Biophysical Journal Volume 66 June 1994 1796–1803

Helix Folding Simulations with Various Initial Conformations

Shen-Shu Sung
Research Institute, The Cleveland Clinic Foundation, Cleveland Ohio 44195 USA

ABSTRACT Using a solvent-referenced energy calculation, a 16-residue peptide with alanine side chains folded into predominantly α-helical conformations during constant temperature (274 K) simulations. From different initial conformations, helical conformations were reached and the multiple energy minima did not become a serious problem. Under the same conditions, the simulation did not indiscriminately fold a sequence such as polyglycine into stable helices. Interesting observations from the simulations relate to the folding mechanism. The electrostatic interactions between the successive amides favored extended conformations (or β strands) and caused energy barriers to helix folding. β-bends were observed as intermediates during helix nucleation. The helix propagation toward the C-terminus seemed faster than that toward the N-terminus. In helical conformations, hydrogen bond oscillation between the $i,i+4$ and the $i,i+3$ patterns was observed. The $i,i+3$ hydrogen bonds occurred more frequently during helix propagation and deformation near both ends of the helical segment.

INTRODUCTION

The widely applied energy-based calculations, including energy minimization, molecular dynamics, and Monte Carlo simulations, play an important role in protein structure studies. When using the all-atom model or the united-atom model, their application to structural studies is restricted, however, by the computing speed. Consequently, calculating the three-dimensional structure of a protein from its amino acid sequence has not been practical without additional experimental data, such as those obtained with NMR measurements. As far as secondary structure study is concerned, the helix unfolding has been simulated successfully (McCammon et al., 1980; Daggett et al., 1991; Tirado-Rives and Jorgensen, 1991; Daggett and Levitt, 1992). With initial nonhelical conformations, high temperatures or thermal perturbations are usually needed to search for helical structures (Wilson and Cui, 1990; Brooks, 1989; Ripoll and Scheraga, 1988). The difficulty arises primarily from the large number of degrees of freedom in protein molecules. The multidimensional conformational hypersurface is very complicated, with an extraordinarily large number of minima. Consequently, before reaching the global minimum conformation, which presumably corresponds to the native structure, the search process is often trapped in local minima. The well known multiple-minima problem (Nemethy and Scheraga, 1977) is a major obstacle in structural calculations.

Lowering the energy barriers can effectively increase the efficiency of the conformation search. In this study the energy barriers were lowered by including the average solvent effect. The calculation of the solvent-referenced energy was based on the experimentally measured change in enthalpy upon folding. A 16-residue peptide with alanine side chains was simulated at the constant temperature of 274 K. Several simulations were carried out with different initial conformations. Various potential functions were also tested, and a simulation of polyglycine was carried out to test the sequence dependence. The observations related to helix folding mechanism are discussed.

MODEL AND METHODS

This section describes the rigid-element algorithm and a solvent-referenced energy calculation. In the rigid-element algorithm (Sung, 1992, 1993), the amide -CONH- groups are kept in the rigid planar *trans* conformation (except for proline, which is not used in this study) to reduce the degrees of freedom and to include backbone hydrogen bonding atoms explicitly. This treatment has previously been applied in the dihedral angle variable method (Go and Scheraga, 1970; Bruccoleri and Karplus, 1987). The dihedral angle method greatly reduces the degrees of freedom, but it severely restricts local motions in the middle portion of the backbone. In contrast, the rigid element algorithm uses nondihedral variables to facilitate independent local motions in the middle portion of the backbone. The -CONH- groups are flexibly connected to the α-carbon atoms, subject to bond length and bond angle constraints. The backbone motion consists of the motion of the α-carbon trace structure and the rotation of the amide groups about the Cα-Cα pseudobonds. The Cα-Cα distance and N-Cα-C angle constraints keep the bond lengths and bond angles in the vicinity of their equilibrium values. For the first and last residues, the rotations about the Cα-C and Cα-N bonds were included in the algorithm. In this study, alanine side chains were used for all residues. Each side chain contains only a methyl group that is represented by a sphere centered at the β-carbon location, as in the united-atom model. The equilibrium positions of the β-carbon atoms were determined by the backbone conformation. Including the β-carbon atoms does not increase the number of degrees of freedom.

The Metropolis Monte Carlo procedure (Metropolis et al., 1953) was used in the simulation. At each step, only one or two rigid amide elements move and all pairwise interactions that do not include the moved elements are not recalculated. The working hypothesis of the Monte Carlo (MC) method in folding simulation is that the folding is a Markov process with Boltzmann transition probabilities and that the native conformations are stochastically stable with respect to random fluctuations (Li and Scheraga, 1987). Using lattice models, the Monte Carlo method has been used successfully to study protein folding (Sikorski and Skolnick, 1989; Skolnick et al., 1988, 1989).

The rigid-element algorithm explicitly includes all atoms involved in important backbone hydrogen bonding. This approach has considerable advantage in that the well refined atomic parameters of the existing all-atom and united-atom models can be readily applied to backbone atoms without major adjustments. In this study, the parameters of the AMBER force field

Received for publication 24 September 1993 and in final form 25 February 1994.

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

0006-3495/94/06/1796/08 $2.00

(Weiner et al., 1984) were used, including electrostatic interactions, the van der Waals (VDW) interactions, and bond stretching and bending energies. The atomic charges on the C, O, N, and H atoms of the carbonyl and amino groups were 0.52, −0.50, −0.52, and 0.25, respectively. For the sake of simplicity, charges on all α-carbons were set equal to 0.25 and those on side chains were set equal to 0, except for charged residues, which were not used in this study. The atomic charges of the OPLS potential functions (Jorgensen and Tirado-Rives, 1988) were also tested.

The 6–12 potential was used in calculating the VDW interaction for nonhydrogen bond atom pairs. Corresponding to thermal expansion, the thermal motion of the atoms tends to increase the atomic distances above the value obtained in energy minimization calculations. To compensate for this increase, the effective VDW radii of atoms may be scaled down (Levitt, 1983). Reduced VDW radii (to 90%) have also been used to compensate for backbone rigidity (Creamer and Rose, 1992). In the present work, all the simulations were carried out at 274 K, and the effective VDW radii were scaled down to 95% of their original values as given by Weiner et al. (1984). For the hydrogen bonding 10–12 potential, the minimum energy was −0.5 kcal/mol at the distance of 2 Å, as given in the AMBER force field. A test calculation without this 10–12 potential was also carried out.

Solvent, either aqueous or hydrophobic, has very important effects on protein folding (Dill, 1990). However, a folding simulation including hundreds or thousands solvent molecules is too time-consuming. On the other hand, the "in vacuo" calculations give larger energy changes upon folding than experimental measurements have indicated (Daggett et al., 1991; Scholtz et al., 1991). The large energy change helps to form compact conformations, but also creates deep energy minima. Actually, the compact form is mainly the result of hydrophobic interactions (Dill, 1990). In this study, the solvent environment is used as the reference for energy calculations. The solvent-referenced energy calculation, or the solvent-referenced interaction, is an approximate treatment and includes a shifted truncation of the VDW interaction and a rescaled electrostatic interaction. It reduces the calculated energy changes and, therefore, also reduces the energy barriers.

In a biological system, the intramolecular VDW interactions of a protein molecule are balanced by the interactions between the protein molecule and solvent molecules, and also balanced by the interactions among solvent molecules. The short-range repulsive interactions caused by close contact have a major effect on the structural arrangement and need to be calculated explicitly. The longer-range attractive interactions provide a nearly uniform background potential and can serve as the reference for the VDW energy calculation. In this study, the VDW interaction was truncated at the minimum energy distances and shifted so that the minimum energy was zero. This shifted truncation is based on the mean field approximation in the van der Waals theory of the liquid-solid transition (Longuet-Higgins and Widom, 1964; Reiss, 1965; Widom, 1967; Weeks et al., 1971; Chandler et al., 1983). A similar treatment has been previously applied to protein structure studies (McCammon et al., 1980).

Similarly, the electrostatic interaction of the in vacuo calculation needs modification to account for the solvent effect. Among the widely used force fields, various atomic partial charges have been assigned and different effective dielectric constants have been used. The AMBER force field was developed using a distance-dependent dielectric constant, $\epsilon = R$, where R is the distance between two atomic charges, to mimic the polarization effect and to compensate for the lack of explicit solvation. Under certain circumstances, $4R$ is suggested (Weiner et al., 1984). Novotny et al. (1989) also suggested the dielectric constant of $4R$ for their empirical free energy potential. However, when solvent molecules are explicitly included, a constant dielectric is probably more appropriate. The OPLS potential functions were developed for liquid simulations with the dielectric constant of 1 (Jorgensen and Tirado-Rives, 1988). In the ECCEP force field, a constant of 2 is used (Momany et al., 1975). In the CHARMM force field, several different treatments have been proposed, including the distance-dependent dielectric constant (Brooks et al., 1983). It is very difficult to represent the complicated interaction in the protein environment using a single dielectric constant. However, experimentally derived enthalpy changes provide a reference for approximately rescaling the dielectric constant. The enthalpies of helix formation range from 0.9 to 1.3 kcal/mol per residue (Ooi and Oobatake, 1991; Scholtz et al., 1991). With AMBER atomic charges, the dielectric $\epsilon = 2R$ gives reasonable values when compared with these data. Therefore, in this study the effective dielectric constant was set to $2R$, according to the measured enthalpy change. In this way, the average competing effects of the solvent molecules for hydrogen bonding are included approximately. The advantage of using the effective dielectric constant $2R$ to rescale the electrostatic interactions is that the ratios of the atomic charges in the force field are unchanged and the conformation with the lowest electrostatic energy remains the same as that with the dielectric constant R. For different force fields, a different effective dielectric constant could be used according to the value of the enthalpy change.

In this study, alanine side chains were used for all residues. The peptide sequence was Ac-AAAAAAAAAAAAAAAA-NH_2. As a theoretical model, the properties of the different side chains in real helices were not included in the calculations. As a specific calculation for polyalanine, this study may be relevant to the conformations of polyalanine under some particular conditions. The peptide itself is not water soluble. Additional solvation energies were not included in the calculation. All simulations were carried out at a constant temperature of 274 K because many synthetic peptides have been studied at this temperature or at temperatures very close to it (Marqusee et al., 1989; Chakrabartty et al., 1991; Scholtz and Baldwin, 1992). The computations were carried out on a SiliconGraphics Personal Iris 4D35. All of the figures of the peptide conformations were produced using the MidasPlus software system from the Computer Graphics Laboratory, University of California, San Francisco, CA (Ferrin et al., 1988).

RESULTS AND DISCUSSION

Fully extended initial conformation

Several simulations were carried out with various initial conformations. The first simulation started from a fully extended conformation with all the ψ, φ angles equal to 180°. It ran for 120 million (120 M) steps. The potential energies during the simulation are shown in Fig. 1. Selected conformations illustrating the folding process are shown in Fig. 2. The conformation underwent an initial relaxation for less than 1 M steps. At step 2 M, two hydrogen bonds formed between residues 2 and 5 and between residues 3 and 6 (a hydrogen

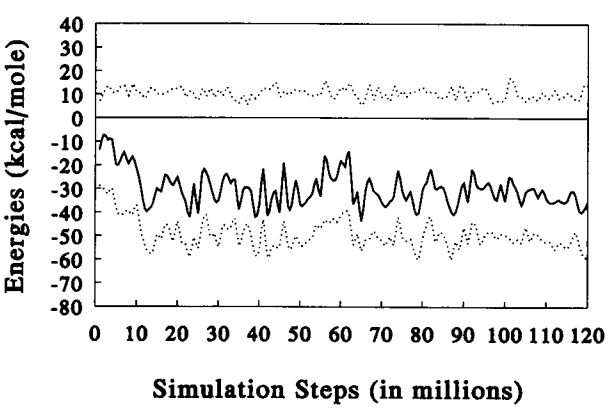

FIGURE 1 The potential energies during the folding simulation with the extended initial conformation. The solid curve represents the total energies, including the nonbonded energies and the bond stretching and bending energies. The lower dotted curve represents the electrostatic energies. The upper dotted curve represents the VDW energies. The total energy drop from step 5 M to step 6 M corresponds to the helix nucleation and the following propagation to residue 11. The energy drop from step 12 M to step 14 M corresponds to the helix propagation to the whole molecule. The energy fluctuations after step 14 M correspond to the equilibrium of the helix-dominated conformations.

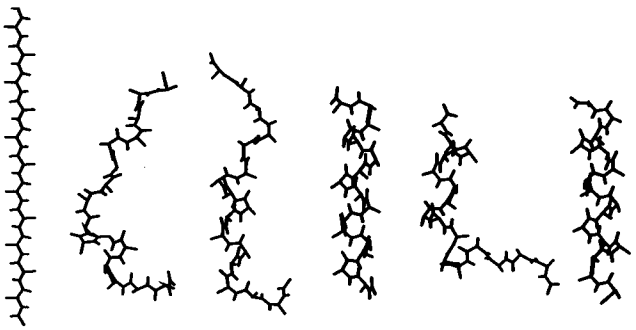

FIGURE 2 The conformations at steps 0, 5 M, 6 M, 14 M, 20 M, and 25 M (left to right) during the folding simulation. The N-terminus is at the bottom of the figure. The starting conformation is a fully extended conformation (first from left, not shown completely). At step 5 M the helical segment started to form (nucleation), at step 6 M it was propagating, and at step 14 M the whole helix formed. The conformation at step 20 M shows the partial deformation of the helix, and that at step 25 M shows the refolding of the helix.

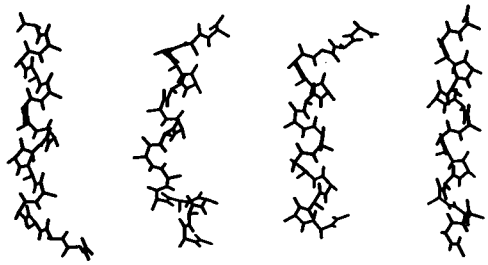

FIGURE 3 The conformations at steps 30 M, 60 M, 90 M, and 120 M during the simulation. The N-terminus is at the bottom of the figure.

bond was assumed when the distance between the carbonyl oxygen and the amino hydrogen was less than 3 Å), but they were not stable. Such hydrogen bonds more often occurred near both ends because there are fewer covalent constraints. However, isolated hydrogen bonds or very short helical segments were often not stable. At steps 3 M and 4 M, the molecule was in a coil conformation without hydrogen bonds. At step 5 M, a turn of helix formed that consisted of three hydrogen bonds between residues 2 and 5, 3 and 6, and 4 and 7. At step 6 M, the helix segment propagated to residue 11. Correspondingly, the total energy reached a minimum at about −20 kcal/mol, as shown in Fig. 1. For very short helical segments, the $i,i+3$ hydrogen bonds often occurred. As the helix propagated, they converted to the $i,i+4$ pattern, but $i,i+3$ hydrogen bonds often remained near both ends. In the following 5 M steps, some hydrogen bonds broke and reformed. From step 12 M to step 14 M, the whole molecule formed an α-helix ($i,i+4$ hydrogen bonds) with 2 to 4 hydrogen bonds in the 3_{10} helix pattern ($i,i+3$) near the ends of the molecule. The electrostatic energy decreased and reached a deeper minimum close to −60 kcal/mol. The total energy reached the minimum at about −40 kcal/mol (Fig. 1).

In the following steps, the overall helical structure remained. Instead of forming a static structure with fixed helical residues and nonhelical residues, all hydrogen bonds were vibrating constantly, and sometimes broke. At step 20 M, the helix had partially deformed (Fig. 2), but at step 25 M a complete helical conformation formed again. Fig. 3 shows the conformations at steps 30 M, 60 M, 90 M, and 120 M. At step 60 M, the C-terminus unfolded and the energies were high. In general, the residues near both ends were more likely to show the $i,i+3$ hydrogen bonding pattern and to unfold. The flexibility of the residues near both ends agreed with experimental observations (Casteel et al., 1993) and theoretical predictions (Poland and Scheraga, 1970). However, the hydrogen bonds in the middle portion of the helix also broke, sometimes. Corresponding to the high peak of −14.1 kcal/mol at step 62 M in Fig. 1, three hydrogen bonds broke in the middle of the helix, as well as two hydrogen bonds near the ends. Different conformations were in equilibrium for α-helix dominated structures. Also, the hydrogen bonds sometimes oscillated between the $i,i+4$ and the $i,i+3$ patterns. Synthetic peptides with a high percentage of helical content may exist in a similar equilibrium (Marqusee et al., 1989; Miik et al., 1992). The total energy of the helix-dominated conformations fluctuated around a constant average value that was 15–20 kcal/mol lower than that of the extended conformation. This energy decrease agreed well with experimental measurements of the enthalpy of helix formation, which range from 0.9 to 1.3 kcal/mol per residue (Scholtz et al., 1991).

Fig. 1 shows many energy barriers in the simulation. To study the electrostatic origin of these energy barriers, a test simulation was carried out without electrostatic interactions between the successive amide units in the primary sequence. In this test, the whole molecule formed a helix within 8 M steps, faster than in the simulation with the successive electrostatic interactions. The total energy decreased more than 30 kcal/mol, and the helix was more stable. The large energy change indicates that the calculations without the electrostatic interactions between successive peptide units were not adequate, but the test suggests an important origin of the energy barrier. In an α-helix, the dipoles of the amide groups are all parallel, whereas in extended conformations or in β strands, they are antiparallel. The long-range electrostatic interactions favor the parallel orientation in helices, but the successive dipole interactions favor the antiparallel orientation in extended conformations or β strands. The successive dipole interactions create an important energy barrier to helix folding. Richardson and Richardson (1990) used the carbonyl oxygen direction as an indication of peptide orientation and used the angle between successive carbonyls as a measure of how the chain twists. Here, the present study showed that the conversion from the antiparallel orientation to the parallel orientation of successive carbonyls has to cross a major energy barrier. Once the orientations were approximately parallel, the helix formed readily.

When the simulation temperature decreased to near 0 K, the potential energy of the helical conformation decreased by about 20 kcal/mol. This energy difference enables the normal

temperature simulation to overcome some barriers that the energy minimization method cannot. The random motion in the MC simulations, representing the entropy effect, drives the molecule to adopt a variety of conformations. Compared with the random coil, the helical conformation allows the ψ, φ angles to vibrate in a much smaller range and is entropically unfavorable. During the folding process, the potential energy decreases at the expense of lowering entropy (Dill, 1990). If the entropy effect dominates, coil conformations occur. The MC simulation does not explicitly calculate the kinetic energy and the entropy, but its average potential energy and structures are temperature-dependent. The entropic effect is implicitly included. The equilibrium of many helical conformations in the simulation is consistent with the fact that proteins are not locked in at incorrectly folded secondary structures. The secondary structures are either stabilized or rearranged into different structures by interaction in the tertiary structure.

Randomly generated initial conformation

In addition to the extended conformation, several other starting conformations were studied. One was generated using random ψ, φ dihedral angles, as shown in Fig. 4. This initial conformation does not have any hydrogen bonds. The simulation ran for 100 M steps. During the first 6 M steps, several unstable hydrogen bonds formed and broke. At step 7 M, a helix turn consisting of three hydrogen bonds formed in the middle portion of the backbone and started to propagate gradually (Fig. 4). At step 8 M, the helical segment contained six hydrogen bonds. From step 15 M, helical conformations dominated. The conformations after step 15 M were similar to those shown in Fig. 3. For example, at step 20 M, the α-helix became partially distorted with three hydrogen bonds in the $i,i+3$ pattern and one hydrogen bond broken between the oxygen of residue 11 and the hydrogen of residue 15 (Fig. 4). The energy changes during the simulation were similar to those shown in Fig. 1. Other random conformations and conformations containing hydrogen bonds were used as initial conformations in different simulations and stable helices also formed.

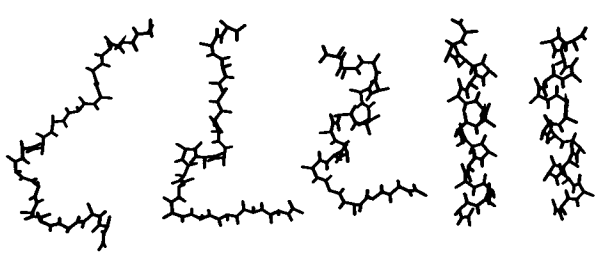

FIGURE 4 The conformations at steps 0, 7 M, 8 M, 15 M, and 20 M (left to right) during the simulation with the random initial conformation (at step 0). The N-terminus is at the bottom of the figure. The conformation at step 7 M shows the nucleation of the helix; that at step 8 M shows the propagation process; that at step 15 M shows the formation of a helix; that at step 20 M shows the partial distortion of the helix near the C-terminus.

Left-handed helical initial conformation

In a previous study of polyalanine (Ripoll and Scheraga, 1988), the conformation search starting from the left-handed helix needed the most extensive runs. Therefore, the artificially folded left-handed helix was tested in this study as a starting conformation. Fig. 5 shows the potential energies during the simulation, and Fig. 6 shows selected conformations that illustrate the structural changes. Left-handed helices rarely occur in nature unless they are very short because of the close approach of the β carbon and the carbonyl oxygen of the naturally occurring L-amino acids. The strong repulsive VDW interaction between the β-carbon and the carbonyl oxygen makes the left-handed helix unstable.

During the first 12 M steps, two hydrogen bonds at the N-terminus broke and some $i,i+4$ hydrogen bonds converted to the $i,i+3$ pattern. From step 13 M to step 17 M, five hydrogen bonds at the N-terminus broke. At step 18 M, the rest of the left-handed helix unfolded. Several hydrogen bonds were left, but no clear helical conformation existed. By step 20 M, all hydrogen bonds had broken. During this period, the VDW energy decreased greatly (Fig. 5), because the β-carbon and the carbonyl oxygen were no longer in close contact; the electrostatic energy increased greatly because the hydrogen bonds were broken. This large increase in electrostatic energy is not common. It is caused by the artificially folded starting conformation. Between step 20 M and step 55 M, isolated hydrogen bonds formed and broke many times, and some right-handed helix segments also formed and unfolded. At step 55 M, the molecule was in a coil state without any hydrogen bonds. Both the electrostatic energy and the

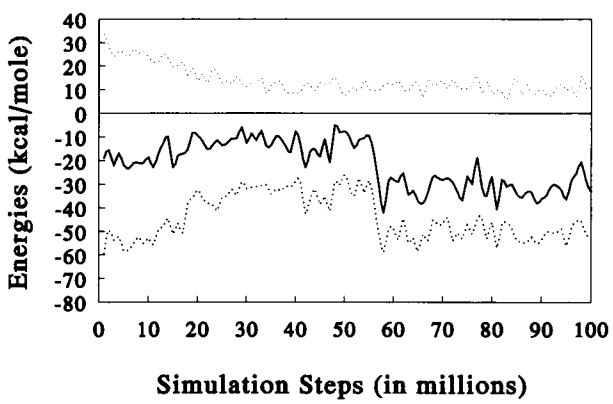

FIGURE 5 The potential energies during the folding simulation started with the left-handed helix. The solid curve represents the total energies. The lower dotted curve represents the electrostatic energies. The upper dotted curve represents the VDW energies. During the first 20 M steps, the unfolding of the left-handed helix greatly increased the electrostatic energy because the hydrogen bonds were broken. In contrast, the VDW energy decreased greatly because the β carbons and the carbonyl oxygens were no longer in close contact. As the result of the two opposing contributions, the total energy increased when the left-handed helix unfolded. The sharp energy drop in both total energies and electrostatic energies from step 55 M to step 58 M corresponds to the formation of the right-handed helix. The energy fluctuations after step 60 M correspond to the equilibrium of the right-handed helix-dominated conformations.

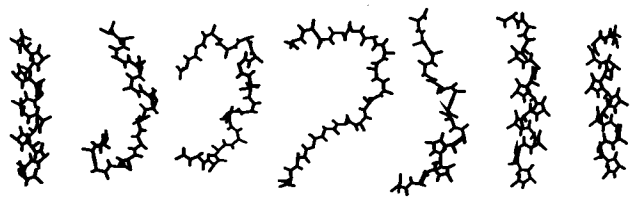

FIGURE 6 The conformations at steps 0, 17 M, 20 M, 55 M, 56 M, 57 M, and 58 M (left to right) during the simulation started from the left-handed helix (at step 0). The N-terminus is at the bottom of the figure. The conformation at step 17 M shows the partially unfolded left-handed helix; that at step 20 M shows the completely unfolded structure; that at step 55 M shows the coil structure before folding into the right-handed helix; those at steps 56 M, 57 M, and 58 M show the nucleation and the propagation of the right-handed helix.

total energy were high (Fig. 5). At step 56 M, four hydrogen bonds in the right-handed helical pattern formed in the N-terminal half of the molecule. From step 56 M to step 58 M, the whole molecule quickly formed a right-handed helix, and both the electrostatic energy and the total energy decreased sharply. After step 58 M, the right-handed helical conformations dominated until the end of the simulation (total 100 M steps). The energy fluctuated around a stable average value of −30 kcal/mol after step 60 M. During the unfolding of the left-handed helix, the total energy increased by about 10 kcal/mol. This long-lasting (tens of millions of steps) potential energy barrier was overcome. During the whole simulation, the total energy decreased by about 10 kcal/mol.

During these simulations, β-bends were often observed as intermediates during the helix nucleation. It has been postulated that the β-bends play an important role in the initiation of protein folding (Lewis et al., 1971; Zimmerman and Scheraga, 1977). Recently, the reverse turns (β-bend structure) were proposed as important intermediates in helix folding on the bases of separate studies of protein structure analysis (Sundaralingam and Sekharudu, 1989) and of molecular dynamics of tripeptides in water (Tobias and Brooks, 1991). In the current study, the β-bend intermediates were directly observed during the helix nucleation. One or two hydrogen bonds formed and broke many times before forming a stable helical segment. Three or more successive hydrogen bonds were more likely to propagate to form a longer helical segment. In addition to the β-bend structure, $i,i+3$ hydrogen bonding also occurred in helical segments. Sometimes, the oxygen of residue i oscillated between the hydrogens of residue $i+4$ and of residue $i+3$ in a helical segment. The $i,i+3$ hydrogen bonding occurred more frequently during the helix propagation or deformation process near both ends of the helical segments. This observation is qualitatively consistent with the experimental study on the 3_{10} helix population by Miik et al. (1992) and with the helix unfolding simulations of Tirado-Rives and Jorgensen that included water molecules explicitly (1991).

During the simulations, helix propagation seemed faster toward the C-terminus than toward the N-terminus. In addition to the differences in atomic sizes and in bond lengths between the CO and NH groups, the difference in partial charges also makes the electrical field in the C end of the helical segment stronger than that in the N end. The dipoles of the peptides following the C end of the helical segment are aligned with the helix faster than those preceding the N end of the helical segment. Because the random motion tends to make the propagation probabilities equal, the difference in the propagation rates was small. Naturally occurring peptides and proteins are not homopolymers, so their helix propagation would be more complicated than polyalanine.

Tests on different potentials and sequences

With the solvent-referenced interaction, the calculation successfully found stable helical conformations. To determine how sensitive the results were to the force field parameters, several tests were conducted. The AMBER force field contains a special 10–12 potential for hydrogen bonding, and a maximum of 0.5 kcal/mol stabilization energy is given to each hydrogen bond in addition to the electrostatic interaction. In a test, the 10–12 interaction was deleted, and the hydrogen bonding interaction became purely electrostatic. The simulation was carried out at 274 K with the extended initial conformation. All other conditions were the same as in the simulation shown in Fig. 1. Fig. 7 shows the energy changes during the simulation. The low energy conformation at step 15 M is an α-helix with 14 residues (Fig. 8), the first and the last residues were not in the helical conformation. This helical conformation unfolded in the subsequent steps. For example, at step 54 M, the conformation did not have any hydrogen bonds. At step 71 M, a helix formed again. Various helical and partial helical conformations were in equilibrium in the steps that followed until the simulation ended at step 100 M (Fig. 8). Without the 10–12 interaction, the total energy of the helical conformations was 5–10 kcal/mol higher than that in Fig. 1, and the helical conformations were less

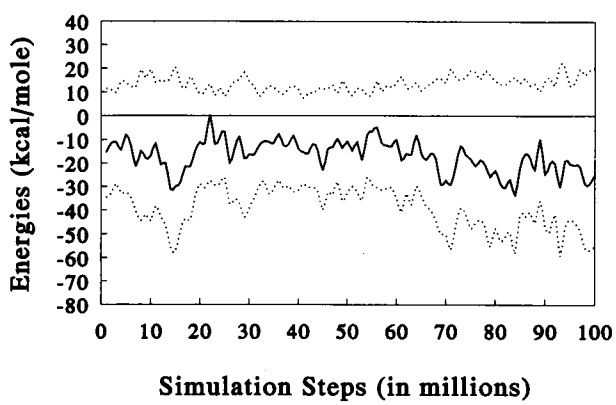

FIGURE 7 The potential energies during the folding simulation without the 10–12 hydrogen bonding interaction. The solid curve represents the total energies. The lower dotted curve represents the electrostatic energies. The upper dotted curve represents the VDW energies. At step 15 M the helical conformation formed. In the following higher energy region, coil conformations dominated. From step 71 M, helical conformations dominated.

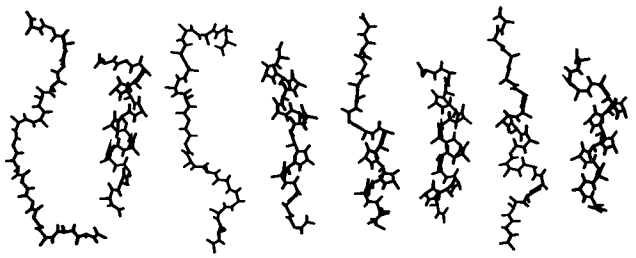

FIGURE 8 The conformations at steps 3 M, 15 M, 54 M, 71 M, 73 M, 84 M, 89 M, and 93 M (left to right) during the simulation without the 10–12 hydrogen bonding interaction. The N-terminus is at the bottom of the figure. The conformation at step 15 M shows the first helical conformation. The conformation at step 54 shows a coil conformation. The conformations from step 71 to step 93 show the helix-dominated conformations in equilibrium.

stable. The special hydrogen bonding potential stabilizes helical conformation, but it is not a necessary condition for the helix folding. Unlike the previous examples, during this simulation both helix folding and unfolding were observed. The low energy states corresponding to the helical conformations were substantially populated, whereas the transitions from the low energy states to high energy coil conformations were observed within the simulation time period. The quantitative distribution of the different states may not correspond to real systems, but it serves as a conformation search example.

To show the importance of the solvent-referenced interaction, simulations without the reference were also carried out for comparison. The unadjusted dielectric constants produce qualitatively different results. Dielectric constants of $\epsilon = 1$ and $\epsilon = 2$ were tested under the same conditions as the simulation in Fig. 1. In both cases, the conformations remained largely extended. Isolated hydrogen bonds formed and broke at times near both ends. In these simulations, the electrostatic interactions were stronger than with the distance-dependent dielectric. The energy barriers caused by the neighboring peptide dipole interactions may have prevented the formation of helical conformations, as discussed earlier. The effective dielectric $\epsilon = R$ was also tested with the same conditions as those in the simulation shown in Fig. 1. This distance-dependent dielectric has taken into account some solvent effects, but was not adjusted according to experimental enthalpy changes. The extended conformation folded into a helix, but the average of potential energy change was about 3 kcal/mol per residue, much larger than the experimental value. The large energy change could cause the multiple-minima problem with other initial conformations. A test on charge parameters was carried out using the charges of the OPLS potential functions (Jorgensen and Tirado-Rives, 1988). All other conditions remained the same as in the simulation shown in Fig. 7. Starting with the fully extended conformation, stable helices formed. However, the OPLS potential function was developed for simulations including water molecules, the same treatment is not expected to work well in all cases. With different initial conformation helices may not form.

The tests on different potentials showed the limitations of the algorithm and the importance of the solvent-referenced interaction. Like other MC methods, the algorithm did not mathematically solve the multiple-minima problem. It cannot cross the high energy barriers in the simulation with $\epsilon = 1$ or $\epsilon = 2$ within reasonable simulation time. The rigid-element algorithm is designed to simplify the calculation for conformation changes. It does not change the overall energy landscape, except for details related to the internal motion of the amide groups. In MC simulations, the transition probability, $\exp(-\Delta E/kT)$, is determined by the barrier height, ΔE. Reducing the energy barrier by one-half will increase the transition probability, e.g., from 1/10000 to 1/100. In other words, an energy barrier that previously needed 10000 steps to cross needs only 100 steps after lowering the barrier by a half. Lowering energy barriers more effectively increases efficiency in searching the global minimum than simplifying the algorithm. The solvent-referenced interaction lowers the energy barriers by truncating the attractive VDW interaction and rescaling the dielectric constant according to experimental enthalpy changes. The efficiency in reaching the global minimum mainly results form the solvent-referenced interaction.

However, lowering energy barriers may destabilize folded structures. The simulated annealing method uses high temperature to overcome energy barriers and low temperature to stabilize the low energy conformations (Kirkpatrick et al., 1983; Wilson and Cui, 1990). Without using artificial temperatures, finding a simple potential function to represent the complicated interactions in proteins is a very important task. This simple potential should be neither so strong as to make the energy barrier very high nor so weak as to destabilize the folded structure. The current study demonstrated the existence of such a proper potential function for helix folding at constant temperature. It includes the average solvent effect, and its results are consistent with experimental enthalpy data. The concept of including the average solvent effect in the energy reference is important. The specific treatment in this study has limitations. It has not been optimized to quantitatively reproduce experimentally observed properties. Different truncation distances for VDW interactions and other forms of the dielectric constant may give even better results. This method has not been tested on other secondary structures, such as β-sheets, and the structures of naturally occurring peptides and proteins. The solvent-referenced interactions included the average effect of the interaction with solvent. The differences in solvation energies of different residues were not included.

Helix folding depends on the amino acid sequence, so tests on different sequences are critical. Glycine does not have nonhydrogen side-chain atoms. A calculation on polyglycine could test whether the method forces the helix formation of the backbone atoms. A 16-residue polyglycine was simulated under the same conditions as those for polyalanine shown in Fig. 1. The initial conformation was fully extended and the simulation was carried out at 274 K. As shown in Fig. 9, the energies fluctuated around −20 kcal/mol and did not decrease

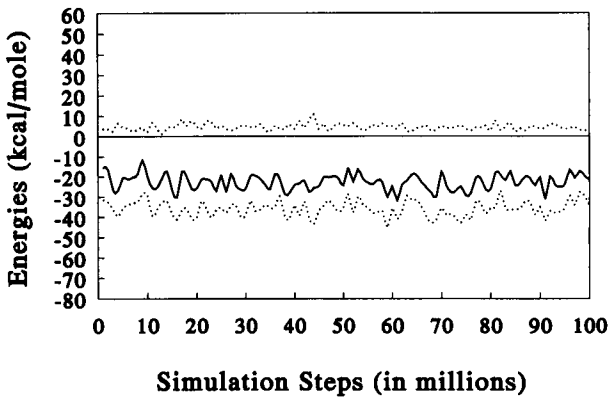

FIGURE 9 The potential energies during the folding simulation of polyglycine. The middle solid curve represents the total energies. The lower dotted curve represents the electrostatic energies. The upper dotted curve represents the VDW energies. The energies fluctuated, and the average value did not decrease.

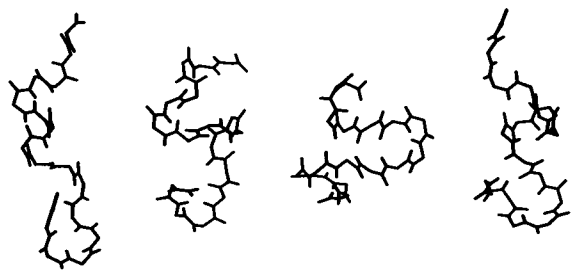

FIGURE 10 The low energy conformations at steps 39 M, 44 M, 59 M, and 91 M (left to right) during the simulation of the polyglycine. The N-terminus is at the bottom of the figure. At step 39 M, a small segment of helix formed in the middle of the backbone, but was unstable. At step 59 M, two antiparallel β-strands with four hydrogen bonds formed, but were unstable. The molecule showed various coil structures and did not form stable helices.

on average during the simulation. Stable helical conformations did not form (Fig. 10). The potential energies of the helix-dominated conformations in Fig. 1 are about −30 kcal/mol. Because in the simulation the alanine side chain was not charged and the attractive VDW interaction was truncated, the side chains did not directly contribute any negative potential energies to the helix stabilization. Polyglycine should have potential energy lower than −30 kcal/mol if it folds into helical conformations. Why did it not fold into helices? As mentioned before, during folding the potential energy decreases at the expense of lowering entropy. The methyl groups of alanine restrict the motion of the backbone in the coil conformations, and the entropy cost for helix folding is low. Random coils of polyglycine have much more freedom, so the entropy cost for folding is high. Fig. 10 shows four low energy conformations during the simulation. Many unstable hydrogen bonds formed and broke. Various conformations were accessible, including transient helical segments (e.g., at step 39 M) and β-strands (e.g., at step 59 M). Energy barriers do not seem to be the reason that a stable helix did not form. The reason that polyglycine did not fold into a helix during the simulation is mainly entropic. In a real system, the different solvation properties may also play an important role. This simulation clearly showed that the method itself does not force the formation of helices.

SUMMARY

Helical conformations were successfully reached with a variety of initial conformations, including the well defined extended conformation, the randomly generated conformation, and the artificially folded left-handed helix. Without using high temperatures or thermal perturbations, the multiple-minima problem did not hinder helix formation. The solvent-referenced interaction, which includes the average solvent effect, was the most important factor in the helix folding simulations. Without using the solvent-referenced energy calculation, helices did not form with dielectric constants $\epsilon = 1$ and $\epsilon = 2$. The simulation does not indiscriminately fold a sequence such as polyglycine into helices. Interesting observations relate to the folding mechanism. The electrostatic interactions between the successive amides in the primary sequence favor extended conformations or β-strands and create energy barriers to helix folding. β-bends were observed as intermediates during helix nucleation, which agrees with other studies (Lewis et al., 1971; Zimmerman and Scheraga, 1977; Sundaralingam and Sekharudu, 1989; Tobias and Brooks, 1991). However, isolated β-bends were not always stable, and their hydrogen bonds could break before helix segments formed. Three or more successive hydrogen bonds in the helical pattern were more likely to propagate to form longer helical segments. For the homopolymer of alanine, the propagation toward the C-terminus seemed faster than toward the N-terminus, but the difference was small. In helical conformations, hydrogen bond oscillation between the $i,i+4$ and the $i,i+3$ patterns was observed. The $i,i+3$ hydrogen bonding occurred more frequently during helix propagation or deformation near both ends of the helix segment, which agrees with those calculations that explicitly include solvent molecules (Tirado-Rives and Jorgensen, 1991).

REFERENCES

Brooks, B. R. 1989. Molecular dynamics for problems in structural biology. *Chemica Scripta*. 29A:165–169.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Compt. Chem.* 4:187–217.

Bruccoleri, R. E., and M. Karplus. 1987. Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers*. 26:137–168.

Casteel, K. M., S. M. Miick, and G. L. Millhauser. 1993. C- and N-terminus helical asymmetry—a clue to forces involved in protein folding? *Biophys. J.* 64:378*a*. (Abstr.)

Chakrabartty, A., J. A. Schellman, and R. L. Baldwin 1991. Large differences in the helix propensities of alanine and glycine. *Nature*. 351:586–588.

Chandler, D., Weeks, J. D., and Andersen, H. C. 1983. Van der Waals picture of liquids, solids, and phase transformations. *Science*. 200:787–794.

Creamer, T. P., and G. D. Rose. 1992. Side-chain entropy opposes α-helix

formation but rationalizes experimentally determined helix-forming propensities. *Proc. Natl. Acad. Sci. USA.* 89:5937–5941.

Daggett, V., P. A. Kollman, and I. D. Kuntz. 1991. A molecular dynamics simulation of polyalanine: an analysis of equilibrium motions and helix-coil transitions. *Biopolymers.* 31:1115–1134.

Daggett, V., and M. Levitt 1992. Molecular dynamics simulations of helix denaturation. *J. Mol. Biol.* 223:1121–1138.

Dill, K. A. 1990. Dominant forces in protein folding. *Biochemistry.* 29: 7133–7155.

Ferrin, T. E., C. C. Huang, L. E. Jarvis, and R. Langridge. 1988. The MIDAS display system. *J. Mol. Graphics.* 6:13–37.

Go, N., and H. A. Scheraga. 1970. Ring closure and local conformational deformations of chain molecules. *Macromolecules.* 3:178–187.

Jorgensen, W. L., and Tirado-Rives, J. 1988. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Amer. Chem. Soc.* 110:1657–1666.

Kirkpatrick, S., C. D. Gelatt, Jr., and M. P. Vecchi. 1983. Optimization by simulated annealing. *Science.* 220:671–680.

Levitt, M. 1983. Molecular dynamics of native protein. *J. Mol. Biol.* 168: 595–620.

Lewis, P. N., F. A. Momany, and H. A. Scheraga. 1971. Folding of polypeptide chains in proteins: a proposed mechanism for folding. *Proc. Natl. Acad. Sci. USA.* 68:2293–2297.

Li, Z., and H. A. Scheraga. 1987. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. USA.* 84:6611–6615.

Longuet-Higgins, H. C., and B. Widom. 1964. A rigid sphere model for the melting of argon. *Mol. Phys.* 8:549–556.

Marqusee, S., V. H. Robbins, and R. L. Baldwin. 1989. Unusually stable helix formation in short alanine-based peptides. *Proc. Natl. Acad. Sci. USA,* 86:5286–5290.

McCammon, J. A., S. H. Northrup, M. Karplus, and R. M. Levy. 1980. Helix-coil transitions in a simple polypeptide model. *Biopolymers.* 19: 2033–2045.

Metropolis N., A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1092.

Miik, S. M., G. V. Martinez, W. R. Fiori, A. P. Todd, and G. L. Millhauser. 1992. Short alanine-based peptides may form 3_{10}-helices and not α-helices in aqueous solution. *Nature.* 359:653–655.

Momany, F. A., R. F. McGuire, A. W. Burgess, and H. A. Scheraga. 1975. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *J. Phys. Chem.* 79:2361–2381.

Nemethy, G., and H. A. Scheraga. 1977. Protein folding. *Q. Rev. Biophys.* 10:239–352.

Novotny, J., R. E. Bruccoleri., and F. A. Saul. 1989. On the attribution of binding energy in antigen-antibody complexes McPC603, D1.3, and Hy-HEL-5. *Biochemistry.* 28:4735–4749.

Ooi, T., and M. Oobatake. 1991. Prediction of the thermodynamics of protein unfolding: The helix coil transition of poly(L-alanine). *Proc. Natl. Acad. Sci. USA.* 88:2859–2863.

Poland, D., and H. A. Scheraga. 1970. Theory of helix-coil transitions in biopolymers. Academic Press, New York. 71 pp.

Reiss, H. 1965. Scaled particle methods in the statistical thermodynamics of fluids. *Adv. Chem. Phys.* 9:1–84.

Richardson, J. S., and D. C. Richardson. 1990. The origami of proteins. *In* Protein Folding. L. M. Gierasch and J. King, editors. AAAS. Washington, D.C. 5–17.

Ripoll, D. R., and H. A. Scheraga. 1988. 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.

Scholtz, J. M., and R. L. Baldwin. 1992. The mechanism of α-helix formation by peptides. *Annu. Rev. Biophys. Biomol. Struct.* 21:95–118.

Scholtz, J. M., S. Marqusee, R. L. Baldwin, E. J. York, J. M. Stewart, M. Santoro, and D. W. Bolen. 1991. 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.

Sikorski, A., and J. Skolnick. 1989. Monte Carlo simulation of equilibrium globular protein folding: α-helical bundles with long loops. *Proc. Natl. Acad. Sci. USA.* 86:2668–2672.

Skolnick, J., A. Kolinski, and R. Yaris. 1988. Monte Carlo simulations of the folding of β-barrel globular proteins. *Proc. Natl. Acad. Sci. USA.* 85:5057–5061.

Skolnick, J., A. Kolinski, and R. Yaris. 1989. Dynamic Monte Carlo study of the folding of a six-stranded Greek key globular protein. *Proc. Natl. Acad. Sci. USA.* 86:1229–1233.

Sundaralingam, M., and Y. C. Sekharudu. 1989. Water-inserted α-helical segments implicate reverse turns as folding intermediates. *Science.* 244: 1333–1337.

Sung, S.-S. 1992. A new rigid body model of protein conformations. *Biophys. J.* 61:348*a*. (Abstr.)

Sung, S.-S. 1993. Helix folding simulations using a new rigid element method. *Biophys. J.* 64:245*a*. (Abstr.)

Tirado-Rives, J., and W. L. Jorgensen. 1991. Molecular dynamics simulations of the unfolding of an α-helical analogue of ribonuclease A S-peptide in water. *Biochemistry.* 30:3864–3871.

Tobias, D. J., and C. L. Brooks III. 1991. Thermodynamics and mechanism of α helix initiation in alanine and valine peptides. *Biochemistry.* 30: 6059–6070.

Weeks, J. D., D. Chandler, and H. C. Anderson. 1971. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.* 54:5237–5246.

Weiner, S. J., P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., and P. Weiner. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Amer. Chem. Soc.* 106:765–784.

Widom, B. 1967. Intermolecular forces and the nature of the liquid state. *Science.* 157:375–382.

Wilson, S. R., and W. Cui. 1990. Applications of simulated annealing to peptides. *Biopolymers.* 29:225–235.

Zimmerman, S. S., and H. A. Scheraga. 1977. Local interactions in bends of proteins. *Proc. Natl. Acad. Sci. USA.* 74:4126–4129.