PROTEINS: Structure, Function, and Genetics 25:202–214 (1996)

Molecular Dynamics Simulations of Synthetic Peptide Folding

Shen-Shu Sung and Xiong-Wu Wu
Research Institute, The Cleveland Clinic Foundation, Cleveland, Ohio 44195

ABSTRACT **Because the time scale of protein folding is much greater than that of the widely used simulations of native structures, a detailed report of molecular dynamics simulations of folding has not been available. In this study, we included the average solvent effect in the potential functions to simplify the calculation of the solvent effect and carried out long molecular dynamics simulations of the alanine-based synthetic peptides at 274 K. From either an extended or a randomly generated conformation, the simulations approached a helix-coil equilibrium in about 3 ns. The multiple minima problem did not prevent helix folding. The calculated helical ratio of Ac-AAQ-AAAAQAAAAQAAY-NH_2 was 47%, in good agreement with the circular dichroism measurement (about 50%). A helical segment with frayed ends was the most stable conformation, but the hydrophobic interaction favored the compact, distorted helix-turn-helix conformations. The transition between the two types of conformations occurred in a much larger time scale than helix propagation. The transient hydrogen bonds between the glutamine side chain and the backbone carbonyl group could reduce the free energy barrier of helix folding and unfolding. The substitution of a single alanine residue in the middle of the peptide with valine or glycine decreased the average helical ratio significantly, in agreement with experimental observations.** © 1996 Wiley-Liss, Inc.

Key words: solvent-referenced potential, multiple minima problem, hydrophobic interaction, side chain-backbone hydrogen bonding, single-residue substitution, helical ratio

INTRODUCTION

Protein folding is a fundamental problem in life sciences, and many experimental and theoretical studies have focused on elucidating the folding mechanism. Molecular dynamics simulations can provide the time evolution of molecular structures, but a detailed report of the molecular dynamics simulation of protein folding has not been available. With lattice models,[1–4] folding has been simulated using the Monte Carlo method or exhaustive enumeration. To obtain detailed information at the atomic level, atom-based models are needed. Because thousands of water molecules are needed in the simulation of a small protein, the atom-based calculations are very time consuming. Dynamic studies of native conformations can afford the inclusion of water because such studies usually run for only a few hundred picoseconds. Starting from an unfolded conformation, folding is a global conformational change in a much larger time scale, from nanoseconds to minutes, or even longer. Including thousands of water molecules makes the folding simulation extremely time consuming. Facing the competing demands for computing power of the complicated protein-solvent system and the large folding time scale, attempts have been made to simplify the system to allow a long simulation or to reduce the time scale by simulating unfolding, including using high temperatures.

If water is not included, the system is simple, and long simulation is possible. However, if the existing force fields are not adjusted, the solvent effect is missing, and thus, the energy change is often exaggerated, as pointed out by Daggett et al.[5] As a result, a simulation is frequently trapped in local energy minima before reaching the native structure. This difficulty is known as the multiple minima problem.[6] To overcome this problem, high temperatures or thermal perturbations are often used,[7–10] including the simulated annealing method.[11] Many of these studies use Monte Carlo or other methods, instead of molecular dynamics. High-temperature molecular dynamics simulations have been used to simulate helix folding, a basic element of the protein-folding problem. In a preliminary account, Brooks[12] reported the helix folding and unfolding of a 13-residue polyalanine during a molecular dynamics simulation at a high temperature of 600 K. The detailed analysis of this folding and unfolding behavior is not yet available.

In a review by Karplus and Shakhnovich,[13] helix folding simulations (by Robert and Karplus) of a 9-residue peptide, $(Ala)_2(Asp)(Ala)_6$, from an ex-

Received August 15, 1995; revision accepted November 20, 1995.

Address reprint requests to Shen-Shu Sung, Research Institute, The Cleveland Clinic Foundation, 9500 Euclid Ave., Cleveland, OH 44195.

tended initial conformation were briefly reported. Simulations in a vacuum ($\varepsilon = 1$) and in aqueous solution suggested similar folding mechanisms, but with substantially different time scales. A detailed description of the study discussing aspects such as the force field, simulation temperature, energy changes, and structure evolution is not available. These pioneering works, including the simulations using high temperatures, have clearly indicated the need for a folding simulation study. On the other hand, the lack of a full report on molecular dynamics simulation of folding may reflect the great difficulty of the problem.

Unfolding simulations are a natural extension of the dynamic simulation of native structures when large deviations from the initial structure are observed. The helix unfolding simulation with or without water has been carried out using both the residue-based model[14,15] and the atom-based model.[5,16–19] With the atom-based model, these simulations are usually carried out for several hundred picoseconds, so explicitly including solvent molecules is feasible and some valuable information on solvent effects has been provided.

Unfolding simulations start with an ordered structure and do not require a specific final structure. In contrast, folding simulations require an ordered final structure from various initial structures. The force field for a folding simulation must satisfy the requirements that the native structure has the lowest free energy accessible from various initial conformations with the experimentally relevant time scale and temperature. These requirements for the force field, in addition to the computation efficiency, make the folding simulation very difficult. At the present time, the atom-based molecular dynamics simulation of folding is still in its infancy.

In our previous work,[20,21] the helix folding of alanine-based peptides was simulated using the Monte Carlo method at normal temperatures (300 and 274 K). To make long folding simulations feasible, water molecules were not included explicitly. Instead, the average solvent effect was included by using a solvent-referenced potential, which reduced the exaggerated energy changes that occur with the "in vacuo" calculation and lowered the energy barriers that cause the multiple minima problem. The multiple minima problem did not prevent helix folding with various initial conformations. In the current study, the concept of the solvent-referenced potential is extended to molecular dynamics simulations of helix folding at (or near) the experimental temperature. The hydrophobic interaction is calculated based on the solvent-accessible surface area. A synthetic peptide, Ac-AAQAAAAQAAA-AQAAY-NH_2,[22] and two related peptides with single substitution by glycine and by valine are studied.

MODEL AND METHODS

To study the basic features of protein folding, an average solvent effect may serve as the reference in energy functions to reduce the computation load of solvent calculations. The solvent effect may be regarded as a function of many variables, such as the time, the spatial location, and the protein conformation. An average solvent effect is a first-order approximation. The different hydrophobic interactions of various residues can be included as a higher order correction. The computational efficiency is a major purpose for using the solvent-referenced potential. It may also reduce the inaccuracy resulting from the cancellation of large energy terms calculated with vacuum reference. The competing effects of the solvent, such as the van der Waals (VDW) attraction and the hydrogen bonding, usually reduce the strength of the interactions overestimated by the "in vacuo" calculation. Therefore, the solvent-referenced potential represents a simplified and reduced interaction. The interactions with specific solvent molecules, such as specific hydrogen bonds with water, are not represented. However, when overall dynamics of protein folding are examined, it is likely that the solvent does not have to be considered explicitly.[13]

In our study, the treatment of the VDW, electrostatic, and hydrophobic interactions are chosen according to the concept of the solvent-referenced potential and are described in the following paragraphs. We tested these treatments on helix folding with the AMBER force field parameters.[23,24] In principle, the concept of the solvent reference can be applied to other force fields, although detailed treatments may be different. The virtue of this solvent-referenced potential rests on its ability to simulate peptide folding and to give results in good agreement with experimental observations.

In biological systems, the intramolecular VDW interactions of a protein molecule are balanced by the intermolecular VDW interactions with solvent molecules. When solvent molecules are not included, the intramolecular VDW interactions should be adjusted accordingly. The longer range attractive VDW interactions provide a nearly uniform background potential and therefore can serve as the reference for the VDW energy calculation. The possible difference between the protein intramolecular VDW attraction and that with water is included in the hydrophobic interaction energy. The short-range repulsion represents the exclusive volume of each atom and needs to be calculated explicitly. Based on this consideration, a shifted truncation is applied to the VDW interaction, as shown in Equation (1), where r is the distance between two interacting atoms and r^* is the minimum energy distance for the given pair of atoms. This treatment is based on the mean field approximation in the VDW theory

of the liquid-solid transition.[25] A similar treatment has been previously applied to protein structure studies.[14] To compensate for the thermal motion, r* is scaled to 95% of the original value of the AMBER force field.

$$E_{VDW}(r) = \begin{cases} \epsilon\left[\left(\frac{r^*}{r}\right)^{12} - 2\left(\frac{r^*}{r}\right)^{6} + 1\right] & r < r^* \\ 0 & r \geq r^* \end{cases} \quad (1)$$

The dielectric screening effect of the solvent could be included by adjusting the dielectric constant. Without including the solvent, the distance-dependent dielectric constant, $\epsilon = r$, was originally suggested in the AMBER force field. However, this dielectric constant overestimates the electrostatic interaction and, therefore, a sigmoidal function has also been suggested for the AMBER force field.[5] This sigmoidal function is based on Debye's theory of ionic saturation[26] and has been developed by Hingerty and co-workers[27] and by Ramstein and Lavery.[28] In the current study, the sigmoidal dielectric function defined in Equation (2) was used. The value of s is 0.3, as used by Daggett et al.[5] D is the bulk solvent dielectric constant, 78. The value of the dielectric function varies smoothly from 2 to 78 as the distance increases,[29] closely following an experimentally derived curve.[30]

$$\epsilon = D - \frac{D-2}{2}[(sr)^2 + 2sr + 2]e^{-sr} \quad (2)$$

The hydrophobic property of the amino acids is a most important factor in protein folding. For homopolypeptides, such as polyalanine, it is possible to include a single average effect in the potential function. When a peptide contains various amino acids, differences in their hydrophobic interaction energies must be represented. The hydrophobic interaction is often assumed to be proportional to the degree to which protein atoms are exposed, a measure such as the solvent-accessible surface area,[31,32] or the hydration shell volume.[33] In the current study, the hydrophobic interaction free energy, ΔG, is the sum of the contribution from each atom, and the contribution from an atom i is assumed to be the product of an atomic solvation parameter $\Delta\sigma_i$ and the solvent-accessible surface area A_i,[31,34] as shown in Equation (3).

$$\Delta G = \sum_i \Delta\sigma_i A_i \quad (3)$$

Wesson and Eisenberg[34] and Schiffer et al.[35] have carried out molecular dynamics simulations with the surface area-based solvation energy. In some of these studies, the atomic solvation parameters were calculated from the free energy of transfer between water and vapor of small organic molecules measured by Wolfenden et al.[36] and modified by Kyte and Doolittle[37] and by Sharp et al.[38] Since we use the average solvent environment as reference, the free energy of transfer between water and a protein-like environment, rather than a vacuum or gas phase, should be used. To mimic the properties of the protein interior, a partially polar solvent is needed, and octanol is probably one of the best choices.[38] Therefore, the free energies of transfer of N-acetyl amino amides from octanol to water are used in the current study. These values have been measured by Fauchere and Pliska[39] and modified by Sharp et al.[38] for volume entropy changes.

Eisenberg and McLachlan[31] have calculated the solvation parameters without the modifications. Following the procedures of Eisenberg and McLachlan, we calculated the solvation parameters with the modifications. The protein atoms are categorized according to six types: C, O, N, O^-, N^+, and S, and the solvation parameters are calculated for each type. Wesson and Eisenberg[34] estimated the average atomic solvent-accessible surface areas of the side chain analogs using the conformations in the Protein Data Bank.[40] In the side chain analogs, the β-carbon atoms are not connected with the backbone and, therefore, have a larger accessible surface area. In our study, the surface areas of the non-carbon atoms of Wesson and Eisenberg were used directly, and the surface areas of carbon atoms were obtained from the total surface areas given by Sharp et al.[38] and those of the non-carbon atoms. The $\Delta\sigma$ values used for C, N, and O atoms in the current study were 25.8, 4.6, and −25.1 cal/mol/Å^2, respectively. The N^+, O^-, and S atom types were not used in this study.

The simulations were carried out using AMBER software.[23] The time step was 0.002 ps with all bond lengths fixed by the SHAKE procedure.[41] The AMBER software we obtained does not contain the solvent-accessible surface area calculation. To evaluate the hydrophobic interaction, the TINKER software[42,43] was used for the solvent-accessible surface area calculation. The surface area-based hydrophobic interaction calculation is more efficient than including water molecules explicitly, but it still increases computing time by about 15-fold compared with that without the hydrophobic interaction calculation. However, the hydrophobic interaction energy change at each step is more than an order of magnitude smaller than the fastest changing energy term (usually, the bond vibration energies). To speed up the simulation, a larger time step (0.02 ps) was used for calculating the hydrophobic interaction, and the results were compared with those calculated every 0.002 ps. The larger time step (0.02 ps) for hydrophobic interaction calculation speeds up the simulation by about sixfold.

Because a detailed report on the molecular dynamics simulation of helix folding is not yet avail-

able, we decided to start with a simple, 16-residue synthetic peptide.[22] It is a neutral, water-soluble peptide with an amino acid sequence of Ac-AA-QAAAAQAAAAQAAY-NH_2, or Ac-$(AAQAA)_3$Y-NH_2. This alanine-glutamine peptide contains a single tyrosine residue, which allows the peptide concentration to be measured accurately by tyrosine absorbance. Its helical content is about 50% as measured by circular dichroism (CD). We carried out simulations of this peptide with both a fully extended initial conformation (all ϕ and ψ dihedral angles are 180°) and a randomly generated initial conformation. Experimental studies have shown that a single-residue substitution on alanine-lysine peptides may change the helical content of the peptide, depending on the helical propensity of the amino acids and the location of the substitution.[44,45] To test how the simulation depends on amino acids sequence, we conducted simulations for two peptides with single substitutions: one is A9V: Ac-AA-QAAAAQVAAAQAAY-NH_2 and the other is A9G: Ac-AAQAAAAQGAAAQAAY-NH_2. For the N-terminal Ac-group, the atomic charges of the carbonyl group were taken from the backbone carbonyl groups, and the charge on the methyl group was determined by making the entire Ac-group neutral. For the C-terminal -NH_2 group, the hydrogen charge was taken from the -NH_2 group in the glutamine side chain, and the charge on the nitrogen was determined by making the entire -NH_2 group neutral. All our simulations were conducted with constant temperature molecular dynamics at 274 K, because many alanine-based synthetic peptides have been studied experimentally at 274 or 273 K. The 1 K difference has little effect on the simulation because both the force field and the molecular dynamics method are approximate. The computations were carried out on the SiliconGraphics Personal Iris 4D35. All figures representing peptide conformations were produced using the MidasPlus software system from the Computer Graphics Laboratory, University of California (San Francisco, CA).[46]

RESULTS AND DISCUSSION

Ac-$(AAQAA)_3$Y-NH_2 Simulation With Extended Initial Conformation

The first simulation started with a fully extended initial conformation with all ϕ and ψ angles at 180°. The energy changes during the simulation, including the total energy, hydrophobic interaction energy, and electrostatic energy, are shown in Figure 1. To describe the conformation distribution, a three-dimensional Ramachandran plot is shown in Figure 2. The third dimension shows the conformation densities. The secondary structures have characteristic regions in a Ramachandran plot: α-helix conformations are centered at (−57°, −47°), 3_{10} helices at (−70°, −5°), and β-sheets near (−120°, 140°).[47] In Figure 2, the highest density is at the α-helix region and is centered at about (−60°, −40°). A considerable density also exists at the 3_{10}-helix region, but very low density is found near the β-sheet region. The small peaks at about (−80°, 60°) and (80°, −60°) are a result of the i − 1, i + 1 hydrogen bonding conformation.

To detect a helix by CD measurement, a minimum length of three helical residues is required,[48] which is the length also necessary to form the α-helical hydrogen bond. Therefore, we defined a helical segment as three or more consecutive residues with dihedral angles within 30° from the standard values, $\phi = -57°$ and $\psi = -47°$. During the simulation, a large number of conformations appeared and interconverted rapidly. Because it is impractical to show all conformations in a figure, we choose to show the locations of the helical segments in 10-ps time intervals in Figure 3. Each vertical line represents a helical segment, and the residue numbers on the vertical axis indicate the location of the helical segments in the sequence. The length of the line segment is proportional to the number of residues in the helical segments.

As shown in Figure 3, isolated helical segments of three or four residues formed from residue 7 to residue 11 within the first 100 ps of the simulation, but they did not propagate to the whole peptide chain immediately. These short helical segments unfolded and refolded many times before forming a complete helix. Many different conformations were observed. Elements of another basic secondary structure, the β-sheet, were also observed. As an example, Figure 4 shows that a β-hairpin conformation occurred at 460 ps. This conformation is not thermodynamically favored for the given sequence and is unstable during the simulation, as indicated by the low density in the β-sheet region of Figure 2. However, it shows that our simulation can cover a reasonably large portion of the conformational space without suffering from the multiple minima problem.

After 1,500 ps, more helical segments appeared. As in our Monte Carlo simulations,[20,21] i, i + 3 hydrogen bonds were also observed during the molecular dynamics simulation, as shown by the density at the 3_{10} helix region in Figure 2. This type of hydrogen bond forms the 3_{10} helix, whereas the i, i + 4 type of hydrogen bond forms the α-helix. The i, i + 3 hydrogen bonds were frequently observed during helix folding and unfolding. This observation is consistent with the conclusion that the 3_{10} helix is the intermediate structure in α-helix folding.[17] However, a complete 3_{10} helix was rarely observed during our simulation. Instead, the i, i + 3 hydrogen bond was more often found with i, i + 4 hydrogen bonds in a helix segment, as shown in the conformation at 2,370 ps in Figure 4. These simulation results are in good agreement with experimental studies.[49–52]

After 3,000 ps, the helix-coil transition ap-

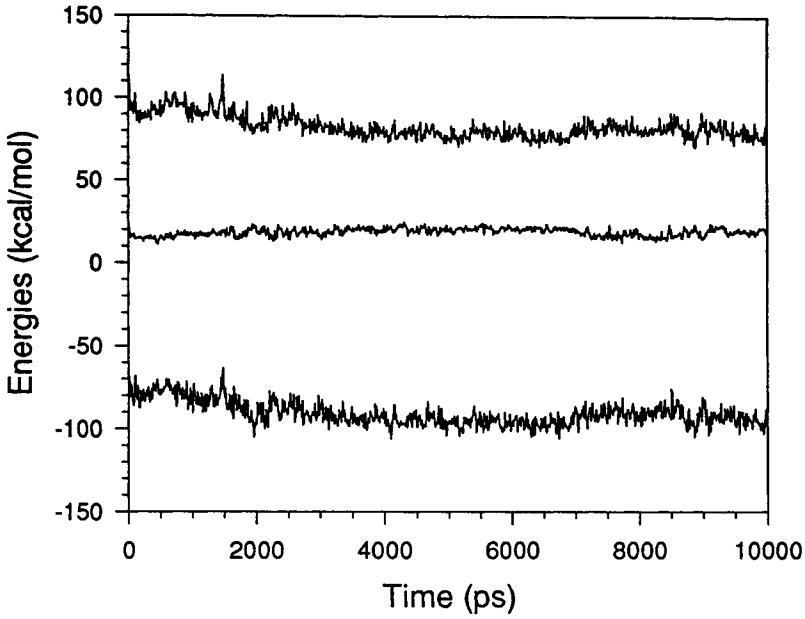

Fig. 1. The energy changes during the simulation of Ac-$(AAQAA)_3$Y-NH_2 with the fully extended initial conformation. The top curve shows the total energy, the middle curve shows the hydrophobic interaction energy, and the bottom curve shows the electrostatic energy. Upon helix folding, the total energy and the electrostatic energy decreased, whereas the hydrophobic interaction energy increased slightly.

proached equilibrium (a much longer simulation may be needed to ascertain the equilibrium further). Figure 4 shows a complete helix occurred at 3,280 ps. Upon helix folding, the total energy and electrostatic energy decreased, and the hydrophobic interaction energy increased slightly. The change in the VDW interaction energy [its attractive part has been truncated according to Eq. (1)] was small during helix folding and is not shown in the Figure 1. Between 3,000 and 7,000 ps, the most frequently observed conformation was a helix with unfolded ends.

Between 7,000 and 9,000 ps, two helical segments often coexisted in a distorted helix-turn-helix conformation, and the total energy increased about 3 kcal/mol. As an example, a compact, distorted helix-turn-helix conformation at 8,090 ps is shown in Figure 4. Its number of hydrogen bonds is only about half that of a complete helix and, therefore, it has a higher electrostatic energy than a complete helix. However, its surface area is small, which provides stabilization energy from the hydrophobic interaction. Using a biased-probability Monte Carlo (BPMC) procedure, Abagyan and Totrov[53] have predicted a similar compact structure containing two interacting α-helical fragments that has an energy 4 kcal/mol higher than the regular α-helix.

To see how hydrophobic interactions contribute to the conformation equilibrium, the solvent-accessible surface areas of three typical conformations were analyzed, including a fully extended conformation at the beginning of the simulation, a complete helix at 4,240 ps, and a compact conformation at 8,090 ps (the distorted helix-turn-helix in Fig. 4). The solvent-accessible surface areas of these three conformations and the contributions from each atom type are shown in Figure 5. Among the three conformations, the extended conformation has the largest solvent-accessible surface area and the highest hydrophobic interaction energy. The complete helix has a smaller surface area and a lower hydrophobic interaction energy than the extended conformation. However, because hydrophilic oxygen atoms are also buried, the total hydrophobic interaction energy decreased only slightly from 25 to 24 kcal/mol when an extended conformation folded into a helix. The compact conformation has a much lower hydrophobic interaction energy of 15 kcal/mol. Compared with the helix, hydrophobic carbon atoms in the compact structure are exposed less to solvent, and hydrophilic oxygen atoms are exposed more. Both changes contribute to a lower hydrophobic interaction energy. For Ac-$(AAQAA)_3$Y-NH_2, the electrostatic interaction favored helices and determined the structure in general, and the hydrophobic interaction favored the compact conformations and determined the structure sometimes (e.g., 7,000–9,000 ps in Fig. 3). The hydrophobic interaction is the cause of the occurrence of the distorted helix-turn-helix conformation. In our simulations, the attractive VDW interactions have been truncated [see Eq. (1) in Model and Methods], and the hydrophobic interaction en-

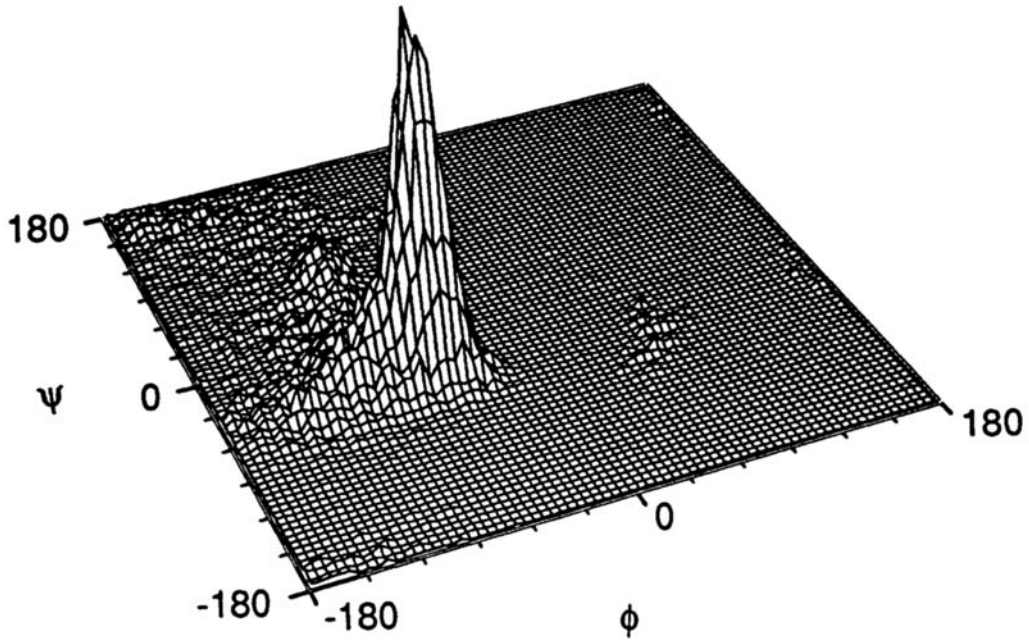

Fig. 2. The ϕ, ψ angle distribution. The vertical dimension shows the relative densities of conformations. The highest peak is at the α-helix region and is centered at about $-60°$, $-40°$. A considerable density also exists at the 3_{10}-helix region ($-70°$, $-5°$), but very low density is found near the β-sheet region ($-120°$, $140°$). The small peaks at about ($-80°$, $60°$) and ($80°$, $-60°$) are a result of the i − 1, i + 1 hydrogen bonding conformation.

ergies, e.g., 25 cal/mol/Å^2 for carbon atoms, are in the reasonable range.[54] However, the experimentally measured hydrophobic interactions vary with experimental methods. The issue of whether the distorted helix-turn-helix conformation is the result of overestimated hydrophobic interactions will be clarified when more experimental observations on the peptide structure are available. Also, in our future studies different hydrophobic interaction energies will be tested.

From 3,000 to 7,000 ps the equilibrium occurred among the conformations containing mainly single helical segments of various lengths (Fig. 3). Occasionally, a couple of hydrogen bonds broke in the middle of the helical segment and the conformation contained two nearly coaxial helical segments, but these two segment often became one segment within a picosecond. This kind of transition occurred at the picosecond (or less) time scale. If the orientations of the two segments changed to nearly antiparallel, a helix-turn-helix conformation formed, which is stabilized by hydrophobic interactions. This kind of transition may be regarded as one between a single secondary structure element and a supersecondary structure, and occurred at a much larger, nanosecond time scale. Our simulation has been extended to 20,000 ps. The distorted helix-turn-helix conformation occurred between 11,000 and 13,000 ps again. During the equilibrium between 3,000 and 20,000 ps, at least two different types of transitions have been observed at quite different time scales.

During simulations, the Gln side chain amides often formed transient hydrogen bonding with the backbone carbonyl oxygen atoms. Figure 6 shows the conformations at 8,670, 8,680, 8,690, and 8,700 ps in the second simulation. The NH_2 group of the Gln-13 side chain formed a hydrogen bond with the backbone carbonyl Ala-9 at 8,670 ps. The NH_2 of the Gln-8 side chain formed a hydrogen bond with the carbonyl of Ala-6 and then shifted to the carbonyl of Ala-5. At 8,700 ps, these hydrogen bonds broke, and a backbone-backbone hydrogen bond in the α-helical pattern formed. Because the side chain usually has more freedom to move than the backbone, side chain-backbone hydrogen bonding may form before backbone-backbone hydrogen bonding during folding. A major component of the free energy barrier of helix folding comes from the entropy decrease. The binding of the side chain of residue i to the carbonyl of residue i − 2, i − 3, or i − 4 brings these residues to a short distance for possible backbone-backbone hydrogen bonding and reduces the kinetic free energy barrier caused by the entropy effect. The competing effect of the side chain with backbone hydrogen bonding in the middle of a helix may thermodynamically destabilize backbone hydrogen bonding. The thermodynamic effect of side chain-backbone hydrogen bonding of the helix boundary residues has been proposed by Presta and Rose[55] using crystal structure analysis. Our simulations showed the kinetic effect of transient hydrogen bonds of the side chains with backbone in the middle of a helix.

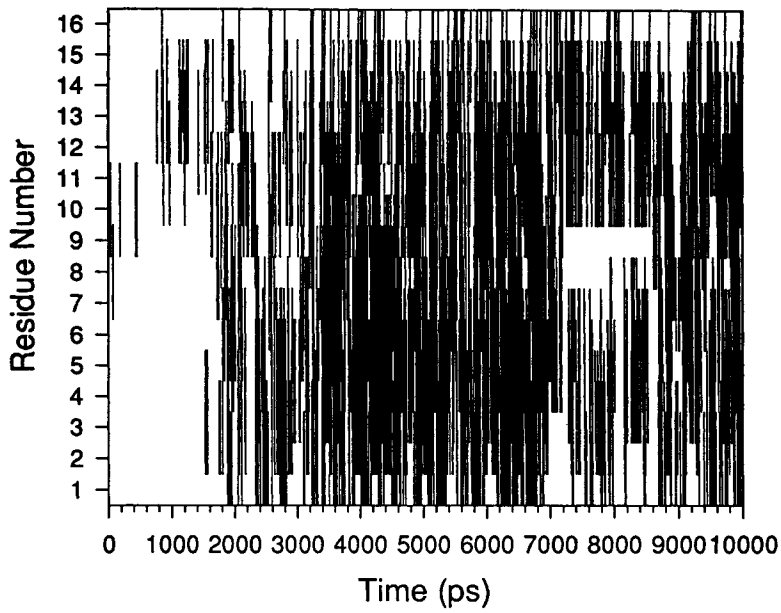

Fig. 3. The locations of helical segments. A helix segment consists of at least three consecutive helical residues whose ϕ, ψ angles are within 30° of the standard values of $\phi = -57°$, $\psi = -47°$. The helical segments in 10 ps intervals are shown. In the initial 3,000 ps, helical segments occurred with increasing frequency. From 3,000 to 10,000 ps, the conformations are in equilibrium. Between 7,000 ps and 9,000 ps, two segments of helix occurred in distorted helix-turn-helix conformations.

After 9,000 ps, the conformation contains mainly a single helical segment again. The average potential energy was 25.5 kcal/mol from 100 to 1500 ps before helix folding and was 10.9 kcal/mol from 4,000 to 10,000 ps after helix folding. The potential energy change was 14.6 kcal/mol, which agrees well with the experimentally measured enthalpy change of 0.9–1.3 kcal/mol/residue upon helix folding.[56,57] The average helical ratio was 47% from 3,000 to 10,000 ps, in good agreement with the CD measurement of about 50%.[22,58] The length of helical segments ranged from 3 to 16 residues, with the most frequently observed helical length being 9–12 residues.

Helix-coil transition theories[59–62] usually assume that the probability of helix initiation is small and mainly depends on the entropy effect. The probability of propagation is larger because of the helix-stabilizing energy. Thus, a short peptide usually has a single helical segment and the probability of having more than one helical segment is very small.[63,64] In our simulation of Ac-$(AAQAA)_3$Y-NH_2, the 16-residue peptide could form more than one helical segment. On average, one helical segment only was observed in 50–60% of conformations, two helical segments in 17–30% of conformations, and three or more helical segments in less than 2% of conformations. This observation implies a larger helix initiation parameter than expected.

We estimated the helix initiation parameter σ. As mentioned previously, we use dihedral angles as the criterion for helical conformations, whereas the original Zimm-Bragg theory uses hydrogen bonds to define the helical segment. Therefore, we first used formulas of the Lifson and Roig theory. The numbers of occurrence of each residue at the coil segments (u'), at the terminal positions of helical segments (v'), and at the middle positions of helical segments (w') were counted directly from the conformations sampled every 10 ps during the time period from 3,000 to 10,000 ps. These numbers of occurrence were then normalized to obtain the weighting factors 1, v, and w. The parameter v is converted to σ, according to the formula $\sigma = v^2/(1 + v)^4$ from Qian and Schellman.[63] Using this procedure, an average value of $\sigma = 0.056$ was obtained. A similar estimated value of the helix initiation parameter was also obtained from the formulas of the Zimm-Bragg theory[60] by converting the dihedral angle definition of helix to the hydrogen bond definition.

These results are rough estimates. At different positions of the sequence, the residues have different calculated σ values, and 0.056 is the average value of residues in the whole peptide. The value of σ should be different for Ala and Gln. However, because enough data were not available to calculate the parameters for Gln separately from Ala, and because the effect of Gln is not limited to the residue itself, the difference in residue types was not considered in averaging. The value of σ has been estimated experimentally[65] and is in the range of 10^{-2}–10^{-5}. Our calculated value is much larger, and the cause

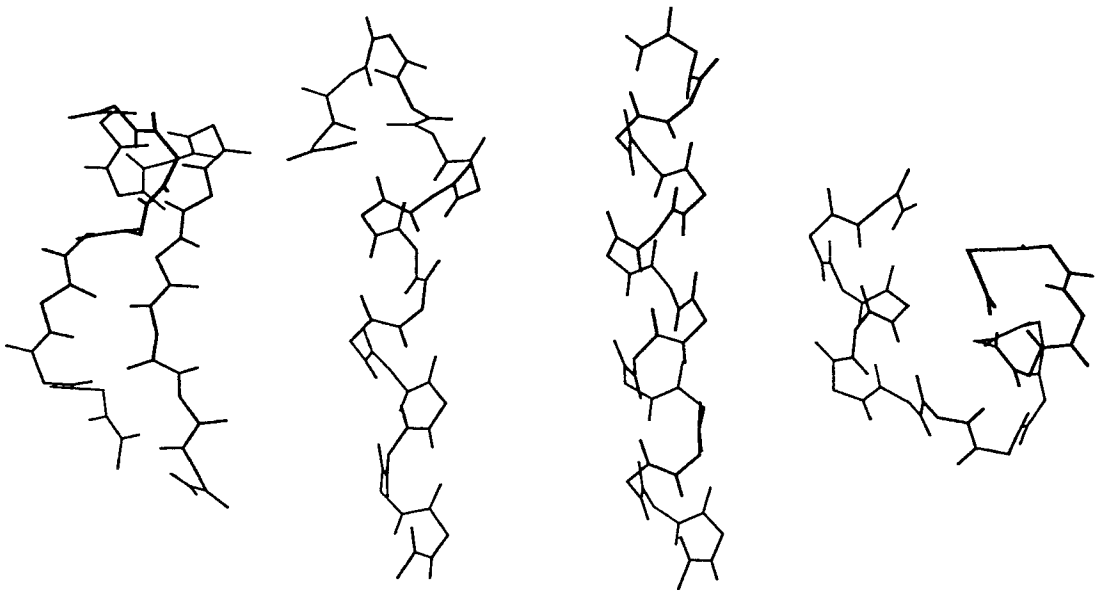

Fig. 4. Selected conformations observed during the simulation. From left to right, the conformations are a β-hairpin conformation at 460 ps, a mixed α-helix and 3_{10} helix segment at 2,370 ps, a complete helix at 3,280 ps, and a compact, distorted helix-turn-helix conformation at 8,090 ps. The side chains are not shown.

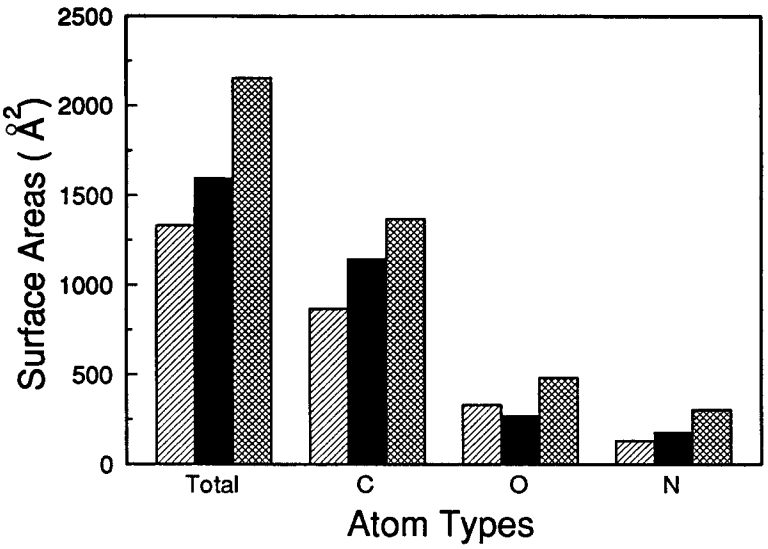

Fig. 5. The solvent-accessible surface areas of the three conformations. The striped bars indicate the compact conformation; the solid bars, the helical conformation; and the hatched bars, the extended conformation. The surface areas of C and N atoms contribute to the positive energy, whereas the surface areas of O atoms contribute to the negative energy.

of the difference from the experimental estimates is not yet clear. Reducing the dihedral angle range of the helix definition by half can make σ smaller, but its order of magnitude remains at 10^{-2}. In the helix unfolding study, Daggett and Levitt[18] found that the intrinsic initiation parameter (calculated from the equilibrium constant $<K>_{ccc \rightarrow chc}$) was independent of temperature below 473 K (to 278 K) and equal to 0.4–0.5, which is also much larger than the experimental estimates. More simulations on different amino acid sequences are needed to better calculate these parameters in our future work.

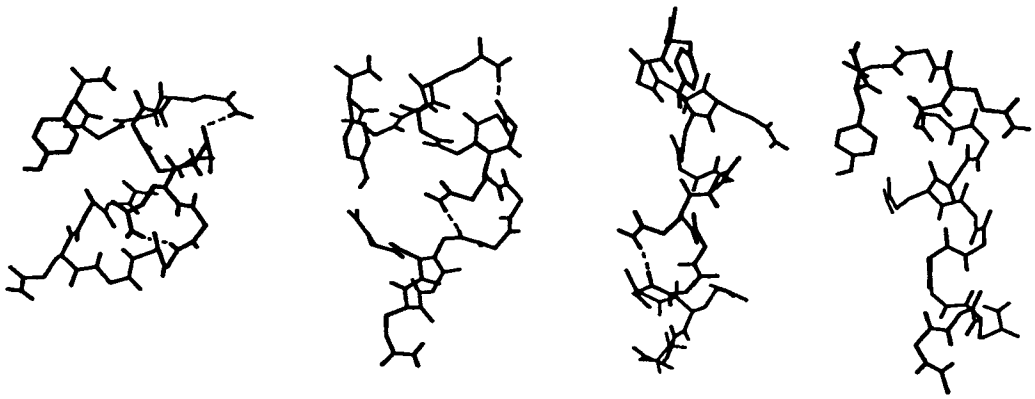

Fig. 6. The side chain-backbone hydrogen bonding at 8,670, 8,680, 8,690, and 8,700 ps (from left to right), indicated by dotted lines. Alanine side chains are not shown. At 8,670 ps the NH_2 group of the Gln-13 side chain bound to the backbone carbonyl Ala-9, and the NH_2 of the Gln-8 side chain bound to the carbonyl of Ala-6 (and shifted to the carbonyl of Ala-5 at 8,680 ps). At 8,700 ps, these side chain-backbone hydrogen bonds broke, and backbone-backbone hydrogen bonds in the α-helical pattern formed.

Tests on Different Initial Conformation and on Smaller Time Step

Another simulation of Ac-$(AAQAA)_3$Y-NH_2 started with a randomly generated initial conformation and ran for 10,000 ps. The energy changes were similar to those shown in Figure 1. The locations of the helical segments are shown in Figure 7. Although the initial conformation is different, the helix-coil transition also approached equilibrium at about 3,000 ps. The real time scale of the helix-coil transition is believed to be on the order of nanoseconds to microseconds.[14] In our simulation, the folding time is at the low end of the helix-coil transition time scale, because the solvent motion-damping effect[66] is not included. From both Figure 3 and Figure 7, it is apparent that the initial folding period is qualitatively different from the refolding at equilibrium. The initial folding needs a much longer time (several nanosecond) than the refolding during equilibrium. The initial folding in the simulation may resemble some features of the non-equilibrium transition that occurs experimentally. Theoretically, the initial folding is important in testing the potential function. If the potential function can not fold various initial conformations into a helix, the reliability of the simulation is in question. The much faster refolding indicates that the unfolded conformations in the simulation are located near the folded region on the reaction coordinates (whose definition has been discussed in detail in Dill et al.[67] and Bryngelson et al.[68]). The distorted helix-turn-helix conformation was observed at about 7,000 ps and at about 10,000 ps. This simulation indicates that the general features of the helix-coil transition described previously do not depend on the initial conformation, and the multiple minima problem did not prevent helix folding.

To save computing time, the simulations discussed previously use the time step 0.02 ps for the hydrophobic interaction calculations. For comparison, a simulation was carried out with the hydrophobic interaction calculated every 0.002 ps, which is the same time step as for other force calculations. During this simulation, the peptide also showed many different conformations before forming a helix. The first helical segment formed at 510 ps. The first complete helix formed at 2,590 ps. Various helical segments were observed after 2,590 ps. The conformations at 0, 1,000, 2,000, and 2,590 ps are shown in Figure 8. Again, the general features of the folding, including the initial folding time, were very similar to the previous simulations. This simulation was stopped after helix formation, at 2,750 ps, because the small time step for calculating the hydrophobic interaction made the simulation very time consuming.

Single-Residue Substitutions With Valine or Glycine

Experimental measurements have shown that a single-residue substitution by valine or glycine in the middle of the sequence greatly reduces the helical content of alanine-lysine peptides.[44,45] To test the sequence dependence of our simulation, the residue Ala-9 was substituted with valine (A9V) and glycine (A9G) in two separate simulations. The simulation conditions (or parameters) were the same as in the simulation of Ac-$(AAQAA)_3$Y-NH_2. The time step for calculating the hydrophobic interaction was 0.02 ps. Both simulations of A9V and A9G started from the fully extended conformation and approached their helix-coil equilibrium in about 3,000 ps. The helical ratio of each residue, defined as the

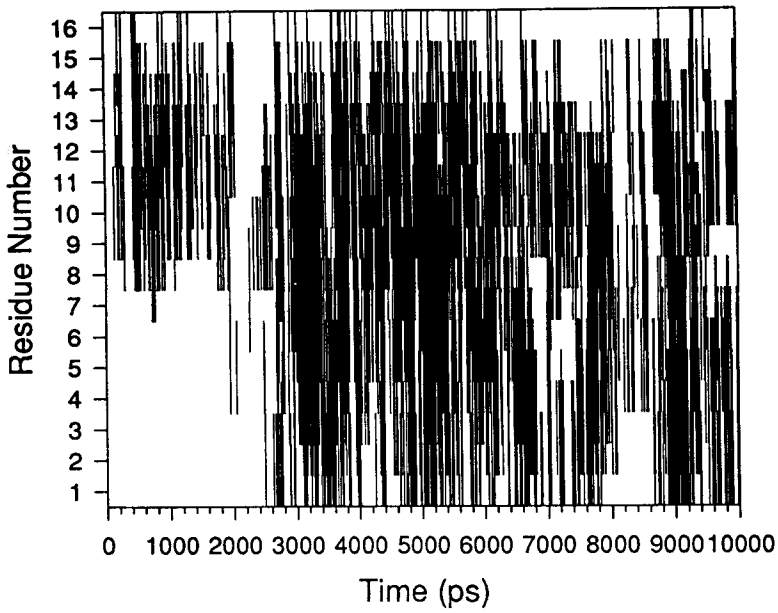

Fig. 7. The locations of helical segments in the simulation of Ac-$(AAQAA)_3$Y-NH_2 with the randomly generated initial conformation. A helix segment consists of at least three consecutive helical residues whose ϕ, ψ angles are within 30° of the standard values of $\phi = -57°$, $\psi = -47°$. The helical segments in 10 ps intervals are shown. Helical conformations formed during the initial 3,000 ps. From 3,000 to 10,000 ps, the conformations are in equilibrium. At about 7,000 ps and about 10,000 ps, two segments of helix occurred.

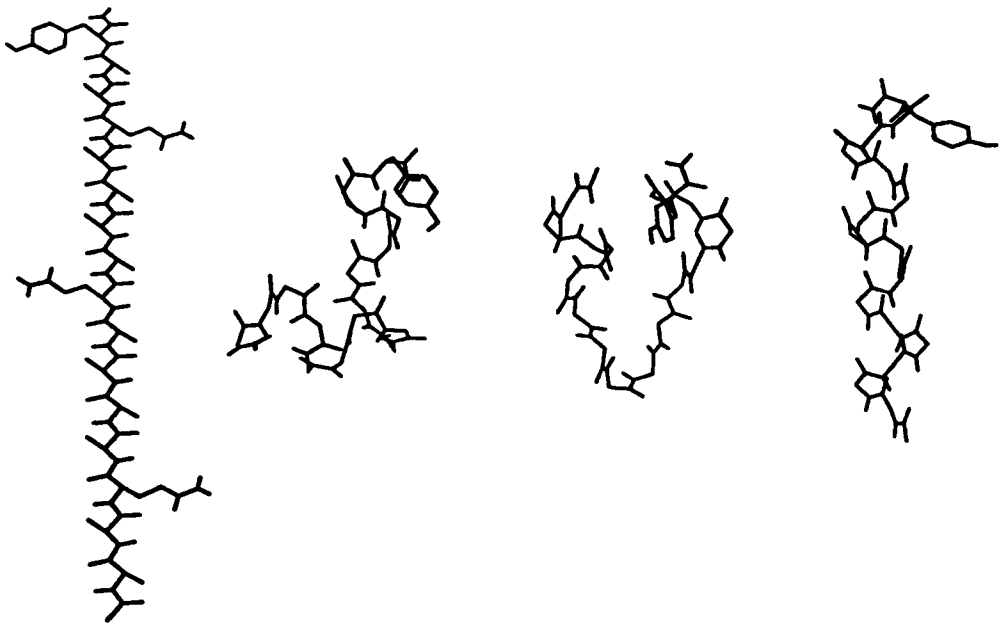

Fig. 8. Conformations at 0, 1,000, 2,000, and 2,590 ps (from left to right) in the simulation with the hydrophobic interaction calculated every 0.002 ps. In the conformations at 1,000, 2,000, and 2,590 ps, the side chains are not shown, except that of tyrosine, which indicates the C-terminus.

percentage of the conformations in which the residue is located in a helical segment of at least three residues, was calculated for the time period from 3,000 to 10,000 ps and shown in Figure 9.

Because the hydrophobic interaction favors the distorted helix-turn-helix conformation, the time-averaged helical ratio is slightly lower for the middle residues of Ac-$(AAQAA)_3$Y-NH_2 in Figure 9.

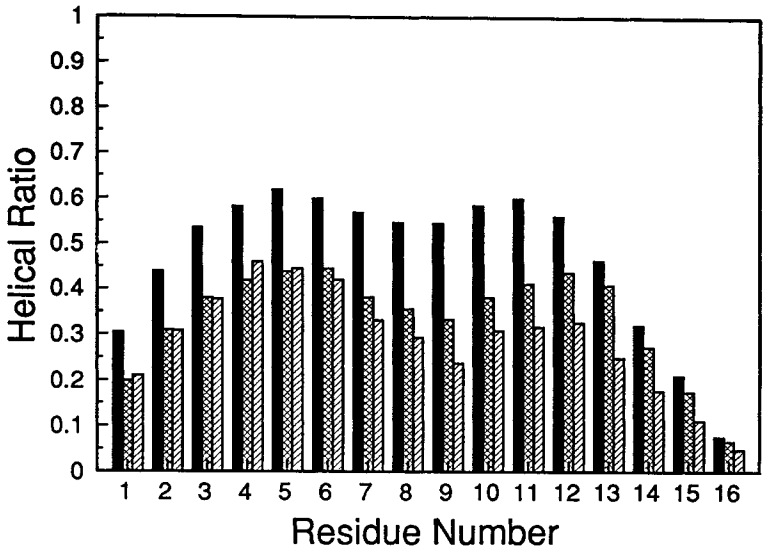

Fig. 9. The helical ratios of the residues, defined as the percentage of the residues that is located in a helix segment of at least three residues. The solid bars indicate the helical ratios of Ac-$(AAQAA)_3$Y-NH_2; the hatched bars, A9V; and the striped bars, A9G.

This observation was not predicted previously by the theoretical analysis of experimental data.[44,45] Our previous Monte Carlo study of alanine-based peptides did not predict the distorted helix-turn-helix conformation either, probably because the group-based hydrophobic interaction used previously is weaker at short distances than the interactions calculated from the surface areas. As mentioned before, the average helical ratio of Ac-$(AAQAA)_3$Y-NH_2 is 47%. The average helical ratios of A9V and A9G are 34% and 29%, respectively. For A9V and A9G, the lower helical ratios of the middle residues are caused by the lower helical tendencies of valine and glycine, in addition to the hydrophobic effect. In experimental studies,[44] a single glycine substitution at residue 9 of a 17-residue alanine-lysine peptide reduced the helical content to 18.5%, about one-third of the value before the substitution. Our simulation produced qualitatively correct results on the order of the helical content, but is less sensitive to the Gly substitution than the experimental study of the alanine-lysine peptides.

In previous studies with various methods,[45,69-72] the order of the helical tendency of different amino acids are not all the same, but the decreasing order for alanine, valine, and glycine is in agreement between the studies. Because our study is a test of the molecular dynamics simulation of peptide folding, only the simple cases that include alanine, valine, and glycine were studied. Their decreasing helical tendencies calculated from the simulations agree with previous studies. A complete study of the helix-forming tendencies of the amino acids will be the subject of another research project. Recently, Okamoto[73] has studied helix-forming tendencies using the Monte Carlo simulated annealing method on homo-oligomers of seven nonpolar amino acids. Our results agree with those of Okamoto on the order of helical tendencies, but our method differs. We use molecular dynamics simulations at normal temperature (274 K). The hydrophobic interactions based on solvent-accessible surface area are included in our simulation. The single amino acid substitutions, instead of the homo-oligomers, are used to compare helical tendencies in our study. The differences arise from our goal to keep the simulation as close to the experimental conditions as possible.

CONCLUSIONS

An important conclusion of our study is that a molecular dynamics method with average solvent effect included in the potential functions can simulate folding of short peptides at experimental temperature. A major difference between this study and many other simulations is that our initial structures are unfolded structures. From either an extended or a randomly generated conformation, the simulations approached a helix-coil equilibrium in about 3 ns. The calculated helical ratio of Ac-$(AAQAA)_3$Y-NH_2 was 47%, which agrees well with the CD measurement (~50%). Unlike our previous rigid-element Monte Carlo studies of the alanine-based peptides,[20,21] in the current molecular dynamics study the peptide backbone is more flexible and all non-hydrogen atoms of the side chains are explicitly represented. During the simulation, the i, i + 3 hydrogen bonds were often observed during α-helix-coil transitions. Other secondary structure elements, such as the β-hairpin, were also observed. A helical segment with frayed ends was the most sta-

ble conformation, but the distorted helix-turn-helix had a non-negligible population in the equilibrium. The transition between the two types of conformations occurred in a much larger time scale than helix propagation. Hydrophobic interaction favors helices more than extended conformations, but compact conformations, such as the helix-turn-helix, could have a lower hydrophobic energy than single helices. The naturally occurring helix-turn-helix motif may have a similar origin in stabilization energies. The transient hydrogen bonds between the glutamine side chain and the backbone carbonyl group could reduce the free energy barrier for helix folding and unfolding. Substituting a single alanine residue in the middle of the peptide with valine or glycine decreased the average helical ratio significantly, in agreement with experimental observations.

REFERENCES

1. Taketomi, H., Ueda, Y., Go, N. Studies on protein folding, unfolding and fluctuations by computer simulation. Int. J. Peptide Protein Res. 7:445–459, 1975.
2. Dill, K. Theory for the folding and stability of globular proteins. Biochemistry 24:1501–1509, 1985.
3. Skolnick, J., Kolinski, A. Simulations of the folding of a globular protein. Science 250:1121–1125, 1990.
4. Shakhnovich, E.I., Gutin, A.M. Influence of point mutations on protein structure: Probability of a neutral mutation. J. Theor. Biol. 149:537–546, 1991.
5. Daggett, V., Kollman, P.A., Kuntz, I.D. Molecular dynamics simulations of small peptides: Dependence on dielectric model and pH. Biopolymers 31:285–304, 1991.
6. Nemethy, G., Scheraga, H.A. Protein folding. Q. Rev. Biophys. 10:239–352, 1977.
7. Ripoll, D.R., Scheraga, H.A. On the multiple-minima problem in the conformational analysis of polypeptides. II. An electrostatically driven Monte Carlo method—tests on poly(l-alanine). Biopolymers 27:1283–1303, 1988.
8. Wilson, S.R., Cui, W. Applications of simulated annealing to peptides. Biopolymers 29:225–235, 1990.
9. Okamoto Y., Fukugita M., Nakazawa T., Kawai H. α-helix folding by Monte Carlo simulated annealing in isolated C-peptide of ribonmuclease A. Protein Eng. 4:639–647, 1991.
10. Carlacci, L., Englander, S.W. The loop problem in proteins: A Monte Carlo simulated annealing approach. Biopolymers, 33:1271–1286, 1993.
11. Kirkpatrick, S., Gelatt, C.D. Jr., Vecchi, M.P. Optimization by simulated annealing. Science 220:671–680, 1983.
12. Brooks, B.R. Molecular dynamics for problems in structural biology. Chem. Scripta 29A:165–169, 1989.
13. Karplus, M., Shakhnovich, E. Protein folding: Theoretical studies of thermodynamics and dynamics. In: "Protein Folding." Creighton, T.E. (eds.). New York: W.H. Freeman, 1992:127–198.
14. McCammon, J.A., Northrup, S.H., Karplus, M., Levy, R.M. Helix-coil transitions in a simple polypeptide model. Biopolymers 19:2033–2045, 1980.
15. Schneller, W., Weaver, D.L. Simulation of α-helix-coil transition in simplified polyvaline: Equilibrium properties and brownian dynamics. Biopolymers 33:1519–1535, 1993.
16. Daggett, V., Kollman, P.A., Kuntz, I.D. A molecular dynamics simulation of equilibrium motions and helix-coil transitions. Biopolymers 31:1115–1134, 1991.
17. Tirado-Rives, J., Maxwell, D.S., Jorgensen, W.L. Molecular dynamics and Monte Carlo simulations favour the α-helical form for alanine-based peptides in water. J. Am. Chem. Soc. 115:11590–11593, 1993.
18. Daggett, V., Levitt, M. Molecular dynamics simulation of helix denaturation. J. Mol. Biol. 223:1121–1138, 1992.
19. Daggett, V., Levitt, M. Protein unfolding pathways explored through molecular dynamics simulations. J. Mol. Biol. 232:600–619, 1993.
20. Sung, S.S. Helix folding simulations with various initial conformations. Biophys. J. 66:1796–1803, 1994.
21. Sung, S.S. Folding simulation of alanine-based peptides with lysine residues. Biophys. J. 68:826–834, 1995.
22. Scholtz, J.M., York, E.J., Stewart, J.M., Baldwin, R.L. A neutral, water-soluble, α-helical peptide: The effect of ionic strength on the helix-coil equilibrium. J. Am. Chem. Soc. 113:5102–5104, 1991.
23. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. Jr., Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. 106:765–784, 1984.
24. Weiner, S.J., Kollman, P.A., Nguyen, D.T., Case, D.A. An all atom force field for simulations of proteins and nucleic acids. J. Comp. Chem. 7:230–252, 1986.
25. Chandler, D., Weeks, J.D., Andersen, H.C. Van der Waals picture of liquids, solids, and phase transitions. Science 200:787–794, 1983.
26. Debye, P. "Polar Molecules." New York: Dover Publications, 1929.
27. Hingerty, B.E., Ritchie, R.H., Ferrell, T.L., Turner, J.E. Dielectric effects in biopolymers: The theory of ionic saturation revisited. Biopolymers 24:427–439, 1985.
28. Ramstein, J., Lavery, R. Energetic coupling between DNA bending and base pair opening. Proc. Natl. Acad. Sci. USA 85:7231–7235, 1988.
29. Okamoto Y. Dependence on the dielectric model and pH in a synthetic helical peptide studied by Monte Carlo simulated annealing. Biopolymers 34:529–539, 1994.
30. Conway, B.E., Bockris, J.O., Ammar, I.A. The dielectric constant of the solution in the diffuse and helmholtz double layers at a charged interface in aqueous solution. Trans. Faraday Soc. 47:756–766, 1951.
31. Eisenberg, D., Mclachlan, A.D. Solvation energy in protein folding and binding. Nature 319:199–203, 1986.
32. Ooi, T., Oobatake, M., Nemethy, G., Scheraga, H.A. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc. Natl. Acad. Sci. USA 84:3086–3090, 1987.
33. Kang, Y.K., Nemethy, G., Scheraga, H.A. Free energies of hydration of solute molecules. 1. Improvement of the hydration shell model by exact computations of overlapping volumes. J. Phys. Chem. 91:4105–4109, 1987.
34. Wesson, L., Eisenberg, D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. Protein Sci. 1:227–235, 1992.
35. Schiffer, C.A., Caldwell, J.W., Kollman, P.A., Stroud, R.M. Protein structure prediction with a combined solvation free energy-molecular mechanics force field. Mol. Simulation 10:121–149, 1993.
36. Wolfenden, R., Andersson, L., Cullis, P.M., Southgate, C.C.B. Affinities of amino acid side chains for solvent water. Biochemistry 20:849–855, 1981.
37. Kyte, J., Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157:105–132, 1982.
38. Sharp, K.A., Nicholls, A., Friedman, R., Honig, B. Extracting hydrophobic free energies from experimental data: Relationship to protein folding and theoretical models. Biochemistry 30:9686–9697, 1991.
39. Fauchere, J.L., Pliska, V. Hydrophobic parameters π of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. Eur. J. Med. Chem. Chim. Ther. 18: 369–375, 1983.
40. Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F., Brice, M.D., Rodger, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. 112: 535–542, 1977.
41. Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J. Comput. Phys. 23:327–341, 1976.
42. Ponder, J.W., Richards, F.M. An efficient newton-like method for molecular mechanics energy minimization of large molecules. J. Comput. Chem. 8:1016–1024, 1987.
43. Kundrot, C.E., Ponder, J.W., Richards, F.M. Algorithms for calculating excluded volume and its derivatives as a

function of molecular conformation and their use in energy minimization. J. Comput. Chem. 12:402–409, 1991.
44. Chakrabartty, A., Schellman, J.A., Baldwin, R.L. Large differences in the helix propensities of alanine and glycine. Nature 351:586–588, 1991.
45. Chakrabartty, A., Kortemme, T., Baldwin, R.L. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. Protein Sci. 3:843–852, 1994.
46. Ferrin, T.E., Huang, C.C., Jarvis, L.E., Langridge, R. The MIDAS display system. J. Mol. Graphics 6:13–37, 1988.
47. Richardson, J.S., Richardson, D.C. Principles and patterns of protein conformation. In: "Prediction of Protein Structure and the Principles of Protein Conformation." Fasman, G.D. (ed.). New York: Plenum, 1989: 1–98.
48. Dyson, H.J., Wright, P.E. Defining solution conformations of small linear peptides. Annu. Rev. Biophys. Chem. 20: 519–538, 1991.
49. Fiori, W.R., Miik, S.M., Millhauser, G.L. Interacting sequence length favors α-helix over 3_{10}-helix in alanine-based peptides: Evidence for a length-dependent structural transition. Biochemistry 32:11957–11962, 1993.
50. Miik, S.M., Martinez, G.V., Fiori, W.R., Todd, A.P., Millhauser, G.L. Short alanine-based peptides may form 3_{10}-helices and not α-helices in aqueous solution. Nature 359: 653–655, 1992.
51. Miik, S.M., Casteel, K.M., Millhauser, G.L. Experimental molecular dynamics of an alanine-based helical peptide determined by spin label electron spin resonance. Biochemistry 32:8014–8021, 1993.
52. Fiori, W.R., Lundberg, K.M., Millhauser, G.L. A single carboxy-terminal arginine determines the amino-terminal helix conformation of an alanine-based peptide. Nature Struct. Biol. 1:374–376, 1994.
53. Abagyan, R. Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J. Mol. Biol. 235:983–1002, 1994.
54. Chothia, C. Hydrophobic bonding and accessible surface area in proteins. Nature 248:338–339, 1974.
55. Presta, L.P., Rose, G.D. Helix signals in proteins. Science 240:1632–1641, 1988.
56. Ooi, T., Oobatake, M. Prediction of the thermodynamics of protein unfolding: The helix coil transition of poly(L-alanine). Proc. Natl. Acad. Sci. USA 88:2859–2863, 1991.
57. Scholtz, J.M., Marqusee, S., Baldwin, R.L., York, E.J., Stewart, J.M., Santoro, M., Bolen, D.W. Calorimetric determination of the enthalpy change for the α-helix to coil transition of an alanine peptide in water. Proc. Natl. Acad. Sci. USA 88:2854–2858, 1991.
58. Scholtz, J.M., Qian, H., Robbins, V.H., Baldwin, R.L. The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. Biochemistry 32:9668–9676, 1993.
59. Schellman, J.A. The factors affecting the stability of hydrogen-bonded polypeptide structures in solution. J. Phys. Chem. 62:1485–1494, 1958.
60. Zimm, B.H., Bragg, J.K. Theory of the phase transition between helix and random coil in polypeptide chains. J. Chem. Phys. 31:526–535, 1959.
61. Lifson, S., Roig, A. On the theory of helix-coil transition in polypeptides. J. Chem. Phys. 34:1963–1974, 1961.
62. Poland, D., Scheraga, H.A. "Theory of Helix-Coil Transitions in Biopolymers." New York: Academic Press, 1970: 1–301.
63. Qian, H., Schellman, J.A. Helix-coil theories: A comparative study for finite length polypeptides. J. Phys. Chem. 96:3987–3994, 1992.
64. Scholtz, J.M., Baldwin, R.L. The mechanism of α-helix formation by peptides. Annu. Rev. Biomol. Struct. 21:95–118, 1992.
65. Scheraga, H.A. On the dominance of short-range interactions in polypeptides and proteins. Pure Appl. Chem. 36: 1–8, 1973.
66. McCammon J.A., Harvey, S.C. "Dynamics of Proteins and Nucleic Acids." Cambridge: Cambridge University Press, 1987.
67. Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D., Chan, H.S. Principles of protein folding—a perspective from simple exact models. Protein Sci. 4:561–602, 1995.
68. Bryngelson, J.D., Onuchic, J.N., Socci, N.D., Wolynes, P.G. Funnels, pathways, and energy landscape of protein folding: A synthesis. Proteins 21:167–195, 1995.
69. Chou, P.Y., Fasman, G.D. Structural and functional role of leucine residues in proteins. J. Mol. Biol. 74:263–281, 1973.
70. Sueki, M., Lee, S., Powers, S.P., Denton, J.B., Konishi, Y., Scheraga, H.A. Helix-coil stability constants for the naturally occurring amino acids in water. 22. Histidine parameters from random poly[(hydroxybutyl)glutamine-co-histidine]. Macromolecules. 17:148–155, 1984.
71. O'Neil, K.T., DeGrado, W.F. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acides. Science 250:646–651, 1990.
72. Lyu, P.C., Liff, M.I., Marky, L.A., Kallenbach, N.R. Side chain contribution to the stability of alpha-helical structure in peptides. Science 250:669–673, 1990.
73. Okamoto, Y. Helix-forming tendencies of nonpolar amino acids predicted by Monte Carlo simulated annealing. Proteins 19:14–23, 1994.