Folding Simulations of Alanine-Based Peptides with Lysine Residues

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

ABSTRACT The folding of short alanine-based peptides with different numbers of lysine residues is simulated at constant temperature (274 K) using the rigid-element Monte Carlo method. The solvent-referenced potential has prevented the multiple-minima problem in helix folding. From various initial structures, the peptides with three lysine residues fold into helix-dominated conformations with the calculated average helicity in the range of 60–80%. The peptide with six lysine residues shows only 8–14% helicity. These results agree well with experimental observations. The intramolecular electrostatic interaction of the charged lysine side chains and their electrostatic hydration destabilize the helical conformations of the peptide with six lysine residues, whereas these effects on the peptides with three lysine residues are small. The simulations provide insight into the helix-folding mechanism, including the β-bend intermediate in helix initiation, the (i, i + 3) hydrogen bonds, the asymmetrical helix propagation, and the asymmetrical helicities in the N- and C-terminal regions. These findings are consistent with previous studies.

INTRODUCTION

Several de novo designed peptides form helical conformations in water (reviewed by Scholtz and Baldwin, 1992). The folding of these short peptides is also a good subject for theoretical and computational studies. However, folding simulations of these peptides have not been reported. The difficulty arises primarily from the large number of degrees of freedom in the peptide-solvent system and the multiple minima on the energy hypersurface (Nemethy and Scheraga, 1977). The helix unfolding has been simulated successfully (McCammon et al., 1980; Daggett et al., 1991b; Tirado-Rives and Jorgensen, 1991; Daggett and Levitt, 1992), but helix-folding calculations (with nonhelical initial conformations) usually need high temperatures or thermal perturbations to search for helical structures (Ripoll and Scheraga, 1988; Brooks, 1989; Wilson and Cui, 1990; Kawai et al., 1991; Okamoto, 1994).

Using a solvent-referenced potential and the rigid-element algorithm, the folding of a 16-residue model peptide with alanine side chains has been successfully simulated at an experimentally relevant temperature (274 K) in a previous study (Sung, 1994). With different initial conformations, including the well defined extended conformation, the randomly generated conformation, and the artificially folded, left-handed helix, the peptide folded into predominantly α-helical conformations. Because the energy barriers were lowered by including the average solvent effect in the solvent-referenced potential, the multiple-minima problem did not prevent helix formation. For a homopolymer, such as polyalanine, the potential representing the average properties of the residues worked well. However, de novo designed peptides are usually heteropolymers. Their structures depend on the amino acid sequence. In addition to the average solvent effect, the specific solvation properties of the various residues in the sequence must be included in the calculation. In the present study of the alanine-based peptides with lysine residues, different hydrophobic effects and electrostatic hydrations of alanine and lysine residues are included.

Marqusee et al. (1989) showed that 16-residue, alanine-based peptides with different numbers of lysine residues form monomeric conformations with various helical contents in aqueous solutions. Based on Circular Dichroism (CD) measurements, the peptide 3K(I), Ac-AAAAKAAAAKA-AAAKA-NH_2, and the peptide 3K(II), Ac-AKAAAAK-AAAAKAAAA-NH_2, showed up to 80% helical content. The peptide 6K(I), Ac-AKAAKAKAAKAKAAKA-NH_2, showed only 15% helical content. Because the helical content differs greatly, a simple model calculation may provide a rationale for this experimental observation. In the present study the folding simulations of 3K(I), 3K(II), and 6K(I) are carried out. The calculated helicities (the ratio of the number of helical conformations to the total number of conformations) are compared with experimental results.

MODEL AND METHODS

Rigid-element algorithm

The rigid-element algorithm and the solvent-referenced energy calculation have been described in detail previously (Sung, 1994). Their main features are briefly mentioned here. 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. The —CONH— groups are flexibly connected to the α-carbon atoms, subject to bond length and bond angle constraints. The Metropolis Monte Carlo (MC) procedure (Metropolis et al., 1953) is

Received for publication 15 September 1994 and in final form 22 November 1994.

Address reprint requests to Dr. Shen-Shu Sung, Research Institute FF3, Cleveland Clinic Foundation, 9500 Euclid Ave., Cleveland, OH 44195. Tel.: 216-444-0056; Fax: 216-444-9263; E-mail: sung@iris1.hh.ri.ccf.org.

0006-3495/95/03/826/09 $2.00

used in the simulation. The backbone motion consists of the motion of the α-carbon trace structure and the rotation of the amide groups about the C_α—C_α pseudo-bonds. 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 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). With lattice models, the MC method has been successfully used to study protein folding (Sikorski and Skolnick, 1989; Skolnick et al., 1988, 1989).

The parameters of the AMBER force field (Weiner et al., 1984) are used for backbone and β-carbon atoms. The alanine side chain is represented by a sphere centered at the C_β location, as in the united-atom model. The 6–12 potential is used in calculating the van der Waals (VDW) interaction for nonhydrogen bond atom pairs. To compensate for the thermal expansion and peptide rigidity, the effective VDW radii of atoms are scaled down to 95% of their original values as given by Weiner et al. (1984). Similar treatments have been used by others (Levitt, 1983; Creamer and Rose, 1992). In addition to the C_β atom, the lysine side-chain atoms are represented by an interaction center with a spherical potential. The center of the sphere is allowed to move within 3.2 Å from the C_β atom, which is the distance between the C_β atom and the geometrical center of the rest of the side-chain atoms in the fully extended conformation. The 6–8 potential suggested by Levitt (1976), is used for the interaction center. The radius for minimum energy of the 6–8 potential is 2.94 Å, calculated from the lysine side-chain volume (Gerstein et al., 1994). The minimum energy of the 6–8 potential is 0.16 kcal/mol, calculated according to the formula given by Levitt (1976). As in the AMBER force field, the 10–12 potential is used for hydrogen-bonding atom pairs, and the minimum energy is −0.5 kcal/mol at the distance of 2 Å.

Solvent-referenced potential

Because the explicit inclusion of water molecules sharply increases the computation load, a solvent-referenced potential was proposed to use the average solvent effect as the reference for energy calculations (Sung, 1994). This approximate treatment includes a shifted truncation of the VDW interaction and a rescaled electrostatic interaction. In a biological system, the intramolecular VDW interactions of a protein molecule are balanced by the interactions with solvent molecules. The longer range attractive interactions provide a nearly uniform background potential and can serve as the reference for the VDW energy calculation. The short range repulsive interactions have a major effect on the structural arrangement and thus need to be calculated explicitly. Therefore, the VDW interaction is truncated at the minimum energy distances and shifted so that the minimum energy is 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 applied previously to protein structure studies (McCammon et al., 1980).

When solvent molecules are not included explicitly, the electrostatic interactions are usually exaggerated (Daggett et al., 1991a; Daggett and Levitt, 1993). Therefore, the electrostatic interaction may be scaled down to account for the solvent effect. According to the enthalpies of helix formation (in the range from 0.9 to 1.3 kcal/mol per residue, see Ooi and Oobatake, 1991; Scholtz et al., 1991), the dielectric $\epsilon = 2R$ is used for AMBER partial charges of the backbone atoms. Thus, the average competing effects of the solvent molecules for hydrogen bonding are included approximately. The advantage of using $\epsilon = 2R$ 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 $\epsilon = R$. For other force fields, a different effective dielectric constant could be used according to the value of the enthalpy change. In a previous study (Sung, 1994), other values of the dielectric constants were tested and their effects were discussed. However, because the lysine side chain is charged and usually stays on the protein surface in contact with water molecules, a different dielectric constant from that for backbone should be used for interactions involving the charged side chains (Warshel and Aqvist, 1991). A high effective dielectric constant (40–100) was suggested for long range electrostatic interactions at low ionic strength between charges on the protein surface (Loewenthal et al., 1993). In the current study, the high effective dielectric constant of 40 is used for interactions of lysine with other groups, and a dielectric constant of 80 is tested for comparison.

Hydrophobic effect and electrostatic hydration

In a previous study (Sung, 1994), alanine side chains were used for all residues in helix-folding simulations. As a theoretical model, the different side-chain properties in real helices were not included. When different sequences are compared, however, the different solvation properties of the residues must be included. The solvation energy consists of several contributions: the entropic effect of the solvent surrounding the protein (the hydrophobic effect), the electrostatic interaction of the charged atoms of the protein molecule with the solvent (the electrostatic hydration), the effect of the thermal motion of solvent molecules, and the specific interactions, such as hydrogen bonding between water and protein atoms. How to include the solvation energy in the molecular mechanics calculations has become the focus of many recent studies. Several methods have been proposed (van Gunsteren et al., 1994). In the current study, the hydrophobic effect and the electrostatic hydration are included using simple approximations to make folding simulations feasible. Other contributions to solvation, such as the specific hydrogen-bonding effect, are not included.

The surface area-based methods are widely used in calculations of the hydrophobic effect. These methods assign a solvation parameter to each atom and then multiply this parameter by the solvent-accessible surface area to calculate a free energy of solvation (Eisenberg and McLachlan, 1986; Eisenberg et al., 1989). However, the surface area calculation is very time-consuming and computationally expensive. Therefore, it is not used in this study, except the approximation proposed by Wodack and Janin (1980), which is tested for comparison.

An exponential decay of the macroscopic hydrophobic forces has been observed, and its application to molecular interactions seems reasonable (Isrealachvili and Pashley, 1982; Pashley et al., 1985). Fauchere et al. (1988) used the pairwise function with exponential decay to estimate the effect of the solvent. Vila et al. (1991) also tested this function and suggested its use in rough conformation searches. Because this pairwise function is much simpler than surface area calculations, it is used in the current study for computational efficiency. A normalization factor is included to improve the accuracy. Otherwise, the calculated energy change of a residue could exceed the free energy of transfer from a completely buried state to a fully hydrated state. The interaction function is derived from a macroscopic interaction, so the group-based solvent effects are calculated for each backbone peptide unit and each side chain. For the hydrophobic effect, a modified formula is chosen in this study,

$$\Delta G_{ij} = -\left(\frac{H_i V_j}{N_i} + \frac{H_j V_i}{N_j}\right)\exp\left(-\frac{r_{ij}}{D}\right), \quad (1)$$

where i and j refer to the groups i and j; r_{ij} is the distance between i and j, and D is a constant assigned to be 1 Å by Fauchere et al. (1988). $D = 2$ Å (used by Vila et al. 1991) is also tested. H_i and H_j are the hydrophobicity values of groups i and j, in terms of the free energy of transfer from hydrophobic to hydrophilic environments. V_i and V_j are the volumes of groups i and j. N_i and N_j are the normalization factors that give the maximum energy change of each group equal to the free energy of transfer, H_i and H_j, respectively. The value of N depends on the cutoff distance, the packing density, and the nearest neighbor effect. In this study, the values of N for $D = 1$ Å are set to 2.6, 1.8, and 2.0 ($\times 4\pi/3$ Å^3) for alanine side chains, lysine side chains, and the —CH_2CONH— (or $\rangle$CHCONH—) groups, respectively.

Octanol is probably one of the best choices for a partially polar solvent to mimic the properties of the protein interior (Sharp et al., 1991). The values of H in the current study are the free energies of transfer from octanol to water, based on measurements of Fauchere and Pliska (1983) with the adjustment of Sharp et al. (1991). Because the data of Fauchere and Pliska (1983) were obtained from the partition of N-acetyl amino acid amides, the

value for a side chain is taken from the difference between this residue and glycine. Each backbone —CH_2CONH— unit is treated as an interaction center located at the carbonyl carbon position. Its H value is set to 0.42 kcal/mol, which is the average of the difference between asparagine and glycine and of the difference between glutamine and alanine. This value is also close to one-half (about 0.5 kcal/mol) that of N-acetyl glycine amide, which contains two —CH_2CONH— units. Because 0.42 is a rough estimate, a value of zero for —CH_2CONH— is also tested. In general, the free energy of transfer from octanol to water results in a smaller energy change than the NMR-derived parameters used by Vila et al. (1992) in the calculations of helix unfolding. This difference seems to be a major reason that the results of the current study are not entirely the same as their results.

The electrostatic interaction of charged residues with solvent, or electrostatic hydration, differs in nature from the solvent entropy effect and is treated separately. Charged residues have much stronger electrostatic interaction with water than neutral residues. This interaction pulls the charged residues to the surface of the protein molecule. The electrostatic hydration can be calculated by solving the Poisson-Boltzmann equation numerically (Gilson et al., 1987), but it is too time-consuming for folding simulations. Recently, Gilson and Honig (1991) proposed a simple approximation to incorporate electrostatic hydration in molecular mechanics calculations:

$$\Delta\Delta G_i = \left(\frac{332}{8\pi}\right)\left(\frac{1}{\epsilon_m} - \frac{1}{\epsilon_s}\right) q_i^2 \left(\sum_{j\neq i}^{N} \frac{V_j}{r_{ij}^4} - \sum_{l\neq i}^{M} \frac{V_l}{r_{il}^4}\right), \quad (2)$$

where $\Delta\Delta G_i$ is the effect of all other atoms on the electrostatic hydration energy of atom i; ϵ_m and ϵ_s are the dielectric constants of the protein and solvent with assigned values of 2 and 80, respectively; q_i is the charge on i; V_l and V_j are the volumes of l and j; and r_{ij} (or r_{il}) is the distance between i and j (or between i and l). The first sum runs over all atoms except i itself. The second sum corresponds to the fully hydrated state and is zero because a fully hydrated atom should be infinitely separated from other protein atoms. This approximation agrees well with the numerical calculations and certain experimental data (Gilson and Honig, 1991). The electrostatic interaction of the partial charges of the backbone atoms with the solvent is included in the hydrophobic effect. The electrostatic hydration, therefore, is calculated for charged side chains only, as in the treatment of Vila et al. (1992). Because the derivation of the formula does not require each particle to be an atom, the lysine side chain is treated as a whole group with the charge of +1.

In the current study, all simulations were carried out at a constant temperature of 274 K because these synthetic peptides were studied experimentally at this temperature (Marqusee et al., 1989). All of the illustrations of the peptide conformations were produced using the MidasPlus software system from the Computer Graphics Laboratory, University of California, San Francisco (Ferrin et al., 1988).

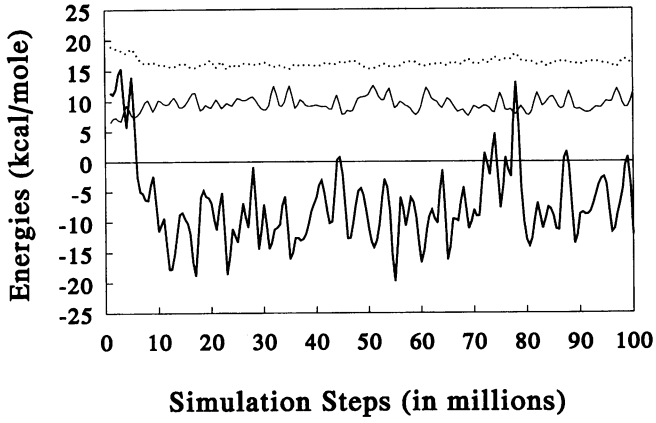

FIGURE 1 The potential energies during the folding simulation of the peptide 3K(I). The thick solid curve (*bottom* curve) represents the total potential energy. The thin solid curve (*middle* curve) represents the electrostatic hydration. The dotted curve (*top* curve) represents the hydrophobic effect.

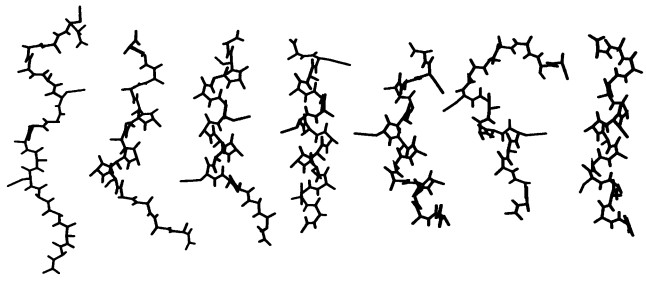

FIGURE 2 The conformations at steps 3 M, 6 M, 8 M, 12 M, 19 M, 78 M, and 80 M (*left to right*) during the folding simulation of 3K(I). N terminus is at the bottom of the figure. Alanine side chains are represented by a C_β atom, and lysine side chains are represented by both a C_β atom and an interaction center for the rest of the atoms. The conformation at step 3 M is an extended coil after relaxation from the fully extended initial conformation. At step 6 M a small helical segment formed in the middle of the molecule. At step 8 M, the helical segment propagated toward the C terminus, and at step 12 M the whole helix formed. The conformations at step 19 M, 78 M, and 80 M show different partially unfolded helices in equilibrium.

RESULTS AND DISCUSSION

Simulation of 3K(I) with an extended initial conformation

The first simulation started with the fully extended conformation (all of the φ, ψ angles equal to 180°) of 3K(I) and ran for 100 million (100 M) steps. The dielectric constant is 40 for interactions with charged lysine side chains. The potential energies during the simulation are shown in Fig. 1, and the conformations illustrating the folding process are shown in Fig. 2. As the simulation started, the molecule relaxed quickly from the fully extended geometry. Unstable hydrogen bonds formed and broke. At step 3 M, the conformation was an extended coil, as shown in Fig. 2. At step 5 M, a hydrogen bond formed between the carbonyl of residue 4 and the amino hydrogen of residue 7 in a β-bend structure. At step 6 M, a helical segment containing six hydrogen bonds formed from residue 4 to residue 13. At step 8 M, the helical conformation propagated to the C-terminus. At step 12 M, a complete helix formed. The first hydrogen bond from the N terminus and the last three hydrogen bonds were in the (i, i+3) pattern, or the 3_{10} helix pattern, whereas the major part of the helix was in the (i, i+4) pattern, or the α-helix pattern (Fig. 2). After step 12 M, helix-dominated conformations were observed. Corresponding to the high energy peak at step 78 M, the helix was mostly unfolded, as shown in Fig. 2. At step 80 M, the helix reformed and the energy dropped. An equilibrium of many conformations observed in the simulation is consistent with the fact that the peptide is in constant motion at the experimental temperature. The total potential energy decreased from about 10 kcal/mol for the extended coil to about −10 kcal/mol for the helix-dominated conformations. This energy change agrees well with the ex-

perimentally measured enthalpy change upon helix folding (Scholtz et al. 1991).

The hydrophobic effect stabilizes the helix relative to the extended conformation. This energy is about 2 kcal/mol lower for helical conformations because helices are more compact than the extended conformation. In contrast, the electrostatic hydration energy destabilizes the helical conformation. This energy increases by about 3 kcal/mol during helix folding because charged side chains are, on average, closer to other residues in helical conformations than in extended conformations. Unlike the residues buried inside a globular protein, the residues in a single helix are still largely exposed. Therefore, the energy changes of both the hydrophobic effect and the electrostatic hydration are small. When a whole protein molecule folds, these energy changes will be larger. The solvation effect, as well as the packing factor, are very important in protein folding (Dill, 1990; Chan and Dill, 1991; Rose and Wolfenden, 1993). The study of an isolated 3K(I) helix does not fully address the role of these factors.

Fig. 3 is a Ramachandran plot for all 16 residues. This plot shows the dihedral angles between step 10 M and 100 M at 1 M step intervals. The data of the first 10 M steps are not shown because during the starting period the initial structure has a large effect and the equilibrium has not been reached. In Fig. 3 the highest density of points is in the helical region of $-90° < \varphi < -45°$ and $-60° < \psi < -15°$. This region is not exactly centered at the ideal α-helix angles ($-57°$ and $-47°$, respectively). Instead, it is close to the helical region in the molecular dynamics study of helix unfolding in water by Daggett and Levitt (1992). The region corresponding to the extended conformations (or β-strands) and the region between the extended and helical regions are also populated, indicating the coil-helix transition. Because the 3_{10} helix is an intermediate during helix folding and unfolding (Tirado-Rives and Jorgensen, 1991; Sung, 1994), the 3_{10} helix region ($-60°, -30°$) is also substantially populated. In addition to the β-bend structure at step 5 M, the (i,i+3) hydrogen bonds and the 3_{10} helix segments often occur near the ends of a helix, such as those seen at the C-terminal region of the helix at steps 12 M and 80 M in Fig. 2. The observations on the β-bend structures, the (i,i+3) hydrogen bonds, and the 3_{10} helices agree with previous findings in both theoretical and experimental studies (Lewis et al., 1971; Zimmerman and Scheraga, 1977; Sundaralingam and Sekharudu, 1989; Tobias and Brooks, 1990; Tirado-Rives and Jorgensen 1991; Miik et al., 1992; Sung, 1994).

Following the assignment of Daggett and Levitt (1992), a residue is considered helical when its φ and ψ angles are within 30° of the ideal α-helix angles of $-57°$ and $-47°$. Using this criterion, from step 10 M to the end of the simulation (100 M steps), an average of 66% of the residue conformations are helical. For each residue the helicity is quite different, shown as solid bars in Fig. 4. The middle residues have higher helicities (close to 90%). The residues near the C terminus have much lower helicities. These findings are consistent with experimental studies (Casteel et al., 1993) and other theoretical results (Chakrabartty et al., 1994). The N-terminal capping by the acetyl group is one reason for the higher helicity near the N terminus than the C terminus. The positively charged lysine side chains may also play a role. The CO groups point toward the C end of the helix. A positively charged side chain in the C-end side of a residue stabilizes the helical conformation of the residue, and that in the N-end destabilizes the helical conformation. The effect of the negatively charged side chains is the opposite. For residues near the N terminus, the lysines are located in their C-end side and stabilize the CO and NH alignment in the helix. For

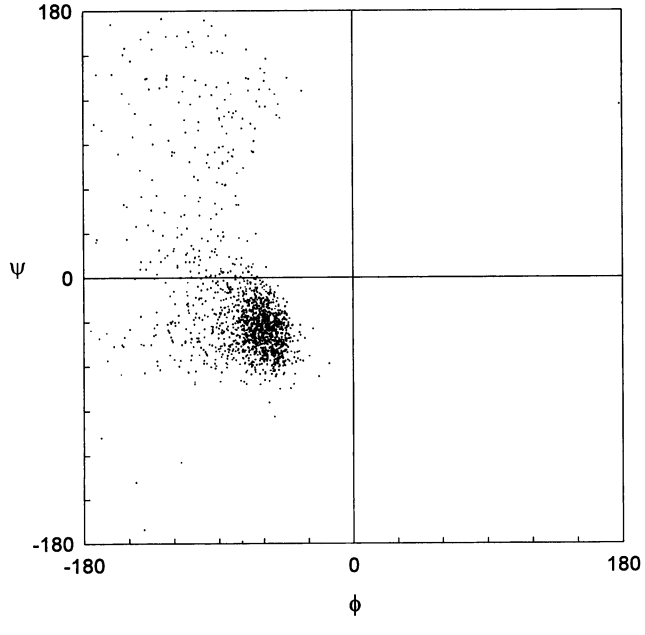

FIGURE 3 The φ, ψ angles during the simulation of the peptide 3K(I) with the fully extended initial conformation. The points represent the dihedral angle pairs of all 16 residues during the simulation between steps 10 M and 100 M at 1 M step intervals.

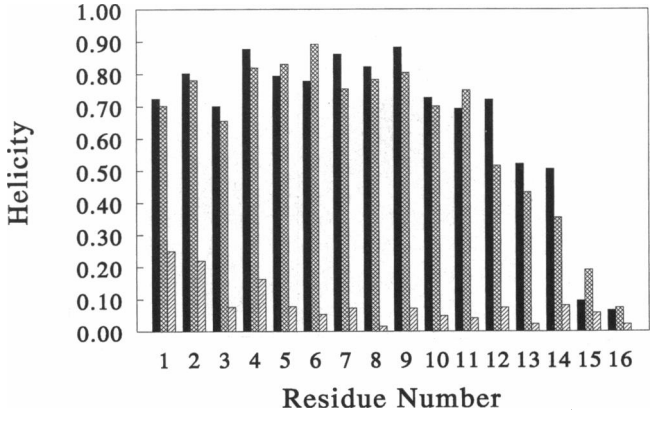

FIGURE 4 The helicities of the residues during the simulation. The solid bars are for 3K(I), the hatched bars are for 3K(II), and the striped bars are for 6K(I).

residues near the C terminus, the lysines are located in their N-end side and destabilize the CO and NH alignment in the helix.

Tests on other initial conformations and different potentials

To see the effect of the initial structure, a conformation with random φ, ψ angles was tested under the same conditions as those for the extended initial conformation. The peptide formed helix-dominated conformations in 10 M steps. The average helicity is 63% from steps 10 to 100 M. A left-handed helix was tested as the initial conformation in a previous study of the polyalanine helix model (Sung, 1994). In the current study, the amino acid sequence is different and different solvation energies are used for alanine and lysine residues. Therefore, this initial conformation is tested again. During the simulation, the left-handed helix unfolded gradually and the right-handed helical segments started to form before the left-handed segment unfolded entirely. The left-handed helix unfolded completely at step 55 M, and a complete right-handed helix formed at step 81 M. The multiple-minima problem did not prevent the simulation from reaching the lowest energy conformation. After step 81 M the right-handed helical conformations dominated in the equilibrium. From step 80 M to the end of the simulation (step 100 M), the average helicity is 62%. Fig. 5 shows the dihedral angles during the simulation. The points around $(+50°,+40°)$ are from the left-handed helices, and those around $(-60°,-40°)$ are from the right-handed helices. As in Fig. 3, the region of extended conformations and the coil-helix transition region are also populated. The distribution of the dihedral angles in the $\varphi < 0°$ region is similar to that in Fig. 3, except that the density in the right-handed helix region in Fig. 5 is smaller because it formed late in the simulation. This similarity indicates a similar coil-helix equilibrium despite the differences in the initial conformations.

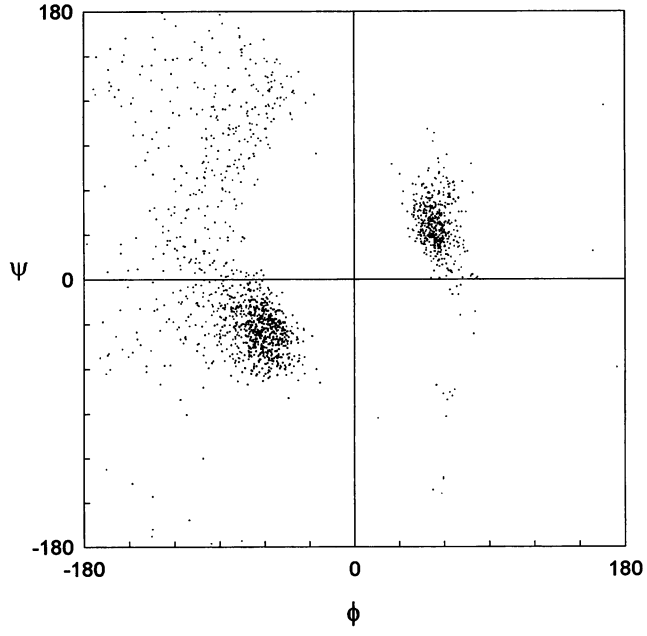

FIGURE 5 The φ, ψ angles during the simulation of 3K(I) that started with the left-handed helix. The points represent the dihedral angle pairs of all 16 residues during the simulation between steps 1 M and 100 M at 1 M step intervals.

The formula used to calculate the hydrophobic effect is highly approximate, but the hydrophobic contribution to the helix stabilization is only about 2 kcal/mol, less than 20% of the total potential energy change. The qualitative results do not seem to depend critically on the approximate formula used to calculate the hydrophobic effect. Nevertheless, other functions of the hydrophobic effect were tested. A different distance dependence of the hydrophobic effect, $\exp(-R_{ij}/2)$, was tested. With this distance dependence, the energy change of the hydrophobic effect was slightly less, and a helix-dominated equilibrium was also obtained. A solvent-accessible surface area-based calculation was tested using the approximation of Wodak and Janin (1980) for computational efficiency. Again, predominant helical conformations were observed. The zero hydrophobicity for —CH_2CONH— was also tested because direct experimental data on the free energy of transfer of the —CH_2CONH— group are not available. The results were similar, and the average helicity of all residues during the period from 10 M to 100 M steps was 60%.

The simulation of 3K(I) was also carried out using a dielectric constant of 80 for charged lysine side chains. The average helicity of all residues was 65%. The electrostatic interactions among the charged side chains did not change much during the helix folding because they are separated by four alanine residues in the sequence and are evenly distributed around the helix axis in space. The lysine side chain charges do not have a large destabilizing effect on the helical conformation of 3K(I). To investigate further the charge effect of the lysine residues, the side-chain charge was artificially set to zero. Consequently, both the intramolecular electrostatic interaction and the electrostatic hydration of the lysine residues became zero. The average helicity of all residues was 69%. In summary, with different initial conformations, the multiple minima problem did not prevent the simulation from reaching the lowest energy conformation. With different electrostatic interactions of the charged lysine side chains, the calculated average helicity of the 16 residues of peptide 3K(I) is in the range between 60 and 70%. The effect of side-chain charges on the calculated helicity of 3K(I) is small.

Simulations of 3K(II)

For another 16-residue, alanine-based peptide with three lysines, 3K(II), a simulation with a fully extended initial conformation was carried out under the same conditions as for 3K(I) shown in Fig. 1. Helix-dominated conformations formed after 7 M steps. The average helicity of the 16 residues of 3K(II) is 63% during the simulation from step 10 M to step 100 M. The helicity of each residue is shown in Fig.

4 as hatched bars. The average helicity of 3K(II) is slightly lower than that of 3K(I), in agreement with experimental observations (Marqusee et al., 1989). The lower helicity might be because the positively charged lysines are closer to the N terminus in 3K(II) than in 3K(I). But the difference is small, and it may well be caused by fluctuation. The Ramachandran plot of 3K(II) is very similar to that of 3K(I) in Fig. 3 and, therefore, is not shown. The left-handed helix was also tested for 3K(II) as an initial conformation. The left-handed helix unfolded, and right-handed helices formed at about 50 M steps. With $\epsilon = 80$ for lysine side chains, the average helicity of the 16 residues is 62%, close to that with $\epsilon = 40$. Again, the effect of the lysine side-chain charges is small for the 16-residue, alanine-based peptide with three lysine residues.

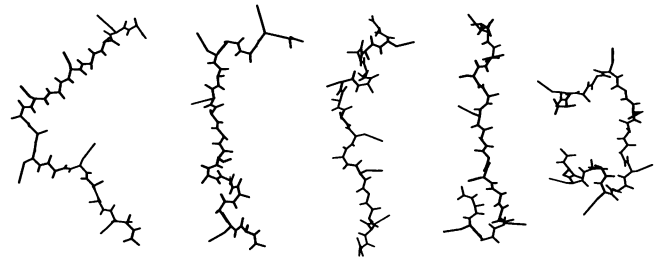

FIGURE 7 The conformations at steps 13 M, 14 M, 72 M, 100 M, and 151 M (*left to right*) during the simulation of 6K(I). The N terminus is at the bottom of the figure. Alanine side chains are represented by a C_β atom, and lysine side chains are represented by both a C_β atom and an interaction center for the rest of the atoms. The conformation at step 13 M is an extended coil after relaxation from the fully extended initial conformation. The conformations at step 14 M, 72 M, 100 M, and 151 M show selected low energy conformations.

Simulations of 6K(I)

For the 16-residue, alanine-based peptide with six lysine residues, 6K(I), a simulation was carried out with the fully extended initial conformation and $\epsilon = 40$. As shown in Fig. 6, the energies of 6K(I) are generally higher than that of 3K(I), mainly because of the repulsion among the charged lysine side chains. The electrostatic hydration energy is also higher. This energy term represents the energy increase of lysines in the peptide conformation from that in a fully hydrated state. Its value is positive. More lysine residues contribute to higher energy change for 6K(I) than 3K(I). Unlike 3K(I), the average energies remain nearly constant during the simulation, indicating no stable folding. During the simulation, isolated hydrogen bonds of the β-bend structure started to appear near the termini at about 3 M steps but were not stable. Corresponding to several energy minima, small helical segments occurred, such as those at steps 14 M, 72 M, and 151 M in Fig. 7. However, not all energy minima corresponded to helical segments. The conformation at step 100 M contained a small β-sheet segment near the N terminus. These secondary structure elements were unstable and disappeared in the following steps. However, they indicate that helical conformations are accessible and energy barriers are not the major reason that stable helices did not form. Still, the simulation of 6K(I) ran for 200 M steps, longer than that of 3K(I), to see whether higher helicity could be reached.

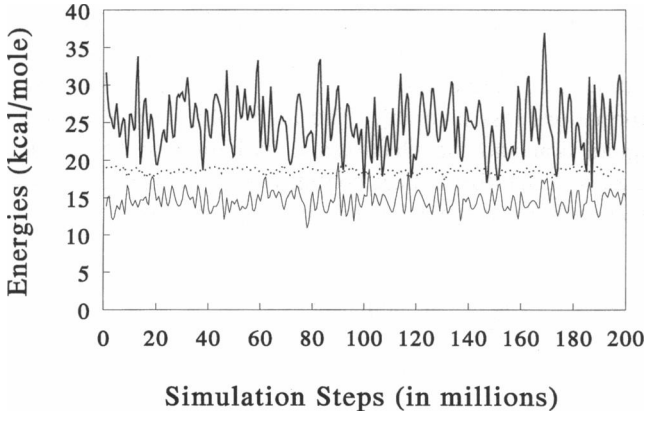

FIGURE 6 The potential energies during the folding simulation of the peptide 6K(I). The thick solid curve (*top* curve) represents the total potential energy. The thin solid curve (*bottom* curve) represents the electrostatic hydration. The dotted curve (*middle* curve) represents the hydrophobic effect.

The helicity of each residue of 6K(I) from 10 M to 200 M steps are shown in Fig. 4 as striped bars. The value for each residue is generally below 25%. Interestingly, the first two residues have the highest values, because of both the N-capping effect and the electrostatic effect of the large number of the positively charged lysines in the C-terminal side of the first two residues. For both the period from 10 M to 100 M steps and the period from 10 M to 200 M steps, the average helicity of the 16 residues is 8%, which is much lower than those of 3K(I) and 3K(II). Experimental measurements have shown that this peptide has a lower helical content, about 15% (Marqusee et al. 1989). The large difference in the calculated helicities between 6K(I) and 3K(I) qualitatively agrees with experimental results. The lower helicity of 6K(I) is also shown in a Ramachandran plot (Fig. 8). The density in the helical region in Fig. 8 is smaller than that shown in Fig. 3. The region corresponding to the extended structure is populated more.

Because a complete helix did not form during the simulation of the peptide 6K(I), it was not known if a coil-helix equilibrium was reached. Therefore, another simulation started with an α-helical conformation of 6K(I) and ran for 200 M steps. During the simulation, the helicity decreased gradually from the initial value of 100%. The helix unfolded completely within 20 M steps. But in the following steps, helical segments formed and unfolded several times. The average helicity finally decreased to about 8% after step 150 M, which is consistent with the helicity calculated with the extended initial conformation. Therefore, the calculated helicity of about 8% is close to the equilibrium value. The intramolecular electrostatic energy difference between the extended conformation and the helical conformation is about 15 kcal/mol for 6K(I) and 20 kcal/mol for 3K(I). The electrostatic hydration energy difference is about -5 kcal/mol for 6K(I) and about -3 kcal/mol for 3K(I). The potential

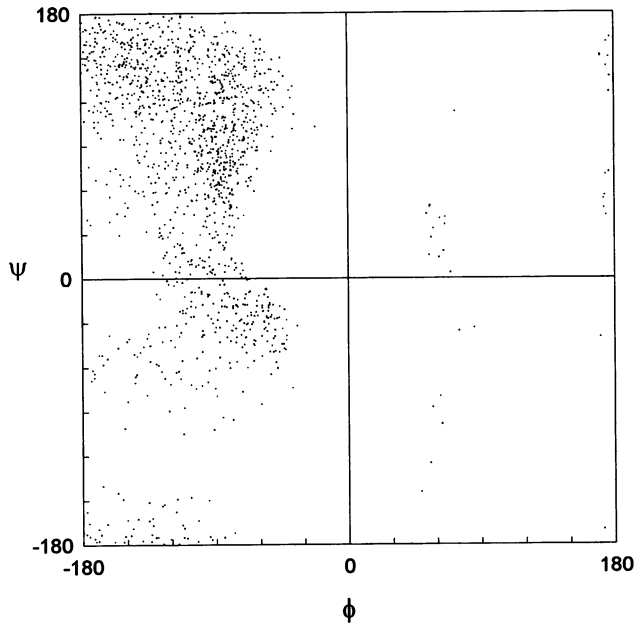

FIGURE 8 The φ, ψ angles during the simulation of 6K(I) with the fully extended initial conformation. The points represent the dihedral angle pairs of all 16 residues during the simulation between steps 10 M and 100 M at 1 M step intervals.

energy difference of the coil-helix transition of 6K(I) seems too small to overcome the entropy effect to reach a high helicity. The charges on the six lysine side chains destabilize helical conformations of 6K(I).

The dielectric constant 80 for charged lysine side chains was tested for 6K(I) with the fully extended initial conformation. The average helicity from step 10 M to 200 M is 14%, which is higher than that with dielectric constant 40. This difference shows that the electrostatic interaction of the charged side chains destabilizes the helical conformation of 6K(I). In contrast, it has little effect on the helix stability of 3K(I). These results are consistent with the experimental observations that the salt concentration has larger effect on the helical content of 6K(I) than on 3K(I) (Marqusee et al., 1989), because the dielectric constant is dependent on the salt concentration.

Further comparison with experimental measurements and discussion of force fields

Quantitatively, the calculated helicities are not directly comparable with the experimentally measured helical contents because they are defined differently. The calculated helicity depends on the dihedral angle ranges that define the helical conformation, whereas the experimental value is measured by the minimum at 222 nm in the CD spectra. Daggett and Levitt (1992) also used a slightly different helical range in analyzing their helix-unfolding simulation results: the range of the dihedral angle φ is $-100°$ to $-30°$, and dihedral angle ψ is $-80°$ to $-5°$. Using this range and the dielectric constant of 40 for charged side chains, the calculated helicity of 3K(I) is 77%, that of 3K(II) is 74%, and that of 6K(I) is 11%. These values are higher than those obtained with the previous helical range and are quantitatively closer to the experimentally measured helical contents. The calculated average helicities are listed in Table 1. However, given the highly approximate nature of the force field, exact quantitative agreement with experiments is not expected. Qualitatively, with either definition the large difference of the calculated helicities between 3K(I) (or 3K(II)) and 6K(I) agrees well with experimental results.

Alanine-lysine peptides were previously simulated using $\epsilon = R$ for backbone atoms (our unpublished results). From extended conformations, 3K(I) and 3K(II) showed helix-dominated final structures and 6K(I) had a much shorter helical segment. Once these conformations formed, interconversions among different conformations were rare and the structures were more static. The major problem in using $\epsilon = R$ is that the calculated energy change is more than twice as large as the experimentally obtained enthalpy change. The exaggerated energy change causes higher energy barriers that prevent frequent interconversions among different conformations and cause the multiple minima problem. Using the solvent-referenced potential, the exaggerated energy change is reduced in part by the larger effective dielectric constant. With $\epsilon = 2R$, more motion is observed. After reaching equilibrium, every residue has some probability in the helical conformation and some probability in the coil conformation. More conformations are accessible. Even the most stable conformations, like the 3K(I) helices, sometimes unfold. The conformations with local minimum energies are unlikely to hold for a long period. Therefore, the multiple minima are no longer a serious problem. The simulation with the solvent-referenced potential represents more realistically the motion of the molecule, and its enthalpy change upon folding is in the experimental range.

SUMMARY

In summary, the simulations show predominantly helical conformations for the alanine-based peptides with three lysines and much lower helicity for the peptide with six lysines. These results are in good agreement with experimental observations. The lysine residues make the alanine-based peptides soluble, but too many lysine residues make helical conformations less favorable. The intramolecular electrostatic interaction and the electrostatic hydration of lysine side chains destabilize helical conformations of 6K(I), whereas these effects on the helix stability of the peptides with three

TABLE 1 The calculated helicities

Helical criteria*	Dielectric constant for lysine side chains	3K(I)	3K(II)	6K(I)
$-87° < \varphi < -27°$ & $-77° < \psi < -17°$	$\epsilon = 40$	66%	63%	8%
	$\epsilon = 80$	65%	62%	14%
$-100° < \varphi < -30°$ & $-80° < \psi < -5°$	$\epsilon = 40$	77%	74%	11%

*Dihedral angle ranges used by Daggett and Levitt (1992).

lysine residues are small. The simulations provide insight into the helix-folding mechanism, which is consistent with a previous helix-folding study using polyalanine model (Sung, 1994) and other studies (Lewis et al., 1971; Zimmerman and Scheraga, 1977; Sundaralingam and Sekharudu, 1989; Tobias and Brooks, 1990; Tirado-Rives and Jorgensen 1991; Miik et al., 1992). The β-bend structure is observed as intermediate in helix initiation. The (i, i+3) hydrogen bonds often occur near the ends of helical segments during folding or unfolding, whereas the middle portion is often α-helical. The helix propagation toward the C terminus seems faster than toward the N terminus, but the residues near the N terminus show higher helicity in the thermodynamic equilibrium.

REFERENCES

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

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., T. Kortemme, and R. L. Baldwin. 1994. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* 3:843–852.

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

Chan, H. S., and K. A. Dill. 1991. Polymer principles in protein structure and stability. *Annu. Rev. Biophys. Biophys. Chem.* 20:447–490.

Chandler, D., J. D. Weeks, and H. C. Andersen. 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. 1991a. Molecular dynamics simulation of small peptides: dependence on dielectric model and pH. *Biopolymers.* 31:285–304.

Daggett, V., P. A. Kollman, and I. D. Kuntz. 1991b. 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.

Daggett, V., and M. Levitt 1993. Realistic simulations of native-protein dynamics in solution and beyond. *Annu. Rev. Biophys. Biomol. Struct.* 22:353–380.

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

Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature.* 319:199–203.

Eisenberg, D., M. Wesson, and M. Yamashita. 1989. Interpretation of protein folding and binding with atomic solvation parameters. *Chem. Scripta.* 29A:217–221.

Fauchere, J. L., and V. Pliska. 1983. Hydrophobic parameters p of amino-acid side chains from the partitioning of N-acetylamino-acid amides. *Eur. J. Med. Chem. Chim. Ther.* 18:369–375.

Fauchere, J. L., P. Quarendon, and L. Kaetterer. 1988. Estimating and representing hydrophobic potential. *J. Mol. Graph.* 6:203–206.

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

Gerstein, M., E. L. L. Sonnhammer, and C. Chothia. 1994. Volume changes in protein evolution. *J. Mol. Biol.* 236:1067–1078.

Gilson, M. K., and B. Honig. 1991. The inclusion of electrostatic hydration energies in molecular mechanics calculations. *J. Comp. Aid. Mol. Des.* 5:5–20.

Gilson, M. K., K. A. Sharp, B. H. Honig. 1987. Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comput. Chem.* 9:327–335.

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

Isrealachvili, J., and R. Pashley. 1982. The hydrophobic interaction is long range, decaying exponentially with distance. *Nature.* 300:341–342.

Kawai, H., Y. Okamoto, M. Fukugita, T. Nakazawa, and T. Kikuchi. 1991. Prediction of α-helix folding of isolated C-peptide of ribonuclease A by Monte Carlo simulated annealing. *Chem. Lett.* 2:213–216.

Levitt, M. 1976. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* 104:59–107.

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.

Loewenthal, R., J. Sancho, T. Reinikainen, and A. R. Fersht. 1993. Long-range surface charge-charge interactions in proteins comparison of experimental results with calculations from a theoretical method. *J. Mol. Biol.* 232:574–583.

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.

Miick, 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.

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

Okamoto, Y. 1994. Helix-forming tendencies of nonpolar amino acids predicted by Monte Carlo simulated annealing. *Proteins Struct. Funct. Genet.* 19:14–23.

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.

Pashley, R. M., P. M. McGuiggan, B. W. Ninham, D. F. Evans. 1985. Attractive forces between uncharged hydrophobic surfaces: direct measurements in aqueous solution. *Science.* 229:1088–1089.

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

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.

Rose, G. D., and Wolfenden R. 1993. Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annu. Rev. Biophys. Biomol. Struct.* 22: 381–415.

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.

Sharp, K., A. Nicholls, R. Friedman, and B. Honig. 1991. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry.* 30:9686–9697.

Sikorski, A., and J. Skolnik. 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.)

Sung, S.-S. 1994. Helix folding simulations with various initial conformations. *Biophys. J.* 66:1796–1803.

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.

van Gunsteren, W. F., F. J. Luque, D. Timms, and A. E. Torda. 1994. Molecular mechanics in biology: form structure to function, taking account of solvation. *Annu. Rev. Biophys. Biomol. Struct.* 23:847–863.

Vila, J., R. I. Williams, J. A. Grant, J. Wojcik, and H. A. Scheraga. 1992. The intrinsic helix-forming tendency of L-alanine. *Proc. Natl. Acad. Sci. USA.* 89:7821–7825.

Vila, J., R. L. Williams, M. Vasquez, and H. A. Scheraga. 1991. Empirical solvation models can be used to differentiate native from near-native conformations of bovine pancreatic trypsin inhibitor. *Proteins Struct. Funct. Genet.* 10:199–218.

Warshel, A., and J. Aqvist 1991. Electrostatic energy and macromolecular function. *Annu. Rev. Biophys. Biophys. Chem.* 20:267–298.

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. Am. 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.

Wodak, S. J., and J. Janin. 1980. Analytical approximation to the accessible surface area of proteins. *Proc. Natl. Acad. Sci. USA.* 77:1736–1740.

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