both gas and water solution phase by computer simulation. The simulations studies were carried out using recently developed two-dimensional umbrella sampling techniques coupled to novel extended system multiple time step molecular dynamics methods. The efficiency of the latter method permitted run lengths greater than 30 ns to be used in the studies. The in vacuo simulation results are compared to recent novel gas-phase experimental studies employing drift tubes coupled to mass spectrometers (the injected-ion mobility/MS technique). Simulation studies of the peptides and neutralizing chloride anions solvated in computer water are compared to solution-phase NMR and ESR experimental work. Solvent effects are elucidated through a comparison of in vacuo and solution-phase simulation results. Comparisons between the results generated for the three different peptides are performed in order to assess the effect of the lysine side-chain and the position of the positive charge on peptide conformational equilibria in the two environments. The results indicate that the 3K peptide is roughly helical in a vacuum as well as in solution. The $[Ala_{16}]^+$ and $[Ala_8]^+$ form compact states in a vacuum in order to solvate the charge on the N-terminus. In water solution, the $[Ala_{16}]^+$ uncompacts but does not become strongly α-helical.

1. Introduction

It is important to study the conformational properties of small to intermediate-length peptides for several reasons. First, it is easy to examine particular problems in conformational equilibrium by designing well-thought-out model systems.[1,2] These model systems can be studied in fine detail both experimentally[1,2] and theoretically[3,4] and thus, can be used to gain insight into the underlying microscopic driving force on a particular problem in conformational equilibrium. Second, conformational equilibrium in peptides of moderate length may be related to larger protein structure. For example, the 3_{10}-helix, an intermediate along the helix-forming pathway, is known to occur at the C-termini of α-helical domains of large proteins.[5] In addition, proteins form well-defined secondary structural units before folding into their final states.[1,3,6] Finally, experiments on model systems have been performed in several different solvents[7−11] including in vacuo[12−14] and therefore a unique opportunity to compare theory to experiment under a diverse set of conditions exists.

Experimental and simulation studies have been performed on small de novo peptides in order to examine the relative populations of α-helices, 3_{10}-helices, random coil, reverse turns, and extended conformers as a function of polypeptide length and composition.[3,4,7−11,15] For example, the conformational equilibria between the α-helix and the 3_{10}-helix in the blocked Ala-3K and Ala-4K polypeptides, 16 and 21 residue units, respectively, with lysine residues inserted periodically has been investigated experimentally.[16,17] It was observed that the shorter 3K peptide favors an admixture of 3_{10} and α-type helices while the 4K peptide favors the α-helix conformation. Conversely, poly-aib peptides are known to form 3_{10}-helices.[18] However, alanine-doped poly-aib peptides make a transition from mixed 3_{10}-α-helical structures to α-helical structures as a function of alanine concentration.[19]

Determining the relative populations of pure α, 3_{10}-helices, and β structures along the folding pathway gives only a partial description of a peptide; conformational equilibrium. These structures represent a few of the many conformers that may form. Indeed, large conformational flexibility has been demonstrated to exist, experimentally. For example, Millhauser et al.[2] have reported the 16-residue alanine-based peptide denoted 3K(I) to exhibit broad ranges of amide I' adsorption in the infrared region.[9] The broad adsorption is due to the large number of unique conformers present at thermal equilibrium. Experimental evidence from X-ray crystallography has also implied significant thermal motion of protein structure in the solid state.[20] The results of this work have led the modern multiconformer interpretation of peptide equilibria; that is, many conformers are present and that considering a single average structure may be misleading.

In this paper, the structural properties of three alanine-based peptides are investigated in vacuo and in water solution at thermal equilibrium ($T = 300$ K) using novel simulation methodology. The new methodology combines extended system, multiple time step molecular dynamics techniques[21−23] with recently developed two-dimensional umbrella sampling tech-

* Author to whom correspondence should be addressed.

10.1021/jp983840x CCC: $18.00
Published on Web 02/24/1999

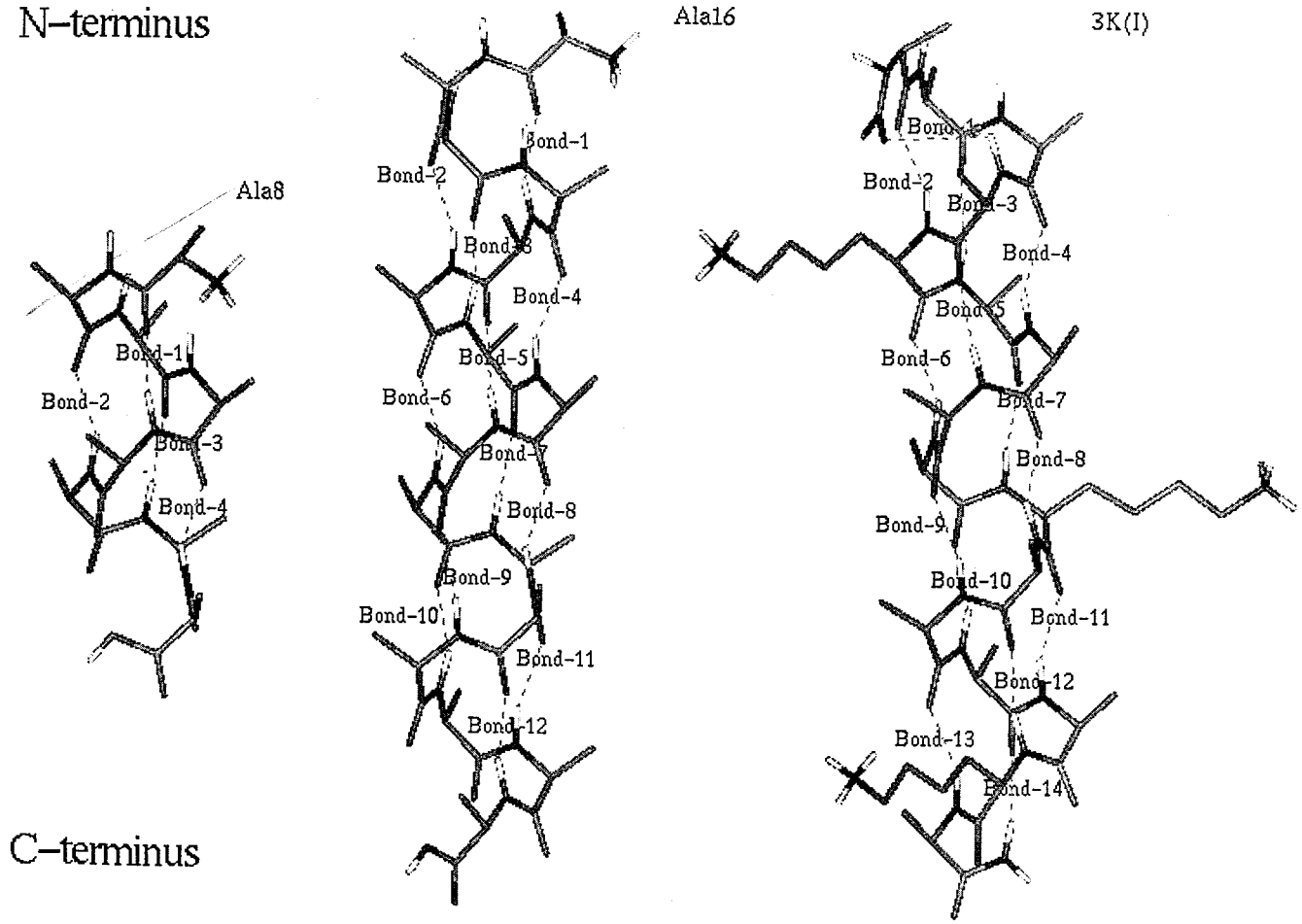

Figure 1. Representations of the $[Ala_8]^+$, $[Ala_{16}]^+$, and $3K(I)^{3+}$ peptides, in ideal α-helical geometries. The α-helical *(i*th*)* → *(i*th + 4) hydrogen bonds are both labeled and drawn with dotted lines. The molecules are depicted without aliphatic hydrogens but include all polar hydrogens. The N-terminus and C-terminus of the peptides are appropriately labeled.

niques to determine peptide configurational equilibrium at a higher level of detail than has previously been possible.[24] The results of both in vacuo and computer water solution simulation studies using runs of greater than 30 ns (due to the efficiency of the multiple time step integration) are compared to the corresponding experimental data, for the following alanine-based systems: $[Ala_8]^+$, $[Ala_{16}]^+$, and $[Ac-(AAAAK)_3A-NH_2]^{3+}$ (denoted 3K(I)). (The solution-phase NMR and ESR experimental data of[7−11,17] and the gas-phase injected-ion mobility/MS data of[13,25] were used.) The 3K(I) peptide carries its charge on the three lysine residues, while the two the polyalanines carry their single charge on the peptide N-terminus. Therefore, the effect of this difference as well as the presence of the lysine side-chains, in general, on peptide structural properties can be investigated both in vacuo and in computer water solution. Finally, the in vacuo results will be compared to those produced in computer water solution in order to examine solvent effects.

In detail, in vacuo simulations of all three alanine-based peptides are performed on two different empirical potential energy surfaces, AMBER95[26] and CHARMM22 at $T = 300$ K.[27] This comparison is interesting because previous studies on the small blocked tripeptide, $CH_3CO-(Ala)_3-NHCH_3$, in vacuo and solution,[28] have shown that there are some differences between the predictions of the two models. Computer water solution, simulation studies of $[Ala_{16}]^+$ plus a neutralizing chloride ion as well as $[Ac-(AAAAK)_3A-NH_2]^{3+}$ plus three neutralizing chloride ions are performed using the CHARMM22 force field[27] under conditions of constant temperature and pressure ($T = 300$ K, $P_{ext} = 1$ atm).

2. Methods

In this paper, a new two-dimensional umbrella sampling method[24] is used in conjunction with new extended system multiple time step molecular dynamics techniques[21−23] to determine the population of peptide conformers present at thermal equilibrium under two empirical force fields, CHARMM22[27] and AMBER95.[26] In particular, two poly-alanine peptides and one substituted poly-alanine peptide, $[Ala_8]^+$, $[Ala_{16}]^+$, and $[Ac-(AAAAK)_3A-NH_2]^{3+}$ (denoted 3K(I)), respectively, are examined in vacuo under constant temperature ($T = 300$ K) and in water solution at constant temperature and pressure ($T = 300$ K, $P_{ext} = 1$ atm). A brief description of force fields, equilibration procedures, and the methodology employed in the computer simulation studies are given below. The new techniques have been incorporated into the PINY-MD computational package which was used to perform the simulation studies.[29]

2.1. Umbrella Sampling. The well-known umbrella sampling method was designed to enhance configurational sampling.[30−32] In umbrella sampling, a series of simulations are performed using not the true direct potential energy function, $\phi(\mathbf{R}, V)$, but the true direct potential energy function plus a biasing function, $\tilde{\phi}(\mathbf{r}, \tilde{\mathbf{r}}_j)$. The biasing potential function is characterized by a set of parameters, $\tilde{\mathbf{r}}_j$, that serve to adjust the strength of the bias in

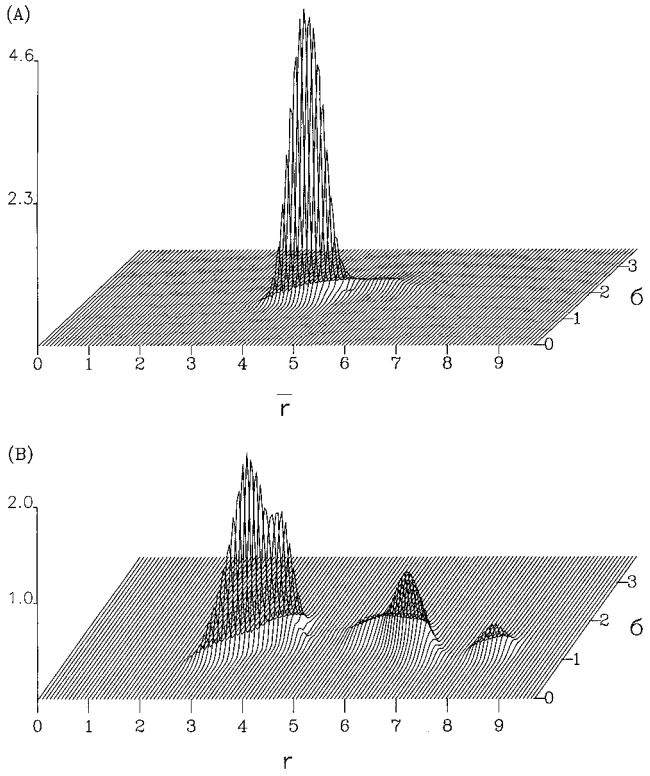

Figure 2. Comparison of the 2-D distribution function, $P(\bar{r}, \sigma)$, for the $3K(I)^{3+}$ peptide determined in vacuo at $T = 300$ K, under (a) the CHARMM22 force field and (b) the AMBER95 force field.

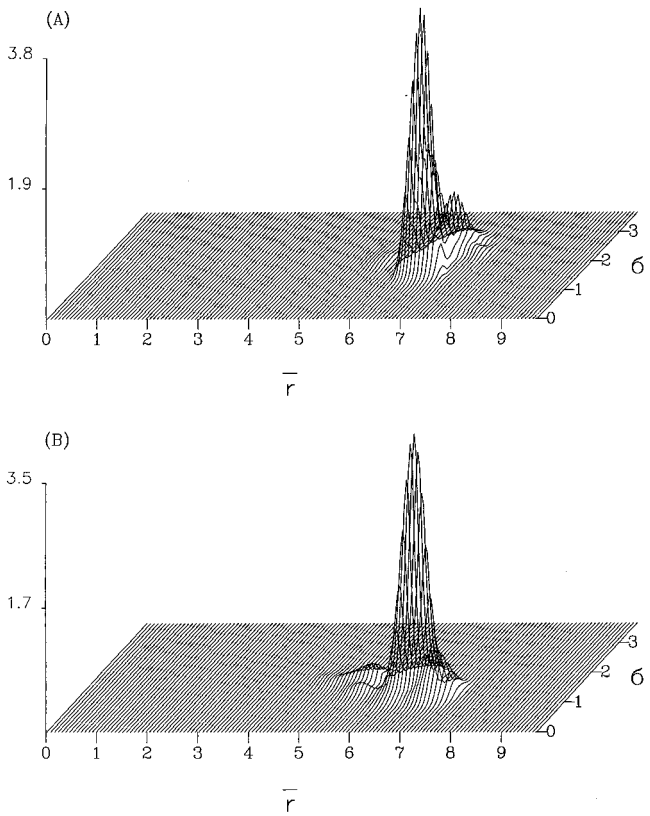

Figure 3. Comparison of the 2-D distribution function, $P(\bar{r}, \sigma)$, for the $[Ala_{16}]^+$ peptide determined in vacuo at $T = 300$ K under the (a) CHARMM22 force field and (b) AMBER95 force field.

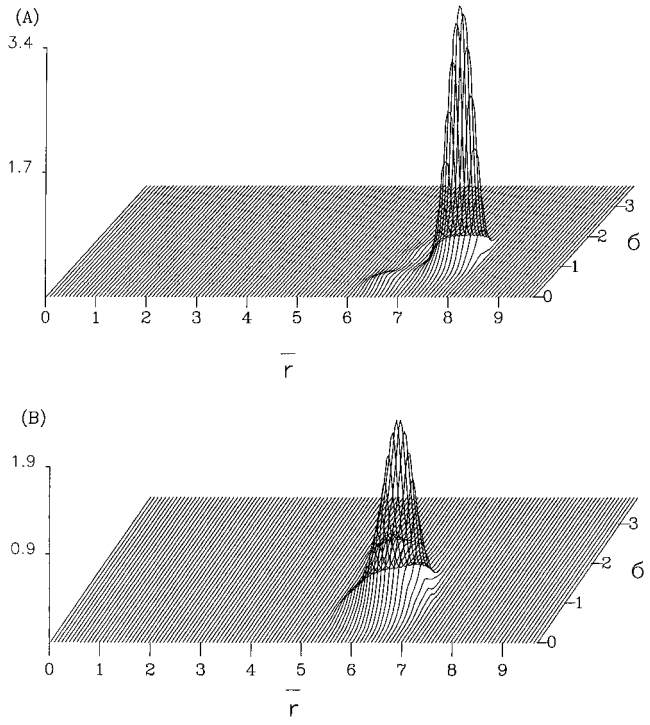

Figure 4. Comparison of the 2-D distribution function, $P(\bar{r}, \sigma)$, for the $[Ala_8]^+$ peptide determined in vacuo at $T = 300$ K under the (a) CHARMM22 force field and (b) AMBER95 force field.

the region of **r** space centered around the $\tilde{\mathbf{r}}_j$. The biasing potential function forces the dynamics to sample regions of configuration space, **r**, that may not otherwise be explored extensively. Thus, a series of calculations at selected $\tilde{\mathbf{r}}_j$ can be performed to "drag" the system through its configuration space "in the shade of the umbrella" formed by the biasing potential function. It is therefore possible to achieve large reductions of computational effort if the "umbrella sampling coordinate", the **r**, contains the rate-limiting pathway(s) through configuration space.

In helical peptides, unfolding/folding occurs when the distance between the amide hydrogen of the *i*th residue and the carbonyl oxygen of the $(i + 4)$th residue in the peptide increases/decreases. Furthermore, if these distances change in a non-uniform manner, the peptide may unfold in one region while remaining folded in another. Therefore, it is natural to define the new umbrella sampling biasing function[24] as follows:

$$\tilde{\phi}(\bar{r},\sigma,\tilde{\bar{r}},\tilde{\sigma}_k) = \frac{f_{\bar{r}}}{2}(\bar{r} - \tilde{\bar{r}}_j)^2 + \frac{f_\sigma}{2}(\sigma - \tilde{\sigma}_k)^2 \qquad (2.1)$$

where $f_{\bar{r}}$ and f_σ, are force constants and the coordinates $\bar{r},\sigma$ are defined to be

$$\bar{r} = \frac{1}{N_H}\sum_{i=1}^{N_H}|\mathbf{r}_i|$$

$$\sigma^2 = \frac{1}{N_H}\sum_{i=1}^{N_H}(|\mathbf{r}_i| - \bar{r})^2 \qquad (2.2)$$

Here, $|\mathbf{r}_i|$ is the hydrogen bonding distance associated with a helical turn, N_H is the number of α-helical hydrogen bonds considered (see Figure 1) and σ is the standard deviation of the specific distances chosen. In selecting hydrogen bonding distances to define the mean and the standard deviation, a constraint is introduced in the two-dimensional space. The

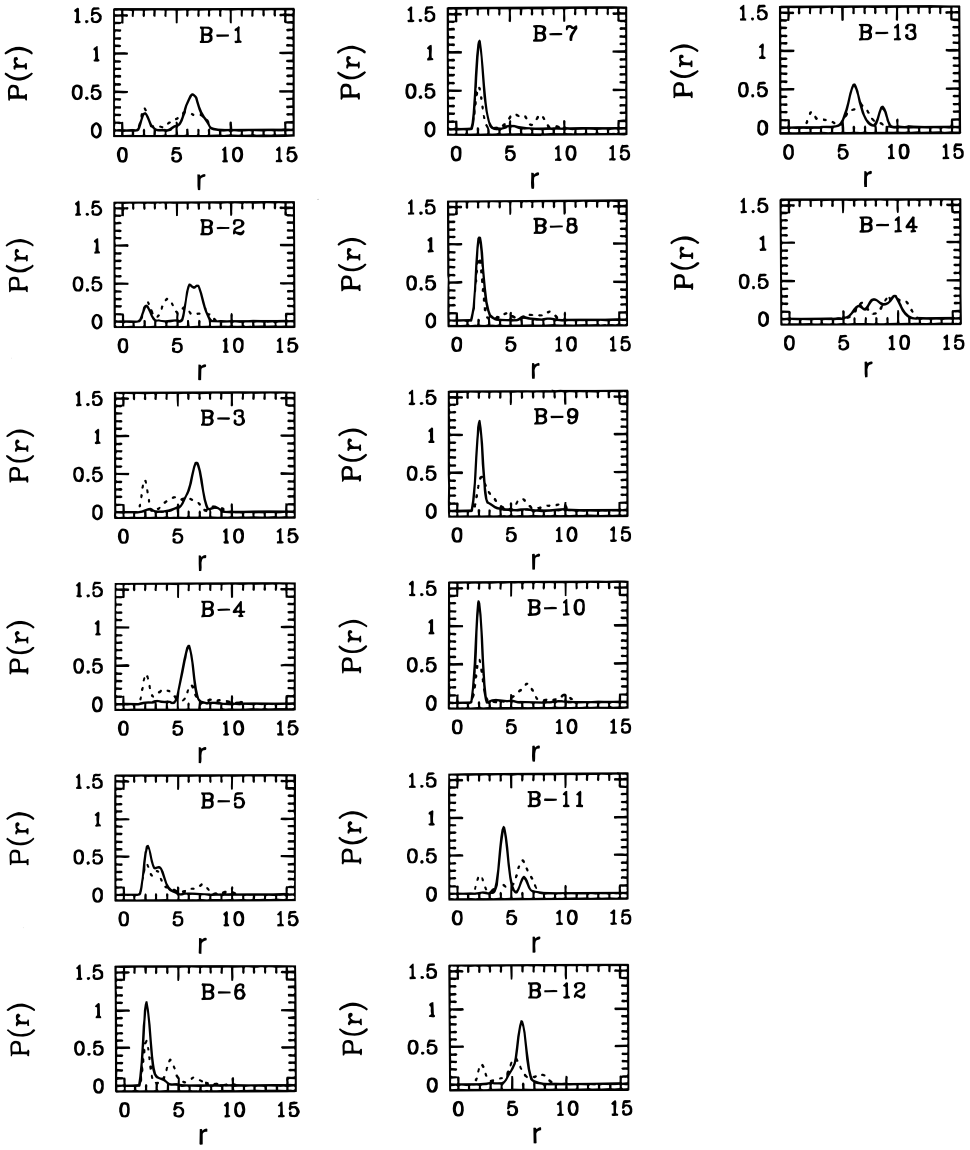

Figure 5. Comparison of the probability distribution functions for each of the 14 α-helical hydrogen bonding distances in the $3K(I)^{3+}$ peptide determined in vacuo at $T = 300$ K, under the CHARMM22 force field (solid 29 line) and the AMBER95 force field (dotted line).

standard deviation, σ, has the upper bound, $\sigma_{max} = \sqrt{N_H - 1}\,\bar{r}$, because $|\mathbf{r}_i| \geq 0$; that is, the standard deviation is maximal when $|\mathbf{r}_i| = N_H\bar{r}$ and the remaining $|\mathbf{r}_i| = 0$. (The constraint occurs naturally and need not be imposed.) This coordinate system is a generalization of the one-dimensional system presented in refs 4, 28. It was motivated by the X-ray and NMR experimental analysis tools[5] and a one-dimensional sampling scheme based on peptide radii of gyration.[33] (See ref 32 for a good review of umbrella sampling methods applied to biological systems.)

As stated above, in the mean-standard deviation coordinate system, various structural changes can be examined, naturally. When $\sigma \approx 0$, each hydrogen bonding distance, $|\mathbf{r}_i|$, along the peptide backbone will be approximately equal in value. Thus, when $\sigma \approx 0$ and $\bar{r}$ is large, the peptide will be uniformly unfolded whereas when $\sigma \approx 0$ and $\bar{r}$ is small, ideal helical structures are present. As the value of σ increases, any hydrogen bonding distance along the peptide backbone may increase/decrease with respect to any other hydrogen bonding distance. This enables the peptide to fold/unfold in a non-uniform manner. Extensive validation and testing of the new umbrella sampling coordinate system/method has been performed on the 3K peptide.[24] Briefly, in the test studies, four independent umbrella sampling calculations were performed using four different types of biasing potential. A careful comparison of the four independent calculations indicated that the methodology is capable of yielding semiquantitatively accurate results on intermediate-size peptide systems.

Finally, the umbrella sampling technique reduces the problem of configurational sampling to the task of finding an effective method for generating unbiased probability distribution functions from the biased simulation data. In these studies, the efficient and accurate weighted histogram method (WHAM)[34−37] is employed to perform this crucial task.

2.2. Molecular Dynamics Method. Gas-phase calculations of the isolated peptides, $3K(I)^{3+}$, $[Ala_{16}]^+$, and $[Ala_8]^+$, were performed under the conditions of NT ($T = 300$ K) using the Nosé−Hoover thermostat chain extended system canonical dynamics method[21,38−40] as implemented in the PINY-MD computational package.[29] (The nonstandard notation "NT" denotes that the "volume" is assumed to be sufficiently large compared to the single molecule studied that any average property of interest is volume independent.) A massive ther-

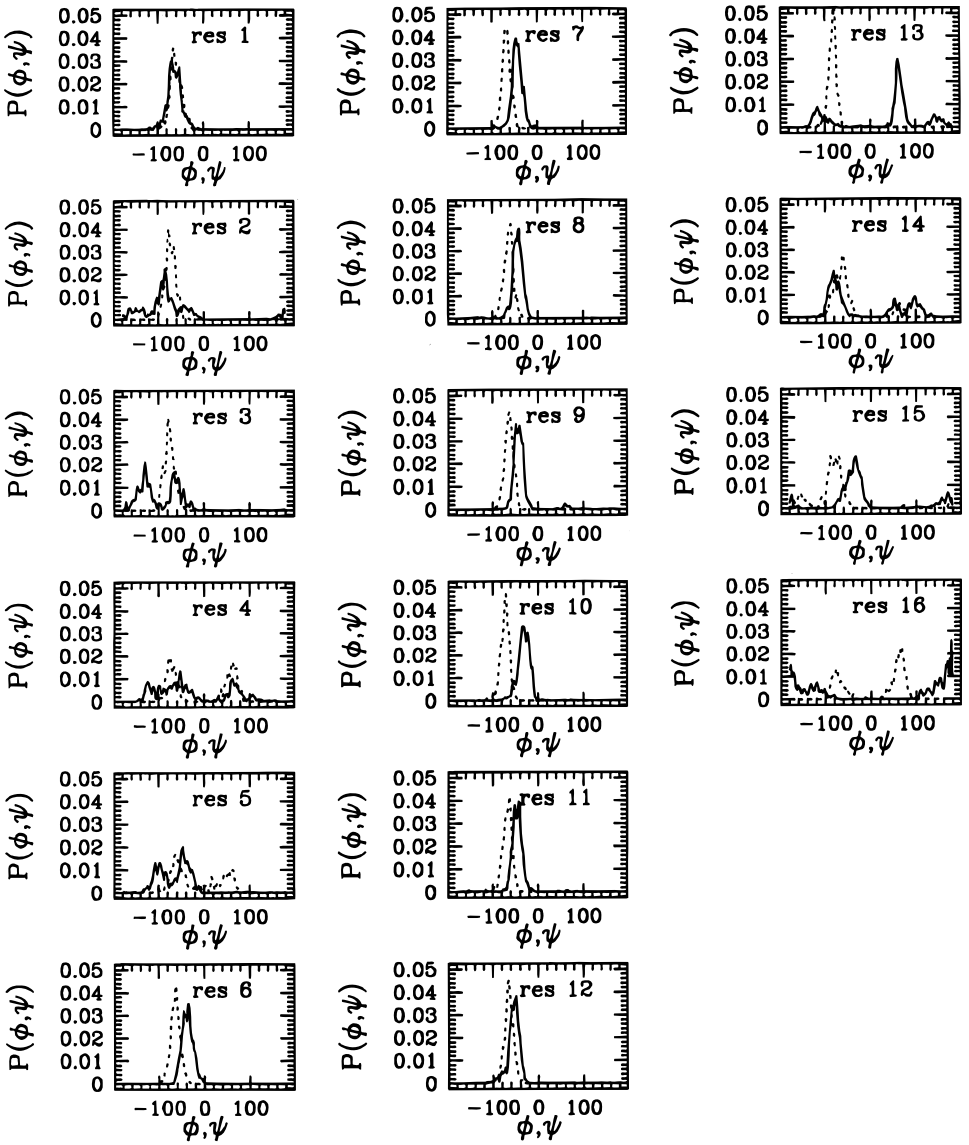

Figure 6. The probability distribution functions for each of the 16 backbone dihedral angles, ϕ (dotted line) and ψ (solid line), in the $3K(I)^{3+}$ peptide determined in vacuo at $T = 300$ K, under the CHARMM22 force field.

mostatting scheme was applied where one Nosé−Hoover thermostat chain is assigned to each degree of freedom in the molecule. A Nosé−Hoover thermostat chain length of $M = 2$ and a Nosé−Hoover thermostat frequency of $r = 50$ fs was employed. (The thermostat frequency determines the thermostat "mass".[21,22,41]) The masses of the hydrogen atoms were increased to that of a carbon atom as only the equilibrium properties of a classical system are considered.[42] These techniques were combined with a simple two-stage decomposition of the force into intramolecular and "intermolecular" components[21,22] which permitted a 3 fs time step to be used. The inner time step was taken to be $\delta t = 0.5$ fs. Here, the term "intermolecular" refers to the fact that atoms separated by more than four bonds interact via the intermolecular force field.

Condensed phase calculations of the $\{3K(I)^{3+}$ peptide $+ 3Cl^-\}$ system and the $\{[Ala]_{16}^+$ peptide $+ Cl^-\}$ system solvated in liquid water were performed under conditions of $NP_{ext}T$ ($T = 300$ K, $P_{ext} = 1$ atm) using an extended system isothermal−isobaric dynamics method[23,39,40] as implemented in the PINY MD computational package.[29] The recently developed reversible multiple time step algorithm[21,22] employed by this package, permitted a 4 fs time step to be utilized with massive Nosé−Hoover chain thermostatting and water molecules treated as rigid bodies using the SHAKE/ROLL and RATTLE/ROLL methods.[22] (The addition of constraints on the covalent N−H bonds of the peptides allowed a larger time step to be employed than in the gas phase.) The two-stage force decomposition was again utilized and masses of all the hydrogen atoms were increased to that of carbon. The inner time step was $\delta t = 1.0$ fs. Each Nosé−Hoover chain was given a length of $M = 2$ and a frequency of $r = 50$ fs. In addition, the volume frequency (used to define the volume "mass"[22,23]) was taken to be $r = 1000$ fs. Finally, the volume, itself, was assigned an independent Nosé−Hoover thermostat chain ($r = 1000$ fs) in order to facilitate the sampling of this important degree of freedom.

2.3. Force Fields. Two force fields were employed in the study, CHARMM22[27] and AMBER95.[26] (The most up-to-date parameter sets were down-loaded from the Web sites, "http : //www.pharmacy.umab.edu/˜alex/www.alex/param.html" and "http : //www.amber.ucsf.edu/amber/amber.htmr" respectively.) The TIP3P[3] water model was employed in the condensed phase calculations (charged peptides and chlorides solvated by

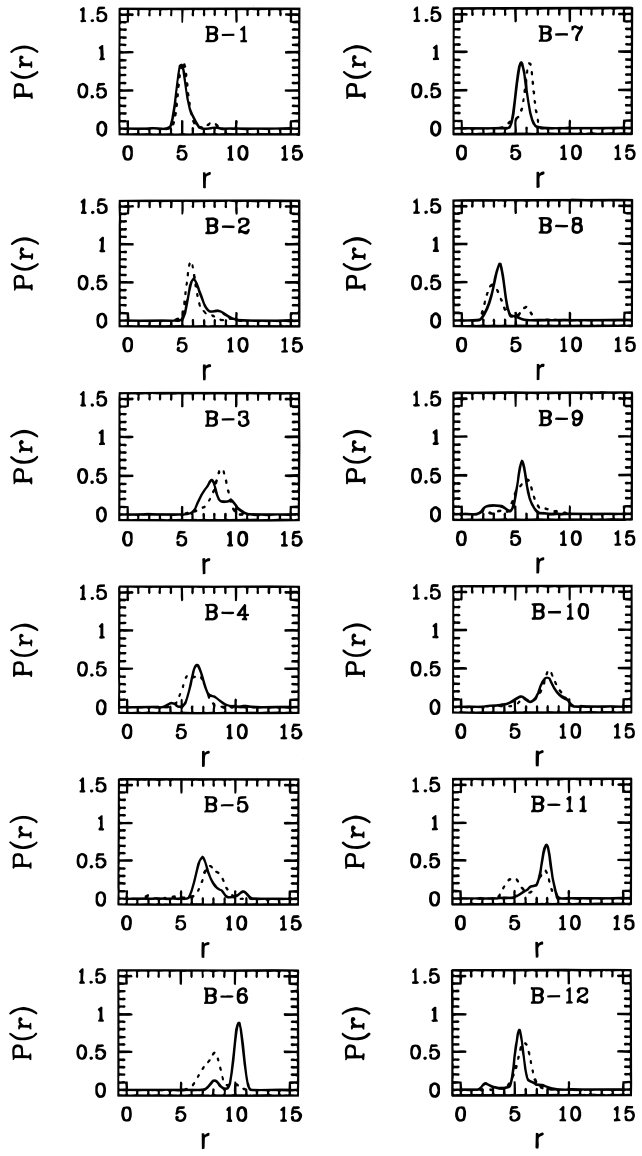

Figure 7. Comparison of probability distribution functions for each of the 12 α-helical hydrogen bonding distances in the $[Ala_{16}]^+$ peptide determined in vacuo at $T = 300$ K, under the CHARMM22 force field (solid line) and the AMBER95 force field (dotted line).

liquid water). Gas-phase calculations were conducted using "cluster boundary conditions" (no periodicity and no real space cutoff). In solution, three-dimensional periodic boundary conditions and Ewald summation were employed. A cutoff of 10 Å was used for the real space interactions.[31] The reciprocal space interactions were treated using efficient particle mesh Ewald techniques.[43]

2.4. Conformer Equilibration. In order to prepare initial conformers for the gas-phase simulation study of each peptide of interest, an ideal right-handed σ-helical configuration was first constructed. The ideal helical configuration was then used to generate equilibrated configurations spanning the full range of mean biasing parameter, $\tilde{r}$, with the standard deviation biasing parameter, $\tilde{\sigma}$, set to zero. This was accomplished by performing a sequence of simulations with an increasing value of $\tilde{r}$, each initiated from the last configuration of the previous run. Each equilibration run was carried out under massive NT dynamics for 450 ps. The resulting set of configurations, which now spans the full range of the mean biasing parameters, was used to initiate equilibration runs at the next larger value of the standard deviation biasing parameter. Again, massive NT dynamics simulations were run for 450 ps per biasing parameter set, $\{\tilde{r}_j, \tilde{\sigma}_k\}$. The procedure was continued until the full range of biasing parameters was spanned. (Twenty mean parameter values in the range 1.3 Å $< \tilde{r} <$ 9.3 Å and seven parameter standard deviation parameter values in the range 0 Å $< \tilde{\sigma} <$ 6 Å were used to create a total of 120 parameter sets.) Production runs of 450 ps per parameter set were initiated from the equilibrated configurations. A total simulation time of 60 ns was, thus, required to produce each surface.

In order to prepare solute phase conformers for both the $[3K(I)]^{3+}$ peptide and the $[Ala_{16}]^+$ peptide, the final configurations obtained in the corresponding in vacuo calculations were placed in a cubic box of side $L = 40$ Å containing 1700 model water molecules and overlapping molecules were removed until approximately 1600 remained. In the $[3K(I)]^{3+}$ peptide study, three of the water molecules were replaced with chloride ions while only one water molecule was replaced with a chloride ion in the $[Ala_{16}]^+$ peptide study. The water molecules removed were selected at random. In addition, the solvated ion's diffusion coefficient is sufficiently small that it is unlikely that the ions will have sampled their equilibrium distribution of position/orientations relative to the peptide during the course of a given simulation. The counterions are included to provide a realistic neutralizing background that is qualitatively similar to the experimental system.[2]

After solvating the peptides, equilibration runs under conditions of $NP_{ext}T$ ($T = 300$ K, $P_{ext} = 1$ atm) and isotropic cell fluctuations were performed for 60 ps using the massive Nosé−Hoover chain thermostat scheme. The production runs were carried out under the same conditions for 240 ps. The typical total simulation time for all the production runs used to generate a two-dimensional umbrella sampling surface is approximately 30 ns.

3. Results

In this section, the results of two-dimensional umbrella sampling computer simulation studies of the conformational equilibria of three alanine-based peptides in vacuo at constant temperature ($T = 300$ K) and in computer water solution at constant temperature and pressure, ($T = 300$ K, $P_{ext} = 1$ atm) are presented. In particular, the properties of three peptides are studied, the $3K(I)^{3+}$, $[Ala_{16}]^+$, and $[Ala_8]^+$ systems. In detail, the predictions of theCHARMM22 force field are compared to those of the AMBER95 for all three peptides in vacuo. In computer water solution, only the CHARMM22 model is examined and an appropriate number of neutralizing chloride anions were included in each study. The 3K(I) peptide carries its charge on the three lysines, while the polyalanines carry their single charge on the N-terminus. Therefore, the calculations will allow the effect of charge placement and the presence of lysine side-chains, in general, on conformational equilibrium to assessed. Comparisons of in vacuo simulations to computer water simulations will permit solvent effects to be investigated. The calculations were performed using the PINY MD computational package.[29] (Validation and testing of the new umbrella sampling/molecular dynamics scheme has been performed elsewhere.[24] Briefly, four independent umbrella sampling calculations were performed using four different types of biasing potential. A thorough comparison of the results indicates that technique is likely capable of yielding semiquantitatively accurate results on intermediate-size peptide systems.)

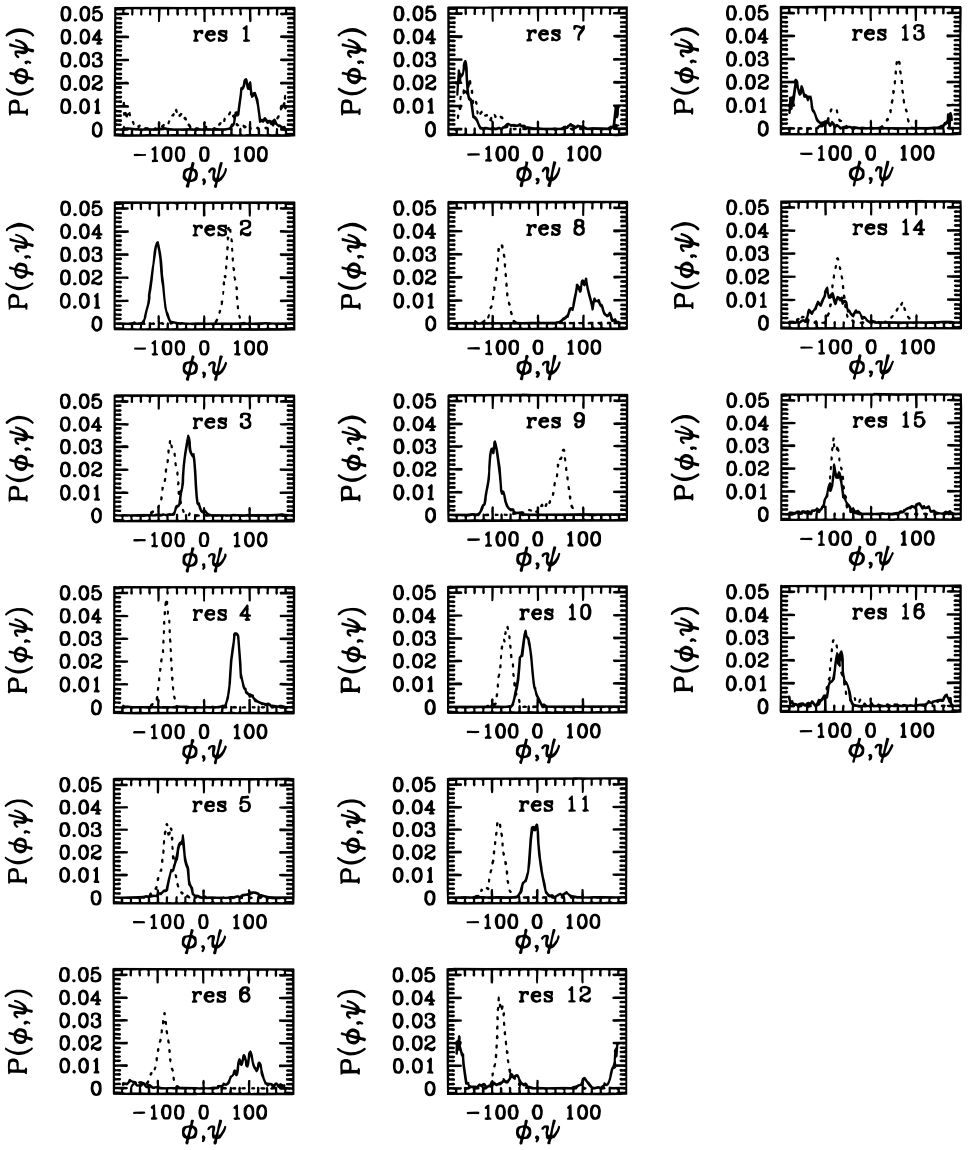

Figure 8. The probability distribution functions for each of the 16 backbone dihedral angles, ϕ (dotted line) and ψ (solid line), in the $[Ala_{16}]^+$ peptide determined in vacuo at $T = 300$ K, under the CHARMM22 force field.

3.1. Gas Phase: Two-Dimensional Probability Distribution Functions. Here, the results of umbrella sampling computer simulations of the three peptides of interest are presented in the form of two-dimensional probability distribution functions, $P(\bar{r}, \boldsymbol{\sigma})$, as a function of the mean and standard deviation of the α-helical hydrogen bonding distances ($r_{O_iH_{i+4}}$) along the peptide backbone (see section 2.1 and Figure 1). The positions of the peaks in the distribution function characterize the structure of a given peptide. If the standard deviation is small, $\boldsymbol{\sigma} \approx 0$, then perfect α-helices are present at small values of the mean ($\bar{r} \approx$ 2 Å), perfect 3_{10}-helices are present at intermediate values of the mean ($\bar{r} \approx 4$ Å), and $\boldsymbol{\beta}$ structures are present at large values of the mean. At large values of the standard deviation, more "disordered" conformers are present.

In Figure 2, parts a and b, the mean-standard deviation probability distribution functions, $P(\bar{r}, \sigma)$, obtained for the $3K(I)^{3+}$ peptide in vacuo under the CHARMM22 and AMBER95 force fields at $T = 300$ K are presented, respectively. The two distribution functions are similar and show that conformers with means in the range 5 Å $< \bar{r} <$ 9 Å are favored. The absence of probability near $\sigma \approx 0$ implies that $3K(I)^{3+}$ does not exhibit strongly ideal structures in vacuo (e.g., all residues α-helical or all residues β as described above). The breadth of the probability distribution function is consistent with the modern multiconformer interpretation of peptide equilibrium. The AMBER95 model generates some probability near $\bar{r} \approx 3$ Å, indicating that it supports more α-helical hydrogen bonds than the CHARMM22 model. These differences in helical propensity between the two force fields are consistent with the results of a previous study on a blocked tri-alanine peptide.[28]

The mean-standard deviation probability distribution functions obtained for the $[Ala_{16}]^+$ peptide in vacuo under the CHARMM22 and AMBER95 force fields are given in Figure 3, parts a and b, respectively. Unlike the $3K(I)^{3+}$ peptide system where a significant portion of the probability in the mean-standard deviation distribution function was found in the region, $\bar{r} < 6$ Å, for the $[Ala_{16}]^+$ peptide system, the probability begins at grow in at $\bar{r} \approx 7$ Å. This indicates that $[Ala_{16}]^+$ has a low α-helical content in vacuo and may have compacted/folded in order to "solvate" the charge on the N-terminus. The AMBER95 force field yields results that are similar to the CHARMM22 force field.

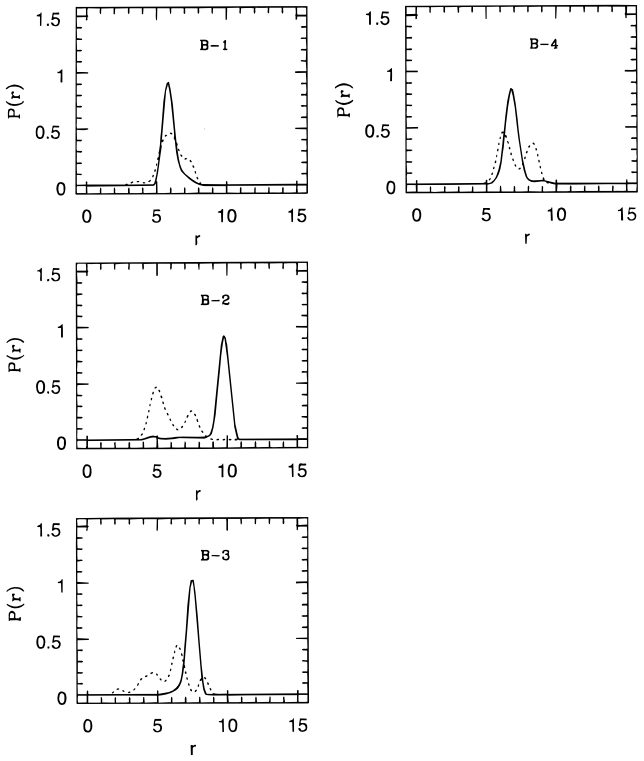

Figure 9. Comparison of the probability distribution functions for each of the four α-helical hydrogen bonding distances in the $[Ala_8]^+$ peptide determined in vacuo at $T = 300$ K, under the CHARMM22 force field (solid line) and the AMBER95 force field (dotted line).

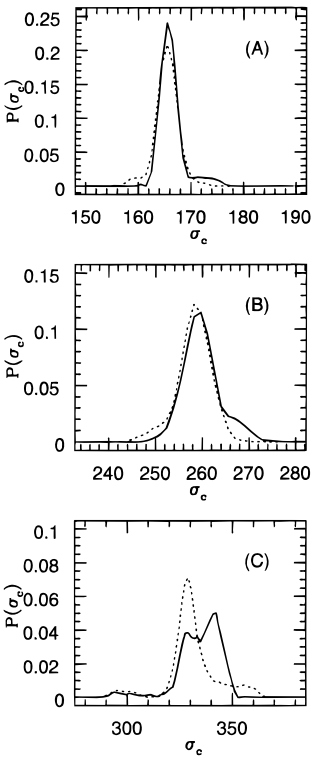

Figure 10. Comparison of the probability distribution functions of the hard sphere collision cross section, σ_c, for (a) the $[Ala_8]^+$ peptide, (b) the $[Ala_{16}]^+$ peptide, and (c) the $3K(I)^{3+}$ peptide determined at $T = 300$ K in vacuo under the CHARMM22 force field (solid line) and the AMBER95 force field (dotted line).

In Figure 4, parts a and b, the mean-standard deviation probability distribution functions, $P(\bar{r}, \sigma)$, for the $[Ala_8]^+$ peptide in vacuo under the CHARMM22 and AMBER95 force fields

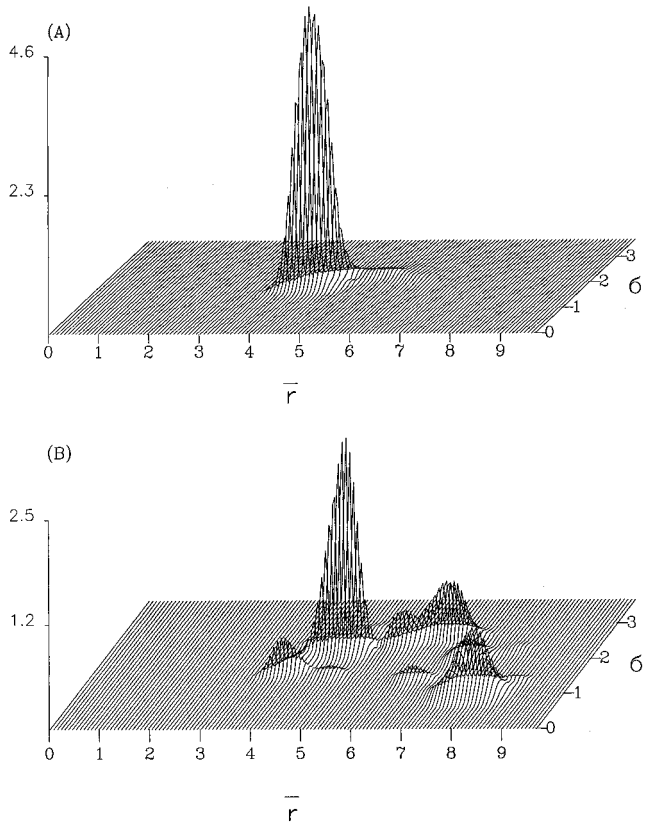

Figure 11. Comparison of the 2-D distribution function, $P(\bar{r}, \sigma)$, for the $3K(I)^{3+}$ peptide determined under the CHARMM22 force field (a) in vacuo at $T = 300$ K, and (b) in water solution with neutralizing chloride anions at $T = 300$ K and $P_{ext} = 1$ atm.

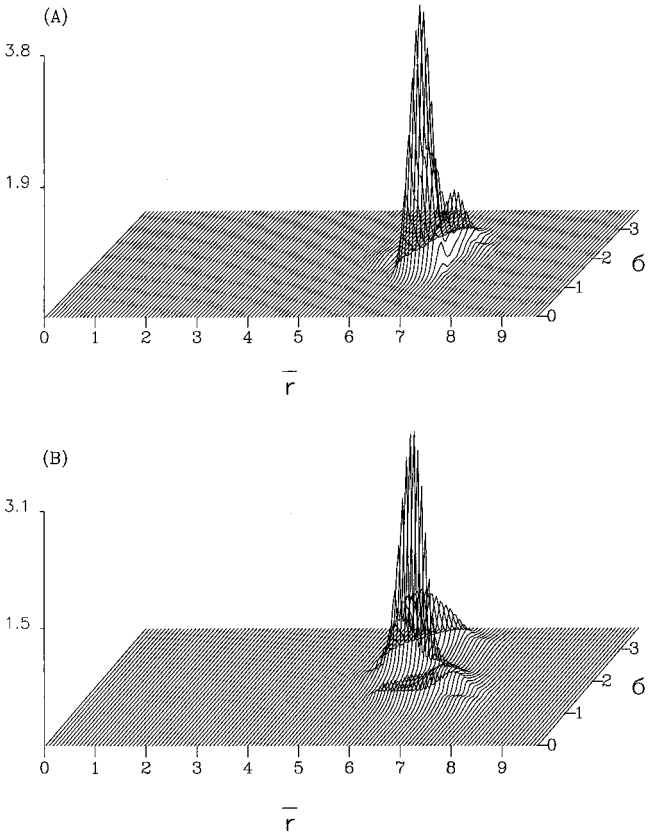

Figure 12. Comparison of the 2-D distribution function, $P(\bar{r}, \sigma)$, for the $[Ala_{16}]^+$ peptide determined under the CHARMM22 force field, (a) in vacuo at $T = 300$ K and (b) in water solution with neutralizing chloride anions at $T = 300$ K and $P_{ext} = 1$ atm.

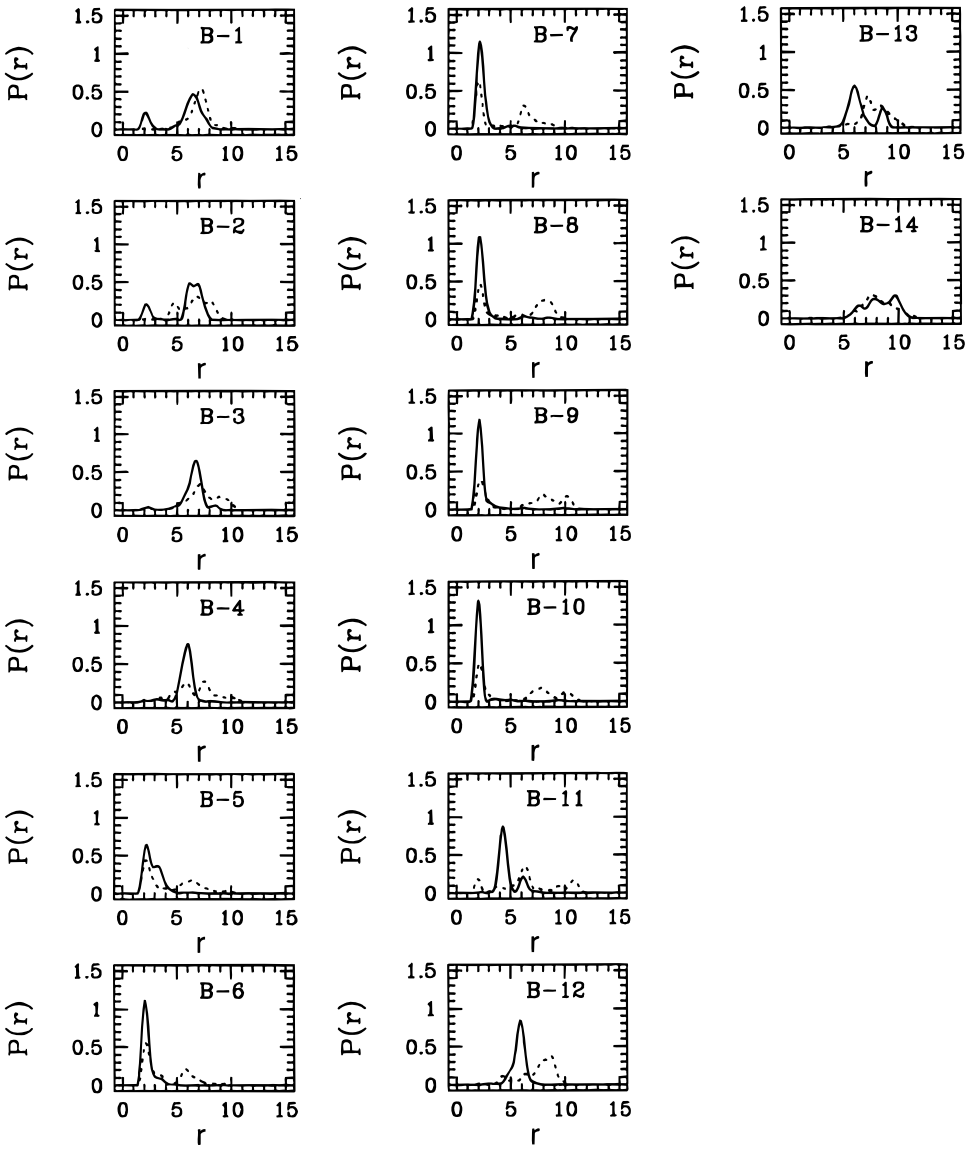

Figure 13. Comparison of the probability distribution functions for each of the 14 α-helical hydrogen bonding distances in the $3K(I)^{3+}$ peptide under the CHARMM22 force field in vacuo at $T = 300$ K (solid line) and in water solution with neutralizing chloride anions at $T = 300$ K and $P_{ext} = 1$ atm (dotted line).

are given, respectively. The two force fields, CHARMM22 and AMBER95, predict that extended conformational states are populated and that this peptide may also have compacted/folded in order to "solvate" the charge on the N-terminus.

3.2. Gas Phase: Conformer Analysis. The mean-standard deviation probability distribution functions, $P(r, \boldsymbol{\sigma})$, presented above for the three peptides of interest yield general information about the peptide conformers populated at thermal equilibrium. More detailed information about the structures can be obtained by examining the α-helical hydrogen bonding distance ($r_{O_iH_{i+4}}$) and the backbone dihedral angle (ϕ and ψ) distribution functions for each individual residue along the polypeptide chain.

In Figure 5, the 14 possible α-helical hydrogen bond distance distribution functions of the $3K(I)^{3+}$ peptide (see Figure 1), determined in vacuo at $T = 300$ K, are presented. The results produced using the CHARMM22 model are represented by the solid lines, and the results produced using the AMBER95 model are given by the dotted line. The predictions of the two force fields are similar. This is consistent with conclusions drawn from the two-dimensional mean-standard deviation distribution functions presented above. In general, the $3K(I)^{3+}$ peptide is α-helical in the middle (bonds 5−10) with a mixture of folded and unfolded states at the peptide termini. AMBER95 predicts a larger α-helical content at the termini and larger 3_{10} population in the middle. The conformer populations can be seen more clearly in Figure 6, where the backbone dihedral angles distribution functions, (ϕ dotted line and ψ solid line) are shown for the CHARMM22 force field, only. Note, in a "perfect" α-helix, the dihedral angles taken on the values $\{\phi = -60°, \psi = -47°$.[5] In addition, a conformer taken from the largest peak in the mean-standard deviation distribution function determined under CHARMM22 is shown in Figure 15a.

The 12 α-helical hydrogen bond distance distribution functions of the $[Ala_{16}]^+$ peptide (see Figure 1) determined under the CHARMM22 and AMBER95 force fields in vacuo at $T = 300$ K are given in Figure 7. It is evident that the peptide has folded into a compact state and is much less helical in vacuo than the $3K(I)^{3+}$ peptide. This is in agreement with the conclusion reached from the mean-standard deviation distribution functions presented above. The backbone dihedral angle

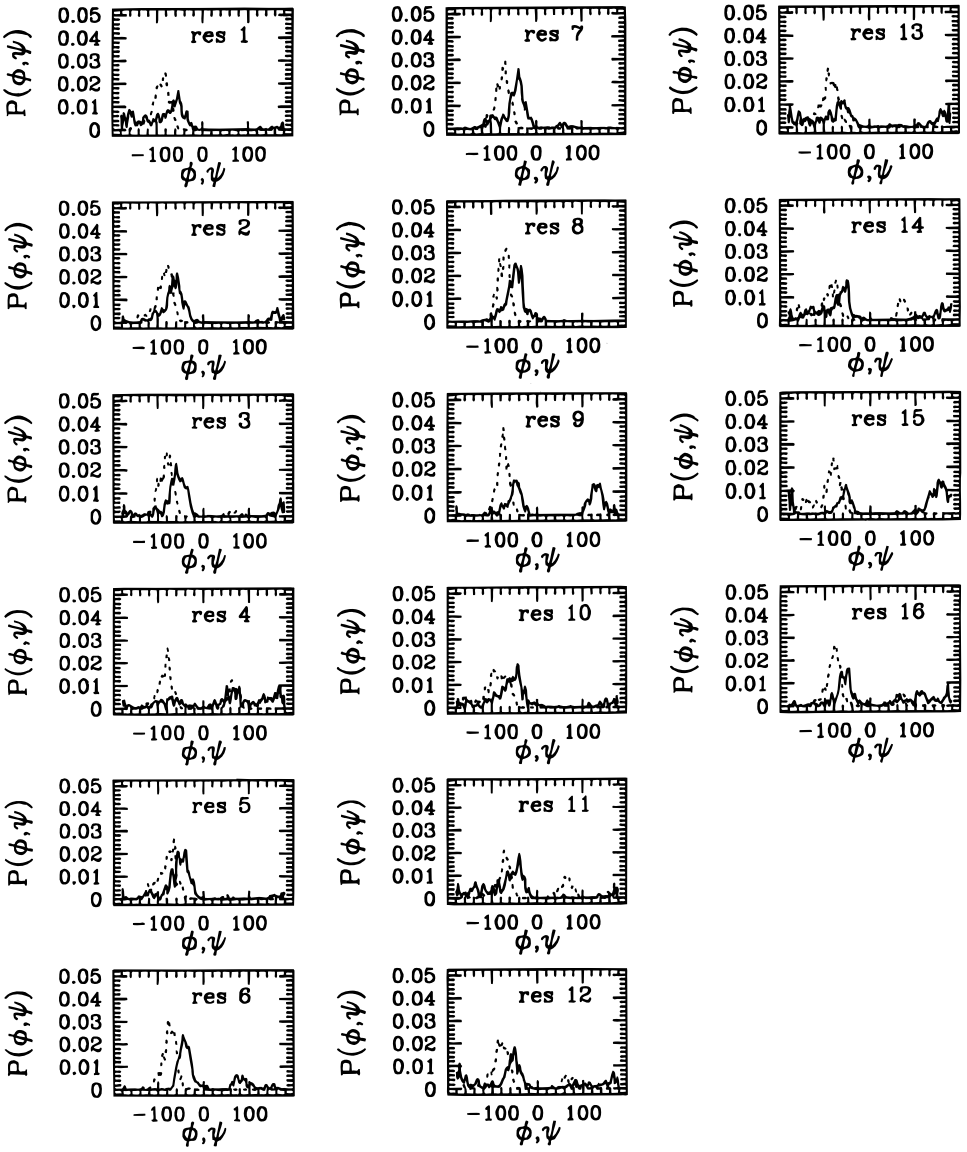

Figure 14. The probability distribution functions for each of the 16 backbone dihedral angles, ϕ (dotted line) and ψ (solid line), in the $3K(I)^{3+}$ peptide determined under the CHARMM22 force field in water solution with neutralizing chloride anions at $T = 300$ K and $P_{ext} = 1$ atm.

distribution functions (ϕ dotted line and ψ solid line) determined under the CHARMM22 force field, only, presented in Figure 8, underscore the lack of helicity in the peptide. In addition, a conformer taken from the peak in the mean-standard deviation distribution function (Figure 18b), clearly, is in a compact state, with only some small helical content present. Note, the distribution of compact conformers characterized by Figures 4, 7, and 8 was generated from a purely α-helical initial condition (see section 2.4). Therefore, the umbrella sampling/molecular dynamics method employed in the calculations seems to be sampling configuration space reasonably well (see also ref 24). However, in systems with rough energy landscapes,[6] it is best to take a conservative view.

In Figure 9, the four α-helical hydrogen bond distance probability distribution functions of the $[Ala_8]^+$ peptide (see Figure 1) are presented. These distribution functions are similar to those found in the $[Ala_{16}]^+$ peptide. None of the bonds remain predominately helical.

3.3. Gas Phase: Collision Cross Sections. Novel experimental techniques have been developed to investigate biomolecular structure in the gas phase.[13,14,25] The injected-ion mobility/MS method has recently been employed to examine the biomolecular structure of the $3KG^{3+}$ peptide (the $3K(I)^{3+}$ peptide with one alanine replaced by glycine), the $[Ala_{16}]^+$ peptide, and the $[Ala_8]^+$ peptide. In this experiment, a biomolecule is subjected to electrospray ionization (ESI), filtered through a mass spectrometer, and then propelled into a drift tube filled with a buffer of inert gas (helium). In the presence of a weak electric field, the biomolecular ion moves through the buffer gas in a finite amount of time that is related directly to its collision cross section. The time scale of the measurement dictates that a single average cross section is determined for each ion, rather than a spectrum of values.

In Figure 10, probability distribution functions of hard sphere collision cross sections, $P(\sigma_c)$, determined under the CHARMM22 and AMBER95 force fields at $T = 300$ K are presented for each of the three peptides studied (Note, σ_c denotes the hard sphere cross section and does **not** refer to a standard deviation of hydrogen bond distances). The helium-"X" hard sphere radii used in the cross section calculations were taken from ref 44. In all cases, the solid line represents the CHARMM22 result and the dotted line represents the AMBER95 result.

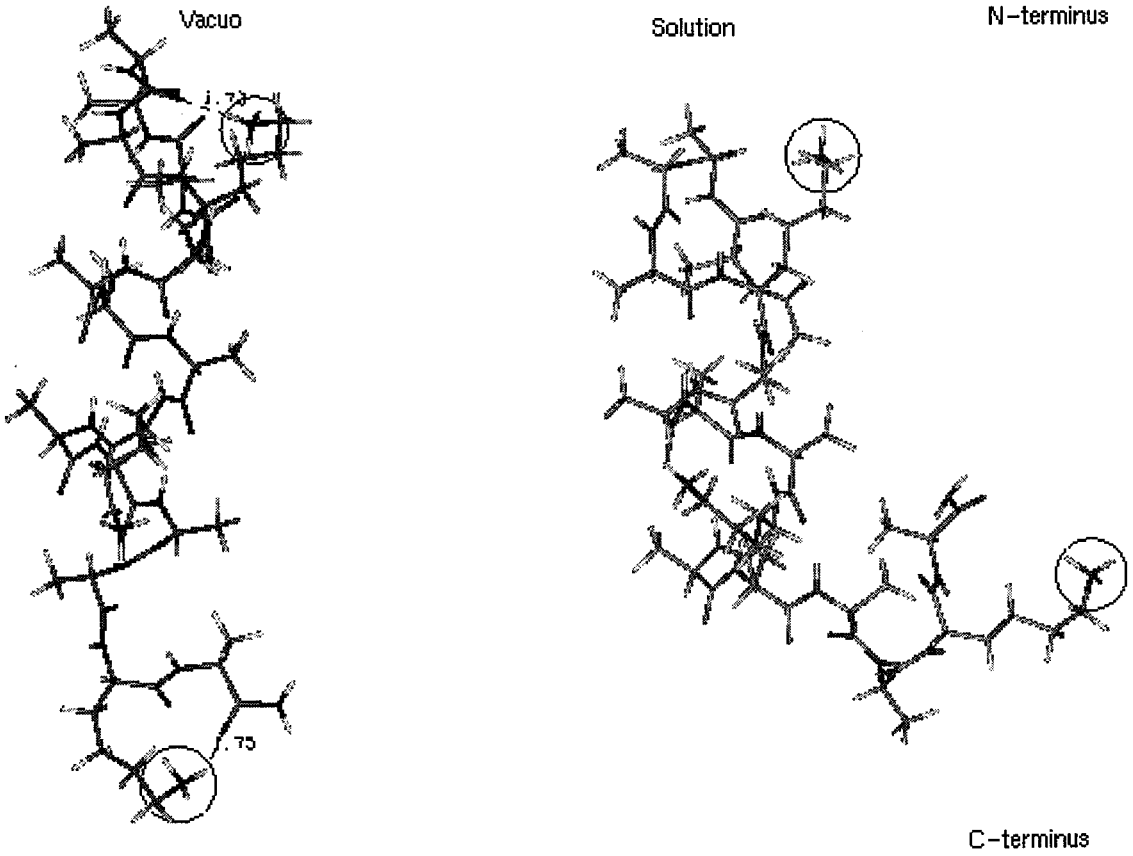

Figure 15. Instantaneous configurations of the $3K(I)^{3+}$ peptide obtained from simulations (a) in vacuo, and (b) in water solution, under the CHARMM22 force field. Both configurations were taken from simulations with the umbrella sampling biasing parameters set to $\tilde{r} = 3$ Å, $\tilde{\sigma} = 2$ Å.

The agreement between the models is good. The experimentally measured collision cross sections for the three peptides studied are $\sigma_c([Ala_8]^+) = 157$ Å^2, $\sigma_c([Ala_{16}]^+) = 257$ Å^2, and $\sigma_c([3KG]^{3+}) = 332$ Å^2. The average theoretical cross sections agree with the experimental results to within 5% for the small peptide and within 3% for the two larger peptides.

3.4. Solution Phase: Two-Dimensional Probability Distribution Functions. The results of constant pressure and temperature umbrella sampling calculations on the $\{3K(I)^{3+}$ peptide + $3Cl^-\}$ system as well as the $\{[Ala_{16}]^+$ peptide + $Cl^-\}$ system solvated in computer water solution under the CHARMM22 force field at the state point, $\{T = 300$ K, $P_{ext} = 1$ atm$\}$, are presented in this subsection. Comparisons to the corresponding in vacuo calculations will serve to illustrate solvent effects on the two peptides.

The mean-standard deviation probability distribution function, $P(\bar{r}, \sigma)$, for the $3K(I)^{3+}$ peptide in vacuo and water solution are both presented in Figure 11. In general, the two distribution functions have peaks in the same regions of mean-standard deviation configuration space, $\{\bar{r}, \sigma\}$. The primary difference between the in vacuo and solution-phase distribution functions is the increase in the extended state population. This indicates that a significant number of conformers in water solution have reduced α-helical content compared to in vacuo. Therefore, the middle of the peptide likely has a higher 3_{10}-helical content in computer water solution than in vacuo.

In Figure 12, the mean-standard deviation probability distribution function, $P(\bar{r}, \sigma)$, for the $[Ala_{16}]^+$ peptide both in vacuo and in water solution are given. Here, the peak in the distribution moves to smaller values of the mean. This indicates that the computer water has broken apart the more compact structures by solvating or partially solvating the charge on the N-terminus and thus increased peptide helicity. (Note, the umbrella sampling method employed in the calculations guarantees that strongly helical structures are indeed sampled, 2 Å $< \bar{r} <$ 4 Å. However, the WHAM unbiasing procedure indicates that these are not populated at thermal equilibrium.)

3.5. Solution Phase: Conformer Analysis. The probability distribution functions for each of the fourteen α-helical hydrogen bond distances in the $3K(I)^{3+}$ peptide (see Figure 1) are presented for the gas and solution phase in Figure 13. In both cases, the middle bonds (bond-5 thru bond-10) remained predominately α-helical (i.e., the distribution function is peaked around $r \approx 2$ Å). However, in solution, the distribution functions show a significantly higher fraction of 3_{10}-helix and extended conformer are populated compared to in vacuo. The backbone dihedral angle distribution functions (ϕ dotted line and ψ solid line), depicted in Figure 14, also show the loss of α-helical structure. (The corresponding gas-phase distribution functions are given Figures 5 and 6.) In addition, an instantaneous configuration of the $3K(I)^{3+}$ peptide taken from the peak of the mean standard deviation probability distribution function in gas and solution phases, respectively, is presented in Figure 15 parts a and b. In total, these results support the conclusions drawn from the mean standard-deviation probability distribution functions that computer water solvent has decreased the helical content of the peptide. This is consistent with the results of a study on a blocked tri-alanine peptide.[28]

The experimental studies of Millhauser et al on the $3K(I)^{3+}$ peptide in water solution[7−11,17] indicate that the first two and

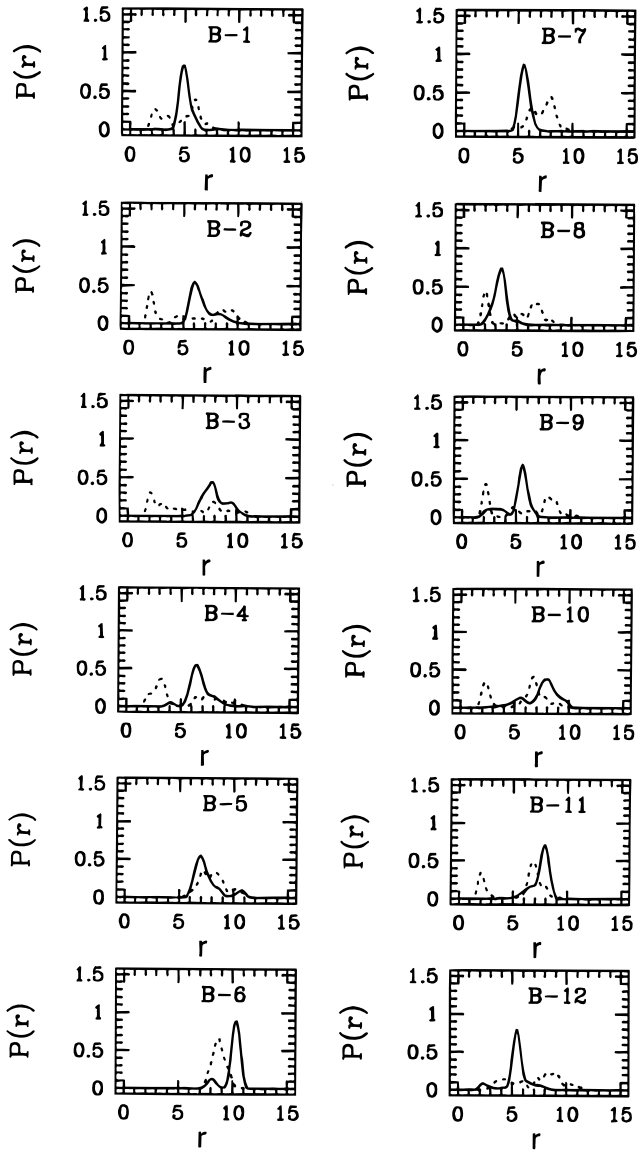

Figure 16. Comparison of the probability distribution functions for each of the 14 α-helical hydrogen bonding distances in the $[Ala_{16}^{+}]$ peptide under the CHARMM22 force field in vacuo at $T = 300$ K (solid line) and in water solution with neutralizing chloride anions at $T = 300$ K and $P_{ext} = 1$ atm (dotted line).

last three residues in the peptide show large deviations from helicity. The results presented here predict a somewhat more "floppy" peptide. (Again, the umbrella sampling method guarantees that strongly helical structures are sampled. However, the unbiasing procedure indicates that strongly helical conformers are not populated at equilibrium.) The average value of the ϕ angle for the middle, helical residues (labeled 5−9) is approximately, $\langle\phi\rangle = -70°$, in agreement with the experimental data. Note, this value is slightly lower than the "perfect" α-helical value $\phi = -60°$.[5] In general, the broad distribution of $\{\phi, \psi\}$ values found in the end residues (see Figure 14) make the nature of the conformers difficult to characterize. For example, both the ϕ and ψ dihedral angles in residue 15 have two peaks. However, the experiments were performed at T = 273 K which likely increases the helix content (from T = 300 K, the simulation temperature).

In Figure 16, the probability distribution function for each of the twelve α-helical hydrogen bond distances in the $[Ala_{16}]^{+}$ peptide (see Figure 1) is presented for both gas and solution phase simulations. In solution, all bonds have significantly higher probabilities near $r \approx 2$ Å than in vacuo indicating a greater helix content. The backbone dihedral angle distribution functions, (ϕ dotted line and ψ solid line), shown in Figure 17, also indicate an increased helix content. An instantaneous configuration of the $[Ala_{16}]^{+}$ peptide taken from the the peak of the mean standard-deviation probability distribution function in solution and gas phases, respectively, are presented in Figure 18, parts a and b. Again, the detailed information supports the conclusion drawn from the mean-standard deviation probability distribution functions, that solvent has opened the compact structure and restored some helicity. However, the peptide is far from a perfect α-helix. Note, initial conformers were generated by solvating the compact gas phase configurations (see section 2.4). Again, the umbrella sampling/molecular dynamics scheme is designed to ensure that helical conformers are sampled. However, these are less favorable then the more complex structures.

4. Discussion

The $3K(I)^{3+}$ peptide samples a broad set of fairly helical conformers with "frayed"/disordered termini both in water solution and in the gas phase. However, peptide helicity is decreased in solution where the computer water molecules successfully compete for hydrogen bonding sites on the peptide backbone. In vacuo, the lysine side-chains which carry the excess charge are "solvated" by the carbonyl oxygens of the peptide backbone and, thus, cause some disruption of the helical structure (see Figure 15). In computer water solution, however, these charges are solvated by the water molecules and the side-chains no longer "hug" the peptide backbone (see Figure 15). In order to measure quantitatively the solvent effect on intramolecular hydrogen bonding between the hydrogen atoms of the NH_3^{+} groups on the lysine side-chains and backbone carbonyl oxygens, the corresponding oxygen−hydrogen probability distribution function is presented in Figure 19 for both phases. The intramolecular hydrogen-bonding peak present in vacuo is missing in water solution, as might be expected from an examination of Figure 15.

In contrast to the $3K(I)^{3+}$ peptide, the unsubstituted 16-residue polyalanine peptide, $[Ala_{16}]^{+}$, samples rather compact structures in the gas phase but becomes more helical in solution. The driving force to solvate the charged N-terminus of the polyalanine peptide with the backbone carbonyl oxygens is apparently sufficient to disrupt the helix and form compact structures in vacuo (see Figure 18). In solution, the peptide N-terminus is solvated by water molecules and a broad distribution of more helical conformers is stablized (see Figure 18). Nonetheless, the peptide remains disordered with some internal hydrogen bonding to the N-terminus.

In the simulation studies presented above, the difference between the peptide structures observed in vacuo and in water solution are caused by the excess hydrogen bonding present in vacuo.[28,45,46] This is in agreement with the predictions of Pauling et al.[45] who postulated that excess hydrogen bonding in vacuo could potentially lead to protein structural changes/rearrangements.

5. Conclusion

Solvent and side-chain effects on three alanine-based peptides have been investigated by computer simulation. Specifically, the $[Ala_8]^{+}$ peptide, $[Ala_{16}]^{+}$ peptide, and the $[Ac-(AAAAK)-3A-NH_2]^{3+}$ peptide have been studied using a novel combination of umbrella sampling and extended system multiple time

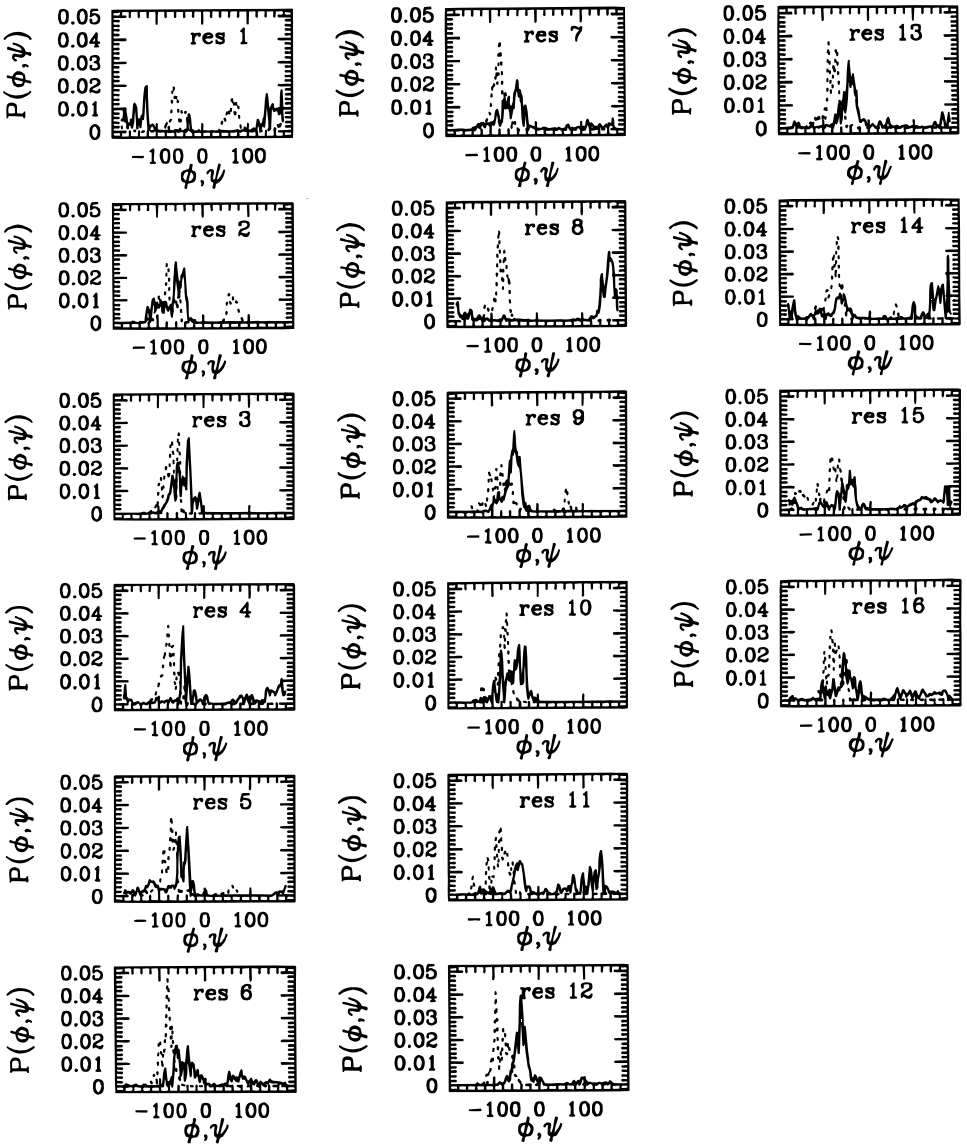

Figure 17. The probability distribution functions for each of the 16 backbone dihedral angles, ϕ (dotted line) and ψ (solid line), in the [Ala_{16}^{+}] peptide determined under the CHARMM22 force field in water solution with neutralizing chloride anions at $T = 300$ K and $P_{ext} = 1$ atm.

step molecular dynamics methods. In general, the three peptides are observed to sample a broad distribution of conformers in agreement with a modern multiconfiguration state picture of peptide thermal equilibrium. In vacuo, the lysine-doped peptide which carries its charge on the three side-chains is fairly helical with "frayed"/disordered termini.[7−11,17] The α-helical content is decreased relative to the 3_{10}-helical content in water solution because the water molecules compete successfully with the backbone for hydrogen bonding sites. In contrast, the unsubstituted polyalanines sample compact structures in vacuo. The charged N-terminus "leans" over to make favorable hydrogen-bonded contacts with the carbonyl oxygens of the peptide backbone, significantly disrupting peptide helicity. In solution, the 16-residue polyalanine samples somewhat more helical conformers. Finally, comparisons between the results produced under the CHARMM22 and the AMBER95 force fields indicates that the AMBER95 force field prefers conformations with a higher helical content in agreement with an earlier study on a blocked tripeptide.[28]

Finally, the combination of multiple time step molecular dynamics techniques and umbrella sampling has enabled the ensemble of important conformers on an atomistic potential energy surface to be obtained for several peptides. The authors have performed tests that indicate that the results are semiquantitatively accurate.[24] Thus, they expect the picture of the three peptides studied to be verified by more sophisticated future work. However, a significant limitation of the extended system molecular dynamics simulation method is its inability to generate large numbers of barrier crossing events in systems that possess rough energy landscapes[6] even when coupled with umbrella sampling techniques. Therefore, the development of new methodology which samples systems at finite temperature in the presence of large barriers to high accuracy, remains crucial. Until these methods become readily available, the convergence of thermally averaged quantities in any biosimulation study must be examined critically.

Acknowledgment. This research was supported by PRF-AC-32139 and NSF-CHE-96-5015. We thank K. Pihakari, A. Counterman, Professor D. Clemmer, and Professor G. Millhauser for useful discussions. In addition, we thank both Professor P. Kollman and Professor A. MacKerrell for placing their force fields on the Web.

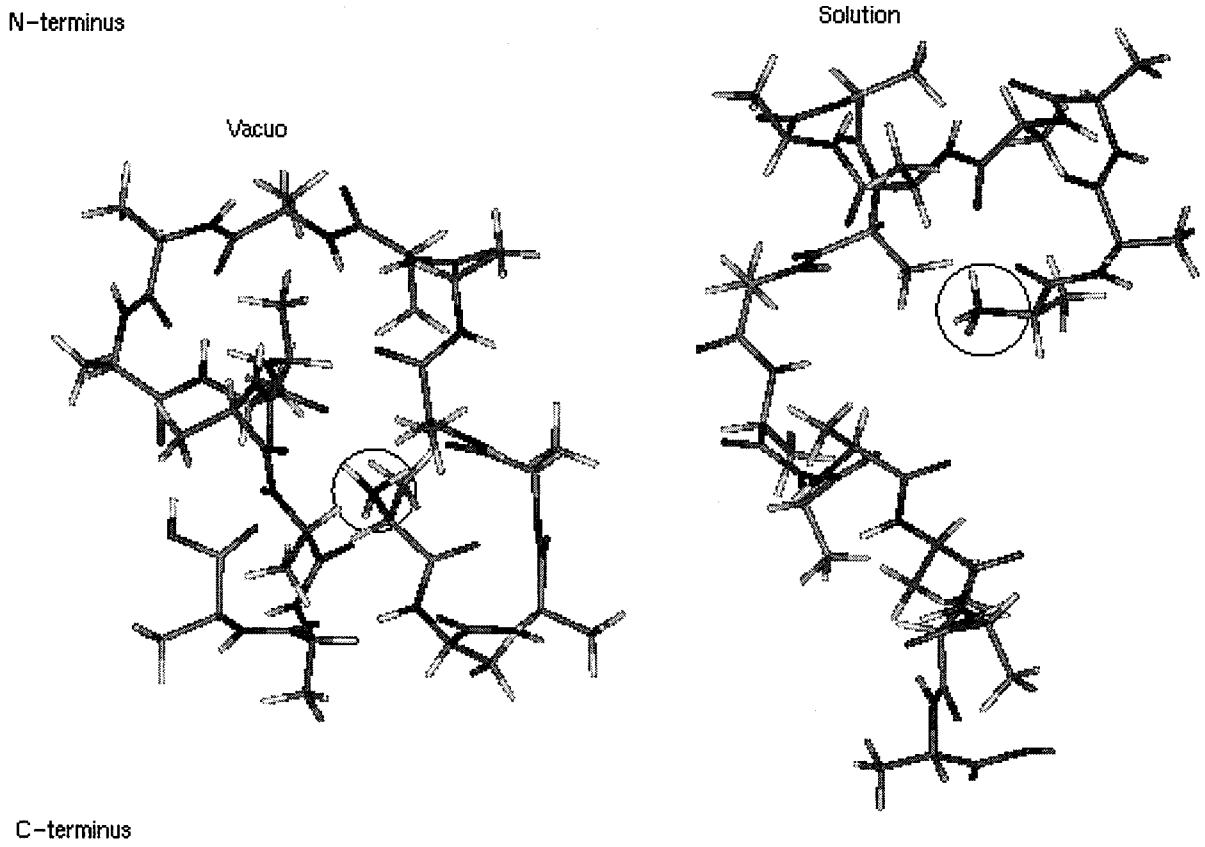

Figure 18. Instantaneous configurations of the [Ala_{16}^{+}] peptide obtained from simulations (a) in water solution, and (b) in vacuo under the CHARMM22 force field. Both configurations were taken from simulations with the umbrella sampling biasing parameters set to {$\tilde{r} = 3$ Å, $\tilde{\sigma} = 2$ Å}.

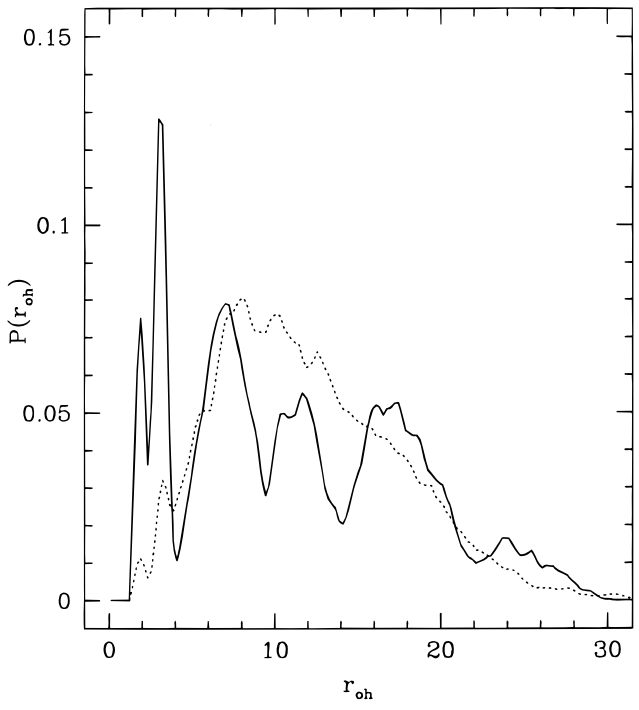

Figure 19. Comparison of the probability distribution function of the distance between the lysine NH_3^+ hydrogens and the backbone carbonyl oxygens, for the $3K(I)^{3+}$ peptide determined in vacuo (solid line) and computer water solution (dotted line) under the CHARMM22 force field at $T = 300$ K.

References and Notes

(1) Rohl, C. A.; Baldwin, R. L. *Biochemistry* **1994**, *33*, 7760.
(2) Millhauser, G.; Martinez, G. *J. Struct. Biol.* **1995**, *114*, 23.
(3) Tirado-Rives, J.; Jorgensen, W. L. *Biochemistry* **1991**, *30*, 3864.
(4) Tobias, D. J.; Brooks, C. L., III. *Biochemistry* **1991**, *30*, 6059.
(5) *Protein Folding*; Creighton, T. E., Ed.; W. H. Freeman and Company: New York, 1992.
(6) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524.
(7) Millhauser, G. L.; Stenland, C. J.; Bolin, K. A.; van de Ven, J. M. *J. Biomol. NMR* **1996**, *7*, 331.
(8) Fiori, W. R.; Millhauser, G. L. *Biopolymers* **1995**, *37*, 243.
(9) Yoder, G.; Pancoska, P.; Keiderling, T. A. *Biochemistry* **1997**, *36*, 15123.
(10) Groebke, K.; Renold, P.; Tsang, K. Y.; Allen, T. J.; McClure, K. F.; Kemp, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 4025.
(11) Smythe, M. L.; Nakaie, C. R.; Garland, R. M. *J. Am. Chem. Soc.* **1995**, *117*, 10555.
(12) Wood, T. D.; Chorush, R. A.; Wampler, F. M., III; Little, D. P.; O'Connor, P. B.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2426.
(13) Liu, Y. S.; Valentine, S. J.; Counterman, A. E.; Hoaglund, C. S.; Clemmer, D. E. *Anal. Chem.* **1997**, *69*, A728.
(14) Bowers, M. T.; Marshall, A. G.; McLafferty, F. W. *J. Phys. Chem.* **1996**, *100*, 12897.
(15) Marqusee, S.; Robbins, V. H.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5286.
(16) Miick, S. M.; Casteel, K. M.; Millhauser, G. L. *Biochemistry* **1993**, *32*, 8014.
(17) Millhauser, G. L. *Biochemistry* **1995**, *34*, 3873.
(18) Toniolo, C.; Benedetti, E. *Trends Biochem. Sci.* **1991**, *16*, 350.
(19) Pavone, V.; Benedetti, E.; Biblasio, B.; Peone, C.; Santini, A.; Bavoso, A.; Toniolo, C.; Crisma, M.; Sartore, L. *J. Biomol. Struct. Dyn.* **1990**, *7*, 1321.
(20) Petsko, A.; Ringe, D. *Ann. Rev. Biophys. Bioeng.* **1984**, *13*, 331.
(21) Tuckerman, M.; Martyna, G. J.; Berne, B. J. *J. Chem. Phys.* **1992**, *97*, 1990.
(22) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. *Mol. Phys.* **1996**, *87*, 1117.

(23) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177.
(24) Samuelson, S.; Martyna, G. J. Two-dimensional umbrella sampling techniques for the computer simulation study of helical peptides at thermal equilibrium: The 3K(I) peptide in vacuo and solution. *J. Phys. Chem.*, accepted.
(25) Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1995**, *117*, 10141.
(26) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Mertz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(27) MacKerell, A., Jr.; Bashford, D.; Bellott, M.; Dumbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kcuhnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(28) Samuelson, S. O.; Tobias, D. J.; Martyna, G. J. *J. Phys. Chem. B* **1997**, *101*, 7592.
(29) Hughs, A.; Tuckerman, M. E.; Samuelson, S.; Tobias, D. J.; Yarne, D.; Klein, M. L.; Martyna, G. J. *Mol. Phys.*, to be submitted.
(30) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578.
(31) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1989.
(32) Scheinerman, F. B.; Brooks, C. L., III. *J. Mol. Biol.* **1998**, *278*, 439.
(33) Boczko, E. M.; Brooks, C. L., III. *Science* **1995**, *269*, 393.
(34) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195.
(35) Kumar, S.; Bouzida, J.; Swendsen, R.; Kollman, P.; Rosenbreg. J. *J. Comput. Chem.* **1992**, *13*, 169.
(36) Boczko, E. M.; Brooks, C. L., III. *J. Phys. Chem.* **1993**, *97*, 4509.
(37) Roux, B. *Comp. Phys. Comm.* **1995**, *91*, 275.
(38) Nosé, S.; Klein, M. L. *Mol. Phys.* **1983**, *50*, 1055.
(39) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
(40) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384.
(41) Tobias, D. J.; Martyna, G. J.; Klein, M. L. *J. Phys. Chem.* **1993**, *97*, 12959.
(42) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1976.
(43) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(44) Wyttenbach, T.; Bushnell, J. E.; Bowers, M. T. *J. Am. Chem. Soc.* **1998**, *120*, 5098.
(45) Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 205.
(46) Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2426.