Self-Guided Molecular Dynamics Simulation for Efficient Conformational Search

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · SEPTEMBER 1998

Impact Factor: 3.3 · DOI: 10.1021/jp9817372

2 AUTHORS:

Xiongwu Wu
National Heart, Lung, and Blood Institute
82 PUBLICATIONS **4,432** CITATIONS
SEE PROFILE

Shaomeng Wang
University of Michigan
291 PUBLICATIONS **14,370** CITATIONS
SEE PROFILE

Self-Guided Molecular Dynamics Simulation for Efficient Conformational Search

Xiongwu Wu and Shaomeng Wang*

Institute for Cognitive and Computational Sciences, Georgetown University Medical Center, The New Research Building, EP07 3970 Reservoir Road, Washington, D.C. 20007

Received: April 3, 1998

Conformational searching is a basic approach in theoretical and computational chemistry, and molecular dynamics [MD] simulation is a powerful computational tool to study structural, thermodynamic, and kinetic properties of molecular systems. In this paper, we report a new MD simulation method with the primary goal of improving dramatically the conformational searching efficiency in an MD simulation in order to study slow events in physics, chemistry, and biology. This new MD simulation method was based upon a new equation of motion in which a guiding force was introduced, in addition to the instantaneous force, to accelerate the systematic motion. This guiding force may be calculated as a time average of the instantaneous force from the trajectory obtained from the very same MD simulation. Bonded substructures were defined for molecular systems in order to enhance effectively both the inter- and intramolecular systematic motions. The improvement of the overall conformational searching efficiency and the possible alterations of ensemble properties and the conformational distribution of the system were investigated through comparative simulations of two peptide systems, an alanine dipeptide and a 16-residue synthetic peptide, both in vacuum and in explicit water. We demonstrated that the self-guided MD simulation method improved conformational searching efficiency but did not significantly alter the ensemble average properties and the conformational distributions of the systems. The successful helix-folding simulations of the 16-residue synthetic peptide both in vacuum and in explicit water clearly showed that the self-guided MD simulations achieved a dramatic enhancement in the conformational searching efficiency and showed an extraordinary ability to overcome energy barriers, as compared to the conventional MD simulations. Our results suggest that this new MD simulation method may be used as a powerful tool to study many important yet slow events or processes in physics, chemistry, and biology.

Introduction

Molecular dynamics [MD] simulation is a powerful computational tool to study structural, thermodynamic, and kinetic properties of molecular systems. MD simulation has been employed to study systems from simple fluids[1] to large proteins and DNA molecules.[2−4] In the past decade, the use of MD simulations to study chemical and biological problems has increased exponentially, primarily owing to the accessibility of relatively inexpensive and powerful computers to chemist and biologist.

Motions in molecular systems take place on a wide range of time scales. Vibrations of covalent bonds take place in 10^{-14} s, whereas large-scale conformational changes such as that in the protein-folding process may take seconds or even minutes. This large range of time scales presents MD simulations a great challenge. In an MD simulation, the classical equations of motion for all particles of a system are integrated over a certain period of time. Because of the limitation in the size of time steps used to solve the equations of motion, the time scale that an MD simulation can access is very short with current available computing power, typically in the range of picoseconds (ps) to nanoseconds (ns). This short time scale severely limits the applicability of MD simulations. To extend the applicability of MD simulations, a number of strategies have been used.

A simple strategy is to improve the computing power, and indeed, as computing power increases, more difficult problems can be studied.[5] Another strategy is to improve MD simulation efficiency. For example, fast motions limit the size of time steps used to solve the equation of motion and make simulations very expensive. Therefore, when fast motions are not of major interest and are separable from other motions, constraints may be used to replace the degrees of freedom of fast motions, so that much larger time steps can be used.[6−8] Other strategies to improve MD simulation efficiency include the use of multiple time steps to save computational load,[9−13] the improvement of numerical integration methods[14] and the use of solvation models to replace explicit solvent molecules.[15] However, the improvement in MD simulation efficiency is far from sufficient to study slow events.

Another very productive strategy is to speed up a slow event so that the event can take place fast enough to be observed on a time scale accessible by an MD simulation. The simulated annealing method is a very successful example in this category.[16−18] In this method, high temperature is used to increase the ability of the system to overcome energy barriers. The system is then slowly cooled to low temperatures and simulated at physiologically relevant conditions. A major advantage of the simulated annealing MD method is that the system can effectively overcome energy barriers and reach different conformational regions at high temperature. A drawback of this method is that it needs multiple simulations to find the global minimum state of the system and to search properly the conformational space of the system.

In the context of conformational search, the motions of a system may be divided into random and systematic motions.

* To whom all correspondence should be sent: (202)-687-2028 (phone); (202)-687-0617 (fax); Email, wangs@giccs.georgetown.edu.

S1089-5647(98)01737-4 CCC: $15.00
Published on Web 08/22/1998

While the random motions are important to the sampling of the detailed conformational distribution of the system, the systematic motions of the system are more relevant to the conformational search over a conformational space of significant size. For a many-body system, the systematic motions are normally much slower than the random motions because of energy barriers and random walk. Energy barriers are created by nonuniform energy surface, and random walk is due to the frequent collisions between atoms. One example of random walk is the diffusion of solute molecules in solution. The constant collisions of solute molecules with surrounding solvent molecules make the solute molecules go through a very complicated and detoured path, which looks like a random walk. Consequently, the diffusion speed of a solute is therefore far slower than its actual thermal motion.

The slow systematic motions are the bottleneck that limits the efficiency of a conformational search during an MD simulation. Therefore, if we can somehow specifically enhance the systematic motions, the efficiency of conformational searching may be improved.

In this work, we wish to present a new MD simulation method, termed the self-guided MD simulation method, with the primary goal of improving the conformational searching efficiency. In this method, an extra guiding force was introduced, in addition to the instantaneous force derived from the force field, into the equation of motion to speed up the systematic motion of the system. In this paper, we first present the derivation and the algorithm of this new method. We then demonstrated the efficiency and examined the behavior of this method through comparative simulation studies of two peptides both in vacuum and in explicit water using the conventional MD and the self-guided MD methods.

The Self-Guided MD Simulation Method

The conformational searching efficiency in an MD simulation depends on how fast the systematic conformational change takes place. Hence, through speeding up the systematic conformational change, or the systematic motion, the conformational searching efficiency in an MD simulation could be improved. In this section, we present a new MD simulation method with a primary goal of enhancing the systematic motion.

Equation of Motion with an Enhanced Systematic Motion. The equation of motion used in the conventional MD simulation under the Newtonian law is shown in eq 1.

$$m_i\ddot{\mathbf{r}}_i = \mathbf{f}_i \tag{1}$$

where m_i is the mass of atom i, $\ddot{\mathbf{r}}_i$ is the acceleration of atom i, and $\mathbf{f}_i$ is the force acting on atom i.

In a many-body system, most of motions change their direction frequently owing to collisions and restraints, such as bond stretching and angle bending. The systematic motion of atom i may be described by its average motion during a certain time period t_l, which we simply call the averaging time thereafter.

$$\bar{\dot{\mathbf{r}}}_i(t) = \frac{\mathbf{r}_i(t) - \mathbf{r}_i(t - t_l)}{t_l} = \frac{1}{t_l}\int_{t-t_l}^{t} \dot{\mathbf{r}}_i(\tau)\, d\tau \tag{2}$$

where $\bar{\dot{\mathbf{r}}}_i(t)$ is the average velocity for atom i from time $t - t_l$ to time t, $r_i(t)$ is the coordinate of atom i at time t, and $\dot{\mathbf{r}}_i(\tau)$ is the velocity of atom i at time τ.

According to eq 1, the equation of the systematic motion can be expressed by eq 3.

$$m_i\bar{\ddot{\mathbf{r}}}_i(t) = \frac{m_i}{t_l}\int_{t-t_l}^{t} \ddot{\mathbf{r}}_i(\tau)\, d\tau = \frac{1}{t_l}\int_{t-t_l}^{t} \mathbf{f}_i\, d\tau = \bar{\mathbf{f}}_i \tag{3}$$

where $\bar{\ddot{\mathbf{r}}}_i(t)$ is the acceleration of the systematic (average) motion of atom i at time t, $\ddot{\mathbf{r}}_i(\tau)$ is the acceleration of atom i at time τ, and $\bar{\mathbf{f}}_i$ is the average force. Equation 3 clearly shows that the acceleration of the systematic motion is governed by the average force.

Therefore, the introduction of the average force into the equation of motion as a guiding force can in principle alter the systematic motion of the system. With a guiding factor in eq 3 and combining it with eq 1, a new equation of motion can be obtained, as shown in eq 4.

$$m_i\ddot{\mathbf{R}}_i = m_i(\ddot{\mathbf{r}}_i + \lambda\bar{\ddot{\mathbf{r}}}_i) = \mathbf{f}_i + \lambda\bar{\mathbf{f}}_i \tag{4}$$

where $\ddot{\mathbf{R}}_i = \ddot{\mathbf{r}}_i + \lambda\bar{\ddot{\mathbf{r}}}_i$ is the acceleration of atom i with altered systematic motion as compared to $\ddot{\mathbf{r}}_i$ in eq 1 and λ is a guiding factor, which can be any value. Since the acceleration of the systematic (average) motion is governed by the average force, a positive value of λ should speed up and a negative value of λ should slow down the systematic motion. A value of 0 for λ returns eq 4 to the classical Newtonian equation of motion, eq 1.

The implementation of eq 4 should result in an MD simulation with an enhanced systematic motion when λ adopts a positive value. However, to perform an MD simulation precisely according to eq 4, one needs to perform an MD simulation according to eq 1 to obtain the average force $\bar{\mathbf{f}}_i$. This needs to be done at every time step, which essentially makes an MD simulation according to eq 4 impractical. To overcome this difficulty, eq 4 was modified as follows.

$$m_i\ddot{\mathbf{R}}_i = m_i\ddot{\mathbf{r}}_i + \lambda m_i\bar{\ddot{\mathbf{R}}}_i = \mathbf{f}_i + \lambda\mathbf{g}_i \tag{5}$$

where $\mathbf{g}_i$ is calculated from eq 6.

$$\mathbf{g}_i = m_i\bar{\ddot{\mathbf{R}}}_i = \frac{1}{t_l}\int_{t-t_l}^{t} (\mathbf{f}_i + \lambda\mathbf{g}_i)\, d\tau \tag{6}$$

$\mathbf{g}_i$ differs from $\bar{\mathbf{f}}_i$ in the way that $\bar{f}_i$ is responsible for the systematic motion of eq 1, while $\mathbf{g}_i$ is responsible for the systematic motion of eq 5. Since $\mathbf{g}_i$ is calculated from the trajectories of eq 5 and is also used to guide the very same motion according to eq 5, $\mathbf{g}_i$ is thus called the self-guiding force. Accordingly, eq 5 is called the equation of the self-guided motion, and an MD simulation performed according to eq 5 is called the self-guided MD simulation.

Equation of Motion with an Enhanced Systematic Motion for Molecular Systems. In many molecular systems, both inter- and intramolecular systematic motions are important for the conformational search. Therefore, an efficient simulation method needs to speed up both inter- and intramolecular systematic motions. Most of atomic motions related to bonded interactions are fast, while a systematic motion generally is very slow. During a conformational change, most of the internal coordinates, such as bond lengths, bond angles, and some of dihedral angles, just oscillate around their equilibrium values. Even though most of these high frequent motions do not contribute to a systematic conformational change, they create significant noise when using a time average velocity to describe the systematic motion. Hence, a guiding force estimated on the basis of the total force in a molecular system may be not

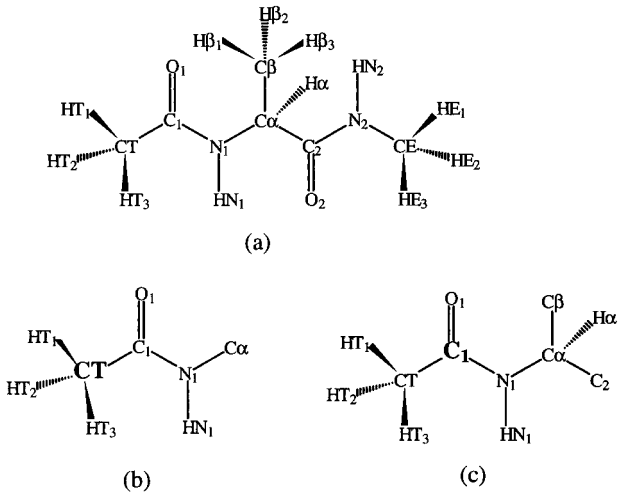

Figure 1. (a) All-atom model of the alanine dipeptide. (b) Bonded substructures of atom CT. (c) Bonded substructure of atom C_1.

very effective to accelerate the systematic motion. Because most of the internal coordinates just oscillate around their equilibrium values, we introduce the concept of bonded substructures to describe the local structural rigidity of a molecule and to derive a guiding force to enhance effectively the systematic motion.

For each atom, i, a bonded substructure, S_i, is defined as a part of the molecule that includes atom i itself, and all other atoms connecting to atom i through either a bond, or a bond angle, or a dihedral angle. In other words, all atoms in a molecule connecting to atom i with three or fewer covalent bonds belong to bonded substructure S_i. Atom i is called the central atom of substructure S_i. Each atom has a bonded substructure and two bonded substructures in a molecule can be partially or even totally overlapped as long as the central atoms are different. To illustrate the concept of bonded substructures, Figure 1 depicts an all-atom model for an alanine dipeptide and two bonded substructures for two atoms in the molecule.

With the concept of the bonded substructures, a conformational change in a molecule can be viewed as the external motion of bonded substructures and the internal motion within bonded substructures. The systematic conformational change of a molecule can be described by the external systematic motion of bonded substructures. The internal motion within bonded substructures is governed by the strong bonded interactions, including bond length, bond angle and dihedral angle potentials. It generally has high frequency and accounts for most of the kinetic energy of the system. The external motion is important for the systematic conformational change but is generally slow. Therefore, the enhancement of the external systematic motion of bonded substructures should effectively improve the conformational searching efficiency for molecular systems in an MD simulation.

To accelerate the external systematic motion of the bonded substructures, a specific guiding force, $\mathbf{g}_i^{(S)}$, was applied to atom i. Analogue to eq 5, the equation of motion for atom i with a guiding force $\mathbf{g}_i^{(S)}$ can be written as

$$m_i\ddot{\mathbf{r}}_i = \mathbf{f}_i + \lambda\mathbf{g}_i^{(S)} \tag{7}$$

Here we use $\ddot{\mathbf{r}}_i$ instead of $\ddot{\mathbf{R}}_i$ to denote the acceleration of atom i according to convention, although $\ddot{r}_i$ now is different from that in the classical Newtonian equation of motion, eq 1. Since the guiding force, $\mathbf{g}_i^{(S)}$, is used to accelerate the systematic motion of bonded substructures, instead of a single atom, its calculation is different from eq 6. Below, we derive the equation for the calculation of this guiding force based upon the definition of the bonded substructures.

For atom i, only the interaction force between atom i and atoms outside S_i is responsible for the external motion of bonded substructure S_i. This interaction force is denoted as $\mathbf{f}_i^{(S)}$, which can be calculated by eq 8.

$$\mathbf{f}_i^{(S)} = \sum_{j\notin S_i}\mathbf{f}_{ij} \tag{8}$$

where $\mathbf{f}_{ij}$ is the force between atoms i and j. The summation is over all the atoms that do not belong to bonded substructure S_i ($j \notin S_i$).

With an extra guiding force, $\lambda\mathbf{g}_i^{(S)}$, the contribution of the combined force on atom i to the acceleration of S_i is as follows.

$$M_i\mathbf{a}_{S_i} = \mathbf{f}_i^{(S)} + \lambda\mathbf{g}_i^{(S)} \tag{9}$$

where M_i is the total mass of S_i and $\mathbf{a}_{S_i}$ is the acceleration of S_i contributed from the combined force on atom i. Such acceleration will apply to all the atoms in S_i. If we neglect the rotation of the bonded substructure, for atom j in S_i, the acceleration of atom j resulted from the above motion of S_i in eq 9 is as follows.

$$\mathbf{a}_{S_i} = \frac{1}{M_i}(\mathbf{f}_i^{(S)} + \lambda\mathbf{g}_i^{(S)}) \tag{10}$$

Because atom j may belong to multiple bonded substructures, the total acceleration of atom j, $\ddot{\mathbf{r}}_j^{(S)}$, resulting from the external motion of these bonded substructures should be the summation over all the bonded substructures to which atoms j belongs. For each bonded substructure S_i to which atom j belongs, its central atom i should also belong to S_j ($i \in S_j$). Hence, we have

$$\ddot{\mathbf{r}}_j^{(S)} = \sum_{i\in S_j}\mathbf{a}_{S_i} = \sum_{i\in S_j}\frac{1}{M_i}(\mathbf{f}_i^{(S)} + \lambda\mathbf{g}_i^{(S)}) \tag{11}$$

Because $\mathbf{g}_i^{(S)}$ was introduced to enhance specifically the systematic motion of atom i resulting from the external motion of bonded substructures, it can be thus calculated from eq 12.

$$\mathbf{g}_i^{(S)} = m_i\bar{\ddot{\mathbf{r}}}_i^{(S)} = \sum_{j\in S_i}\frac{m_i}{M_j t_1}\int_{t-t_1}^{t}(\mathbf{f}_j^{(S)} + \lambda\mathbf{g}_j^{(S)})\,\mathrm{d}\tau = \sum_{j\in S_i}\frac{m_i}{M_j}(\bar{\mathbf{f}}_j^{(S)} + \lambda\bar{\mathbf{g}}_j^{(S)}) \tag{12}$$

where $j \in S_i$ indicates that atom j belongs to substructure S_i. $\bar{\ddot{\mathbf{r}}}_i^{(S)}$ is the time average of the total acceleration of atom i resulting from the systematic motion of these bonded substructures.

Hence, a self-guided MD simulation with an enhanced (altered) systematic motion of bonded substructures for a molecular system can be performed according to eq 7. Equation 7 is called the equation of motion with an enhanced systematic motion for molecular systems. The force $\mathbf{f}_i$ on atom i is calculated in exactly the same way as that in the conventional MD simulation. The guiding force $\mathbf{g}_i^{(S)}$ on atom i is calculated according to eq 12. For monoatomic molecular systems, the bonded substructure becomes a single atom. In this case, eq 12 reduces to eq 6 and eq 7 reduces to eq 5.

Estimation of the Guiding Force. To improve computational efficiency for the calculation of the time average for

any property P, eq 13 was proposed.

$$\bar{P}(t) = \frac{\delta t}{t_l}\sum_{\tau=t-t_l}^{t} P(\tau) = \frac{\delta t}{t_l}\left(\sum_{\tau=t-t_l}^{t-\delta t} P(\tau) + P(t)\right) = \frac{t-\delta t}{t_l}\left(\frac{\delta t}{t-\delta t}\sum_{\tau=t-t_l}^{t-\delta t} P(\tau)\right) + \frac{\delta t}{t_l}P(t) \cong \left(1-\frac{\delta t}{t_l}\right)\bar{P}(t-\delta t) + \frac{\delta t}{t_l}P(t) \quad (13)$$

where $\bar{P}(t)$ is the time average of property P at time t, $P(\tau)$ is the value of P at time τ, δt is the size of one MD time step, and t_l is the averaging time.

Equation 13 shows that the time average of P at the current time step, $\bar{P}(t)$, can be estimated from the time average of the previous time step, $\bar{P}(t-\delta t)$, and the value of P at the current step, $P(t)$. This approximation can effectively save both memory and CPU time.

Accordingly, the guiding force in eq 12 can be approximately calculated according to eq 14.

$$\mathbf{g}_i^{(S)}(t) = \left(1-\frac{\delta t}{t_l}\right)\mathbf{g}_i^{(S)}(t-\delta t) + \frac{\delta t}{t_l}\sum_{j\in S_i}\frac{m_i}{M_j}(\mathbf{f}_j^{(S)}(t) + \lambda\mathbf{g}_i^{(S)}(t)) \quad (14)$$

t_l should be larger than or equal to the size of one MD time step, δt.

The Self-Guided MD Simulation Algorithm. A self-guided MD simulation is performed by numerically solving eq 7. In our implementation, the leapfrog algorithm[19,20] was employed, as shown in eqs 15 and 16.

$$\dot{\mathbf{r}}_i\left(t+\frac{\delta t}{2}\right) = \dot{\mathbf{r}}_i\left(t-\frac{\delta t}{2}\right) + (\mathbf{f}_i(t) + \lambda\mathbf{g}_i^{(S)}(t))\frac{\delta t}{m_i} \quad (15)$$

$$\mathbf{r}_i(t+\delta t) = \mathbf{r}_i(t) + \dot{\mathbf{r}}_i\left(t+\frac{\delta t}{2}\right)\delta t \quad (16)$$

where $\dot{\mathbf{r}}_i(t)$ is the velocity of atom i at time t, $\mathbf{f}_i(t)$ is the force on atom i at time t, $\lambda\mathbf{g}_i^{(S)}(t)$ is the guiding force on atom i at time t, and λ is the guiding factor.

Because of the extra guiding forces, now atoms move along a trajectory departing from the classical Newtonian dynamics. For microcanonical ensemble simulation, the total energy of the system should be conserved. This can be accomplished by scaling the velocities of atoms in the system after each time step. The velocities at the next half-time step, $t + \delta t/2$, are scaled according to eq 17.

$$\dot{\mathbf{r}}_i'\left(t+\frac{\delta t}{2}\right) = \chi_E\dot{\mathbf{r}}_i\left(t+\frac{\delta t}{2}\right) \quad (17)$$

where χ_E is called the energy conservation scaling factor, which was introduced to make the total energy conserved during each time step.

Hence, with the scaling factor χ_E, the kinetic energy at $t + \delta t/2$ is

$$E'_k\left(t+\frac{\delta t}{2}\right) = \frac{1}{2}m_i\dot{\mathbf{r}}_i'^2\left(t+\frac{\delta t}{2}\right) = \chi_E^2 E_k\left(t+\frac{\delta t}{2}\right) \quad (18)$$

According to eq 7, the kinetic energy change at each time step without the scaling factor χ_E can be expressed as

$$\delta E_k = E_k\left(t+\frac{\delta t}{2}\right) - E_k\left(t-\frac{\delta t}{2}\right) = \sum_i m\dot{\mathbf{r}}_i(t)\,\delta\dot{\mathbf{r}}_i = \sum_i m\dot{\mathbf{r}}_i(t)\,\ddot{\mathbf{r}}_i(t)\,\delta t = \sum_i(\mathbf{f}_i + \lambda\mathbf{g}_i^{(S)})\dot{\mathbf{r}}_i(t)\,\delta t \quad (19)$$

while the potential energy change is

$$\delta E_p = -\sum_i \mathbf{f}_i\,\delta\mathbf{r}_i = -\sum_i \mathbf{f}_i\dot{\mathbf{r}}_i(t)\,\delta t \quad (20)$$

With the scaling factor χ_E in eq 17, the total energy is thus conserved. Hence, the kinetic energy change should be equal to the negative potential energy change.

$$\delta E'_k = E'_k\left(t+\frac{\delta t}{2}\right) - E_k\left(t-\frac{\delta t}{2}\right) = \chi_E^2 E_k\left(t+\frac{\delta t}{2}\right) - E_k\left(t-\frac{\delta t}{2}\right) = -\delta E_p = \sum_i \mathbf{f}_i\dot{\mathbf{r}}_i(t)\,\delta t \quad (21)$$

From eqs 19 and 21, the following equation can be derived to calculate the scaling factor χ_E

$$\chi_E = \left(\frac{E_k\left(t-\frac{\delta t}{2}\right) + \sum_i \mathbf{f}_i\dot{\mathbf{r}}_i(t)\delta t}{E_k\left(t-\frac{\delta t}{2}\right) + \sum_i(\mathbf{f}_i + \lambda\mathbf{g}_i^{(S)})\dot{\mathbf{r}}_i(t)\delta t}\right)^{1/2} \quad (22)$$

where $\dot{\mathbf{r}}_i(t)$ is calculated by

$$\dot{\mathbf{r}}_i(t) = \dot{\mathbf{r}}_i\left(t-\frac{\delta t}{2}\right) + \frac{\delta t}{2m_i}(\mathbf{f}_i + \lambda\mathbf{g}_i^{(S)}) \quad (23)$$

For constant temperature simulation, Berendsen et al. proposed a velocity scaling method[21] that uses a velocity rescaling factor

$$\chi_B = \left(1 + \frac{\delta t}{t_T}\left(\frac{T}{\mathscr{T}} - 1\right)\right)^{1/2} \quad (24)$$

to maintain a constant temperature, where $\mathscr{T}$ is the current kinetic temperature and t_T is a preset time constant. In this work, we combined the energy conservation scaling factor χ_E and Berendsen's velocity rescaling factor χ_B to obtain a velocity scaling factor, χ_T for constant temperature simulation using the self-guided MD method.

$$\chi_T = \chi_B\chi_E \quad (25)$$

Alternative to the scaling method described above, a constant temperature simulation using the self-guided MD method can also be accomplished using the constraint method to keep the kinetic energy constant.[22−25] The constrained equation of the self-guided motion can be written as

$$m_i\ddot{\mathbf{r}}_i = \mathbf{f}_i + \lambda\mathbf{g}_i^{(S)} - \xi\dot{\mathbf{r}}_i \quad (26)$$

Where ξ is a kind of "friction coefficient", which varies to constrain the kinetic energy to a constant value. This constraint method has also been implemented but was not used in the simulations presented in this paper. Thus, the detailed algorithm of its implementation is not discussed here.

To summarize, the self-guided MD simulation is performed as follows.

(1) At time step t, calculate the force on each atom, $\mathbf{f}_i$. This calculation is exactly the same as that in a conventional MD simulation.

(2) Calculate the guiding force according to eq 14. Since eq 14 is an approximation, for convenience, the calculation is actually done using eq 14′.

$$\mathbf{g}_i^{(S)}(t) = \left(1 - \frac{\delta t}{t_l}\right)\mathbf{g}_i^{(S)}(t - \delta t) + \frac{\delta t}{t_l} m_i \sum_{j \in S_i} \frac{\mathbf{f}_j^{(S)}(t) + \lambda \mathbf{g}_j^{(S)}(t - \delta t)}{M_j} \tag{14′}$$

At the beginning, $\mathbf{g}_i^{(S)}(0)$ is simply set to 0.

(3) Forward the velocities to the next half-time step, $t + \delta t/2$, according to eq 15. For constant-energy simulations, the velocities are scaled according to eq 17. For constant-temperature simulations, the velocities are scaled by the scaling factor calculated according to eq 25.

(4) Forward the positions to the next time step, $t + \delta t$, according to eq 16. If bond lengths or other internal coordinates are to be fixed, the SHAKE algorithm developed by Ryckaert et al.[6] is applied to get the constrained positions and velocities.

(5) Repeat steps 1−4 until the end of simulation.

Simulation Conditions

In this paper, the self-guided MD simulation method was applied to two peptides, an alanine dipeptide and a neutral, 16-residue synthetic peptide. The structure of the alanine dipeptide is shown in Figure 1a. The alanine dipeptide system has been extensively studied by MD simulations and other computational methods.[26−34] The neutral, 16-residue synthetic peptide has a sequence of CH_3CO-AAQAAAAQAAAAQAAY-NH_2, where A refers to an alanine, Q refers to a glutamine, and Y refers to a tyrosine. This synthetic peptide was determined experimentally using the circular dichroism (CD) spectrum to have ca. 50% helix secondary structure in aqueous solution.[35,36] All-atom models were used to represent these peptides. For the purpose of comparison, we also carried out comparative simulations to these two peptides using the conventional MD method. Identical conditions were used when carrying out the conventional MD and the self-guided MD simulations except that in the conventional MD simulations, the guiding factor λ was set to be 0 in eq 7.

The self-guided MD simulation algorithm has been implemented in both the CHARMM[37,38] and the AMBER[39,40] programs. The MD simulations were performed using the CHARMM[37,38] (version 24) with its version 22 force field parameters or the AMBER (version 4.0) with its all-atom force field parameters. Both force field parameters[37−40] include bond length, bond angle, dihedral angel, and improper dihedral angle interactions for bonded interaction, Lennard-Jones 6−12 potential, and electrostatic interaction for nonbonded interaction. The AMBER force field[39,40] also includes an explicit hydrogen bond potential. In all simulations, a constant dielectric constant of 1 was used. The results obtained using the AMBER force field were all reported. For brevity, the results obtained using the CHARMM force field were reported only for the simulation of the 16-residue peptide in explicit water at 300 K using the conventional MD and the self-guided MD simulation methods.

Fully extended conformations were used as the initial conformation for these peptides. The aqueous solution for a peptide was built by immersing it into an equilibrium water box and deleting overlapping water molecules that were within 2.45 Å of the peptide. The TIP3P water model[41] was used to represent water molecules.

For simulations in vacuum, nonbonded interactions were calculated with no cutoff. For simulations in explicit water, nonbonded interactions were calculated with a cutoff value of 14 Å and a switching function[37] was applied from 8 to 12 Å to smooth the interaction change across the cutoff. The cubic periodic boundary condition was applied to aqueous solution systems. A time step of 0.002 ps was used in all simulations. A neighbor list was created for interaction calculation and updated every 10 MD time steps. The SHAKE algorithm[6] was applied to constrain all bonds. All simulations were conducted at constant temperature. The velocity scaling method was used to maintain a constant temperature and t_T in eq 24 was set to be 1 ps in all simulations.

Results and Discussion

The fundamental idea of the self-guided MD simulation method is to use a guiding force to achieve an enhanced systematic motion. However, introduction of an arbitrary force into the equation of motion in general will only alter the conformational distribution of the system but cannot improve the conformational searching efficiency in MD simulations. In the following section, we investigated the improvement in the conformational searching efficiency and the possible alteration in ensemble averages and the conformational distribution of the system through comparative simulations of the two peptide systems at different conditions using the conventional MD and the self-guided MD methods.

The Alanine Dipeptide System. *1. Simulations in Vacuum.* This peptide has a total of 22 atoms with two peptide bonds (Figure 1a). Two dihedral angles, ϕ for $C_1-N_1-C_\alpha-C_2$ and ψ for $N_1-C_\alpha-C_2-N_2$, in this molecule can freely rotate, and the rotation of these two torsion angles results in different conformations for the molecule. The conformation of this molecule can be described by the values of ϕ and ψ. Conformations with different ϕ and ψ angles have different potential energies because of the interactions between atoms in the molecule. Therefore, the conformational distribution is not uniform throughout the $\phi-\psi$ space. The $\phi-\psi$ distribution is characteristic for each amino acid, and this distribution can be shown by a Ramachadran plot, which depicts the conformational densities in a 2D plot with ϕ being the X-axis and ψ being the Y-axis.

Two comparative 10 000 ps simulations were performed on the peptide at 700 K with the AMBER force field.[39,40] For the purpose of comparison, a temperature of 700 K was chosen for both simulations because it was found that, at lower temperatures such as 300 K, a 10 000 ps simulation using the conventional MD method cannot obtain a proper sampling of the entire conformational space and a much longer simulation time would be required to obtain a proper sampling. The parameters for the self-guided MD simulation were chosen as $\lambda = 0.1$ and $t_l = 0.2$ ps. Trajectories were recorded every 0.2 ps in both simulations for analysis.

First, the ensemble average properties obtained from these two simulations were evaluated. The ensemble averages of temperature, the fluctuation of temperature, the total energy, the fluctuation of total energy, the potential energy, the fluctuation of the potential energy, the kinetic energy, and fluctuation of the kinetic energy obtained from both simulations are provided in Table 1. As can be seen, the values obtained from the conventional MD and the self-guided MD methods are essentially the same, indicating that the self-guiding force does not change significantly these ensemble average properties.

Next, the conformational distributions in the $\phi-\psi$ space (the Ramachadran plot) obtained from conventional MD and the self-

TABLE 1: Comparison of Ensemble Averages between the Conventional MD and the Self-Guided MD Methods for the Alanine Dipeptide Systems

	temperature (K)		total energy (kcal/mol)		kinetic energy (kcal/mol)		potential energy (kcal/mol)	
	average	fluctuation	average	fluctuation	average	fluctuation	average	fluctuation
conventional MD (vacuum)	699.94 ± 0.41	100.51 ± 1.80	11.97 ± 0.60	1.96 ± 0.38	27.12 ± 0.02	3.89 ± 0.07	−15.16 ± 0.59	4.54 ± 0.17
self-guided MD (vacuum)	699.69 ± 0.46	100.73 ± 1.89	11.86 ± 0.69	1.84 ± 0.41	27.11 ± 0.02	3.90 ± 0.07	−15.26 ± 0.69	4.50 ± 0.18
conventional MD (water)	302.94 ± 0.34	8.91 ± 0.27	−1661.3 ± 5.3	9.7 ± 2.2	396.42 ± 0.44	11.66 ± 0.35	−2057.7 ± 5.2	15.2 ± 1.4
self-guided MD (water)	303.15 ± 0.34	8.88 ± 0.28	−1658.2 ± 5.1	9.7 ± 2.1	396.69 ± 0.45	11.63 ± 0.36	−2054.9 ± 5.1	15.1 ± 1.4

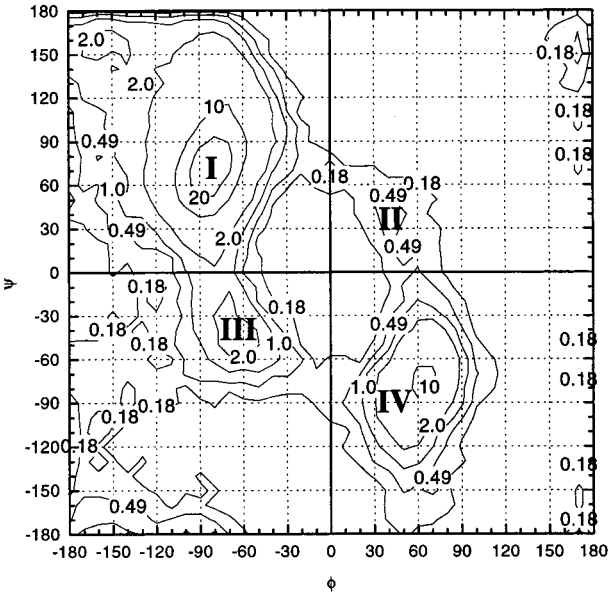

Figure 2. $\phi-\psi$ dihedral angle distribution (contour map) of the alanine dipeptide obtained from a 10 000 ps MD simulation in vacuum at 700 K using the AMBER force field.

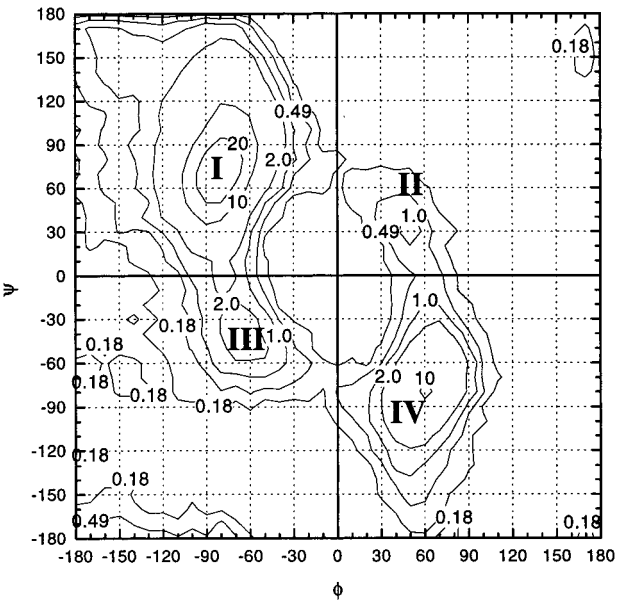

Figure 3. $\phi-\psi$ dihedral angle distribution (contour map) of the alanine dipeptide obtained from a 10 000 ps self-guided MD simulation in vacuum at 700 K using the AMBER force field. In the simulation, $\lambda = 0.1$ and $t_l = 0.2$ ps.

guided MD simulations are plotted in Figure 2 and Figure 3, respectively. Both maps revealed that there are four high-density regions, which are labeled as regions I−IV in Figures 2 and 3, respectively. These four high-density regions correspond to four local minima for the alanine dipeptide. In both contour maps, the ϕ and ψ values for these four local minima are (−75, 75), (55, 35), (−65, −45), and (65, −75), respectively. According to the size of each region, region I was defined as the area within 50° of peak I, region II as the area of 10° within peak II, region III as the area of 10° within peak III, and region IV as the area of 30° within peak IV, respectively. When these two contour maps were superimposed on each other, no significant difference was found. A small difference was observed for region II, where the self-guided MD simulation results in a sharper peak than the MD simulation. These results demonstrated that the conventional MD and the self-guided MD simulations yield almost identical results for the alanine dipeptide system in both the ensemble averages and the conformational distribution of the system.

TABLE 2: Transition Frequency between Local Minima in the Conventional MD and the Self-Guided MD Simulations

	I ↔ II	I ↔ III	II ↔ IV	III ↔ IV
conventional MD (vacuum)	10	302	59	29
self-guided MD (vacuum)	31	374	139	35
conventional MD (water)	11	29	3	1
self-guided MD (water)	16	47	5	3

The conformational searching efficiency of these two simulation methods was then compared. Four local minima were found in the conformational space as shown in Figures 2 and 3. Energy barriers exist between these four local minima, as evident from the low conformational density in the contour maps. The system needs to overcome these energy barriers in order to transfer from one region to another and to sample properly the $\phi-\psi$ conformational space during a simulation. Therefore, the frequency of the transitions between local minima is a measure of the ability to overcome the energy barriers as well as the conformational searching frequency. The transitions between local minima were counted for both simulations, and the results were summarized in Table 2. It is clear that more transitions between the minima occurred in the self-guided MD simulation than in the conventional MD simulation. These results suggest that an improved conformational searching efficiency was achieved using the self-guided MD simulation method.

2. Simulations in Aqueous Solution. The alanine dipeptide was solvated using 212 TIP3P water molecules, enclosed in a cubic box. The box size was set to make the density of the system 1.0 g/cm^3. Two comparative 5000 ps simulations were carried out at 300 K for the alanine dipeptide in aqueous solution using the conventional MD and the self-guided MD methods with the AMBER force field.[39,40] Systematic motion in explicit water is generally much slower than that in vacuum, and, for this reason, a relatively bigger λ value of 0.5 was used to increase the guiding effect of the self-guiding force and a relatively longer t_l of 2 ps was chosen for the self-guided MD

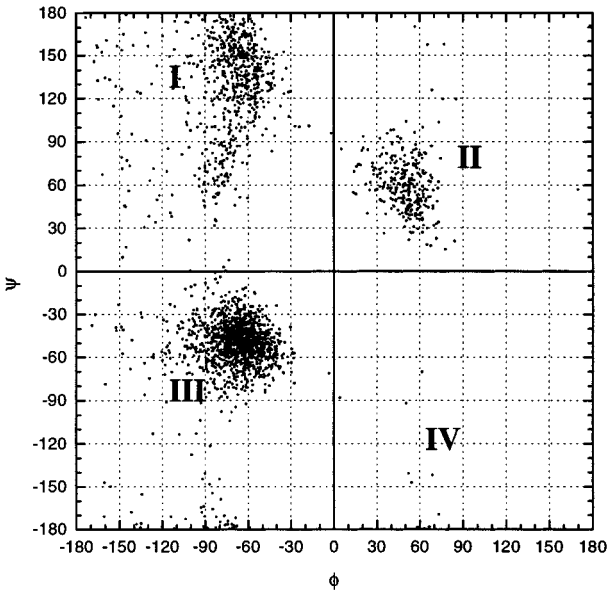

Figure 4. $\phi-\psi$ dihedral angle distribution of the alanine dipeptide obtained from a 5000 ps MD simulation in explicit water at 300 K using the AMBER force field.

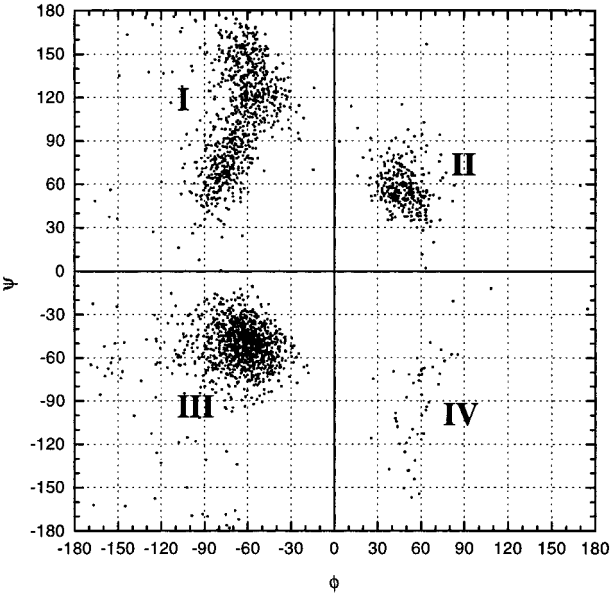

Figure 5. $\phi-\psi$ dihedral angle distribution of the alanine dipeptide obtained from a 5000 ps self-guided MD simulation in explicit water at 300 K using the AMBER force field. In the simulation, $\lambda = 0.5$ and $t_l = 2$ ps.

simulation. Trajectories were recorded every 2 ps in both simulations for analysis.

The same ensemble average properties examined for the simulations in vacuum were evaluated, and the results are also provided in Table 1. Again, the values obtained from the two simulations are essentially the same, indicating that the self-guiding force does not change significantly the ensemble average properties for the system in aqueous solution.

The conformational distributions in the $\phi-\psi$ space obtained from these two simulations are plotted in Figures 4 and 5, respectively. As can be seen, the conformational distributions from these two simulations are very similar to each other. Region III, which is located at the same position, around (−60, −60) on both maps, has the highest conformational density. Two other regions with high conformational density, I and II, are also located at the same position on both maps. Region IV, which can be easily identified as a local minimum in Figure 5, was only accessed a few times during the conventional MD simulation, as shown in Figure 4. The major difference between these two maps is that the conformations are more concentrated around the peaks in Figure 5 than in Figure 4, and this is particularly obvious for conformations in regions I and II. This suggests that the self-guided MD simulation accesses more often the low-energy conformations than the conventional MD simulation. However, it is important to note that the positions of the global minimum and other local minima are the same in both maps.

The conformational searching efficiency for the simulations in aqueous solution was then compared. As can be seen from Figures 4 and 5, four regions of local minima were found in the conformational space and energy barriers exist between these four local minima, as evident from the low conformational density in the contour maps. Four regions are defined as region I, 75° around (−60, 105); region II, 45° around (45, 60); region III, 50° around (−60, −60); IV, 50° around (60, −100). The transitions between the four regions were counted for both simulations, and the results are also summarized in Table 2. It is clear that more transitions between the four regions occurred in the self-guided MD simulation than in the conventional MD simulation.

Since each local minimum only covers certain range of ϕ and ψ angles, another way to examine the frequency of transitions between local minima is to examine the time profiles of ϕ and of ψ, which are plotted in Figure 6a for ϕ and 6b for ψ, respectively. As can be seen from Figure 6a, more frequent transitions occurred for the ϕ angle during the self-guided MD simulation than during the conventional MD simulation, and from Figure 6b, similar results were obtained for the ψ angle. These results clearly showed that the guiding force in the self-guided MD simulation method improves the conformational searching efficiency for the alanine dipeptide in the aqueous solution.

The 16-Residue Synthetic Peptide System. When a system is large, the conformational space of the system becomes more complicated, and during an MD simulation, the system is easily trapped into one local minimum or another. The proper sampling of the entire conformational space and the identification of the global minimum of a large system become an extremely difficult or impossible task to accomplish using the conventional MD simulation method at moderate temperature. In this section, we wish to report our simulation studies on the synthetic 16-residue peptide[35,36] both in vacuum and in explicit water. This peptide is neutral and water-soluble and has been determined to have ca. 50% helix secondary structure in aqueous solution.[36]

The helix is a common secondary structural motif found in proteins, and the mechanism of helix−coil interconversion is key to understanding the protein-folding problem.[42] One very direct approach to the mechanism of helix folding is through MD simulation. However, because of energy barriers and random walk, the helix-folding process of peptides is a difficult task for the conventional MD simulation. MD simulation has so far been successful to study the relative fast process such as the folding dynamics of a pentapeptide in water[43] and the conversion of 3_{10} helix to α-helix of a 16-residue, alanine-based peptide in water.[44] For folding simulations of larger peptides, modified force fields had to be used to circumvent the energy barrier problem.[45,46] Since our self-guided MD simulation method was developed primarily to speed up the conformational searching efficiency, it is, therefore, of great interest to

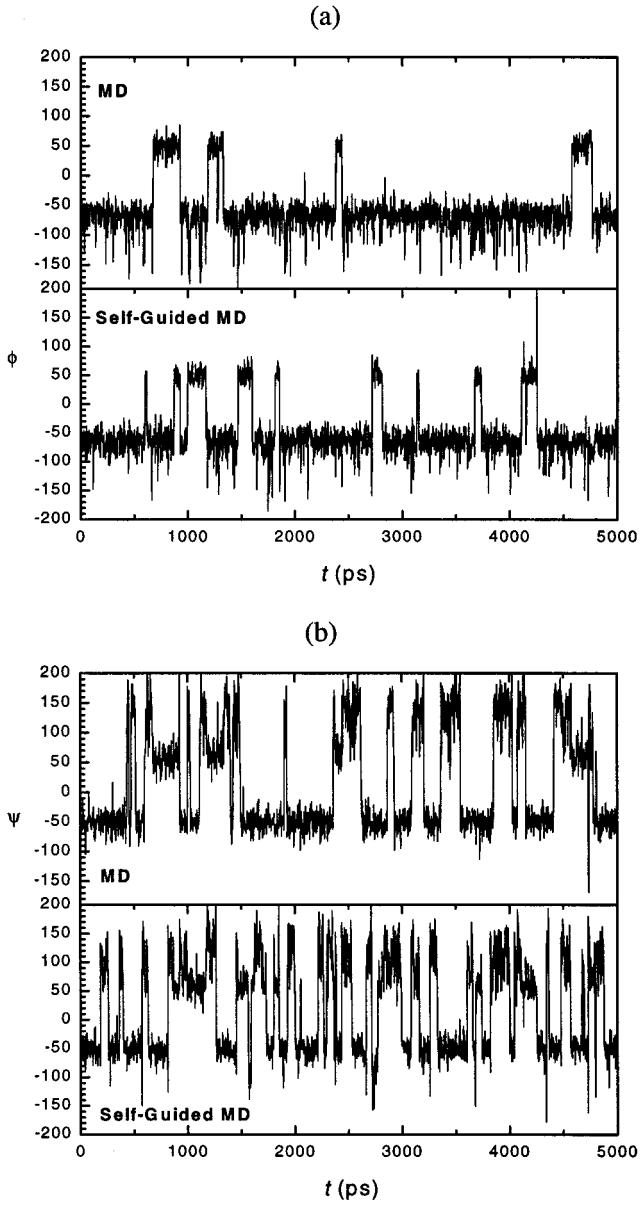

Figure 6. (a) ϕ and (b) ψ dihedral angle time profile of the alanine dipeptide obtained from 5000 ps MD or self-guided MD simulations in explicit water at 300 K using the AMBER force field.

investigate its efficiency through the helix-folding simulations of this 16-residue peptide.

1. Simulations in Vacuum. Two comparative 10 000 ps simulations were performed at 300 K using the conventional MD and the self-guided MD simulation methods using the AMBER force field.[39,40] In the self-guided MD simulation, λ of 0.1 and t_l of 0.2 ps were used. Trajectories were recorded every 10 ps during the simulations for analysis.

In the conventional MD simulation starting from a fully extended conformation, the peptide collapsed to a compact structure within 100 ps (Figure 7). Continued simulation to 10 000 ps did not result in much conformational change, indicating that the peptide was trapped in a local minimum. In the self-guided MD simulation at 300 K starting from the same fully extended conformation, the peptide folded into a helix structure within 100 ps (Figure 8). With continued simulation to 10 000 ps, significant conformational changes occurred in local regions of the peptide, but overall the peptide maintained the helix structure.

To further investigate the ability to overcome energy barriers, another self-guided MD simulation was performed starting from the "trapped structure" of the peptide obtained at the end of the 10 000 ps conventional MD simulation. It was found that within 1000 ps, the peptide folded into a complete helix structure (Figure 9), similar to that obtained from the self-guided MD simulation starting from the fully extended conformation (Figure 8). Continued self-guided MD simulation to 10 000 ps showed that the peptide underwent some conformational changes in local regions, but overall the helix structure was maintained.

The potential energies of the peptide in these three simulations are plotted against the simulation time in Figure 10. As can be seen, in the conventional MD simulation, for the first 500 ps, the potential energy of the peptide decreased quickly and reached −505 kcal/mol. After 500 ps, the potential energy of the peptide fluctuated between −505 and −515 kcal/mol throughout the remaining period of the 10 000 ps simulation. In the self-guided MD simulation starting from the extended conformation, the potential energy of the peptide reached −520 kcal/mol and the complete helix formed within 100 ps. The potential energy of the peptide fluctuated between −520 and −530 kcal/mol throughout the remaining period of the 10 000 ps simulation. In the self-guided MD simulation starting from "the trapped structure" obtained from the conventional MD simulation, it was found that the potential energy of the peptide first went up to −495 kcal/mol from −506 kcal to overcome a large energy barrier and then decreased quickly to −505 kcal/mol. At 500 ps, the potential energy went up again to −494 kcal/mol to overcome another large energy barrier and then decreased to −522 kcal/mol at ca. 1000 ps simulation, when a complete helix formed. After this point, the potential energy of the peptide fluctuated between −520 and −530 kcal/mol, similar to the self-guided MD simulation starting from the fully extended conformation. From Figure 10, it can be seen that the potential energy jumps are much more frequent in the self-guided MD simulations than in the conventional MD simulation.

This example demonstrates clearly that the self-guided MD simulation method has an extraordinary ability to overcome the energy barriers and to search out low-energy states of the system.

2. Simulations in Aqueous Solution. Folding of this 16-mer peptide in aqueous solution through MD simulations would be difficult to achieve because the systematic conformational change of the peptide in aqueous solution is very slow owing to the strong and random collisions of the peptide with its surrounding solvent molecules. Helix folding has been the subject of many simulation studies.[43,44,47−52] Successful de novo helix folding of large peptides in explicit water through MD simulations has not to the best of our knowledge been reported so far, presumably because the systematic conformational change of large peptides in explicit water is very slow.

A total of 1879 TIP3P water molecules were used to solvate the peptide with its extended conformation (Figure 11), and the density of the system was set to be 1.0 g/cm^3. The peptide was placed along the diagonal of the cube to minimize the box size. At the beginning of the simulation, the closest distance between the peptide and its images was 7.8 Å. As the simulation proceeded, the peptide began to fold and the closest distance between the peptide and its images became bigger. Overall, the interactions between the peptide and its images are not significant for the system.

Comparatively, 200 ps simulations were performed at 300 K with both the CHARMM[37,38] and AMBER[39,40] force fields using both the conventional MD and the self-guided MD methods.

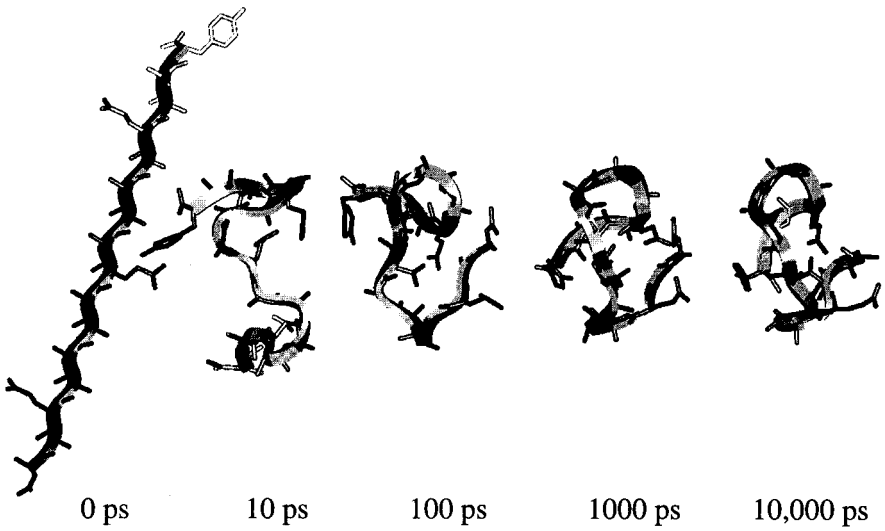

Figure 7. Snapshots of the conformations of the 16-residue peptide obtained from the 10 000 ps conventional MD simulation in vacuum at 300 K using the AMBER force field, starting from the fully extended conformation. For clarity, hydrogen atoms were not shown.

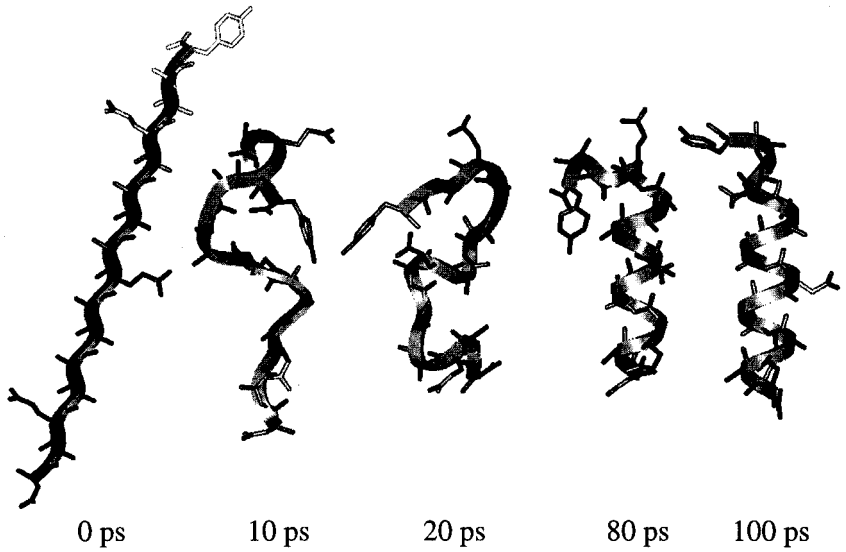

Figure 8. Snapshots of the conformations of the 16-residue peptide obtained from a 10 000 ps self-guided MD simulation in vacuum at 300 K using the AMBER force field, starting from the fully extended conformation. For clarity, hydrogen atoms were not shown. In the simulation, $\lambda = 0.1$ and $t_l = 0.2$ ps.

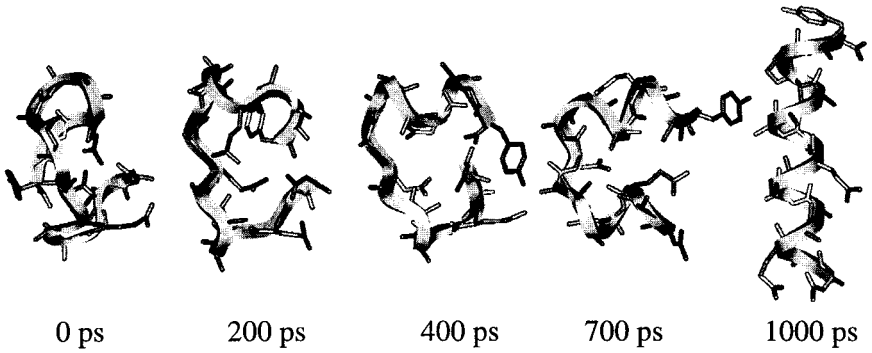

Figure 9. Snapshots of the conformations of the 16-residue peptide obtained from a 10 000 ps self-guided MD simulation in vacuum at 300 K using the AMBER force field, starting from a local energy minimum conformation obtained from the MD simulation in vacuum at 300 K. For clarity, hydrogen atoms were not shown. In the simulation, $\lambda = 0.1$ and $t_l = 0.2$ ps.

In the self-guided MD simulation, λ of 0.5 and t_l of 2 ps were used. Each of these 200 ps simulations took ca. 80 CPU hours on the SGI indigo2 system with a single R10000 processor.

During the 200 ps conventional MD simulations using either the CHARMM[37,38] or AMBER[39,40] force fields, no significant conformational changes occurred for the peptide. Snapshots of the structure of the peptide during the 200 ps conventional MD simulation using the CHARMM[37,38] force field were shown in Figure 12. The structure of the peptide was merely relaxed, and its N-terminal end was bent slightly. To observe significant

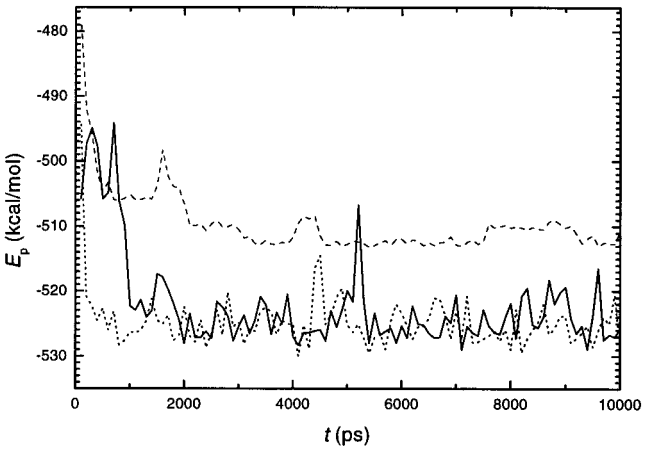

Figure 10. Potential energies of the 16-residue peptide during simulations in vacuum at 300 K using the AMBER force field. The points shown in this figure are subaverages over a period of 100 ps. The dashed line is the result obtained from the conventional MD simulation, the dotted line is the result obtained from the self-guided MD simulation starting from a fully extended conformation, and the solid line is the result obtained from the self-guided MD simulation starting from "the trapped structure" obtained at the end of the 10 000 ps conventional MD simulation.

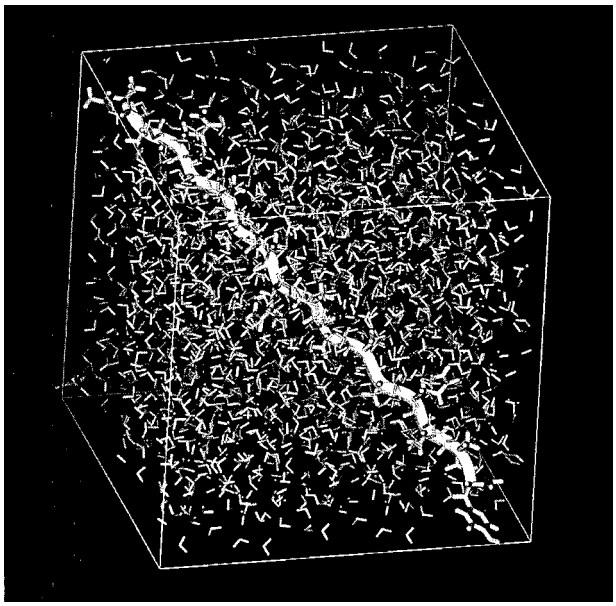

Figure 11. System of the 16-residue peptide in 1879 explicit waters. Cubic periodic boundary condition was applied in the simulations.

conformational changes of the peptide, much longer simulation time is required.

In contrast, significant conformational changes of the peptide were observed during the 200 ps self-guided MD simulation using either the CHARMM[37,38] or AMBER[39,40] force fields.

Below we described in detail the conformational change of the peptide during the 200 ps self-guided MD simulation using the CHARMM[37,38] force field. At 18 ps, the first hydrogen bond between the carbonyl group (CO) of A9 (A9 refers to alanine residue at the ninth position in the sequence) and the amido group (NH) of A12 was formed. At 20 ps another hydrogen bond between the CO of A7 and the NH of A10 was formed, and at 22 ps the first α-helix turn was formed between residues 8 and 12. The formation of this α-helix turn promoted a hydrogen bond between the CO of A10 and the NH of Q13 at 24 ps. At 26 ps, a 3_{10} helix segment was formed from residues 11 to 14. This helix segment propagated toward the

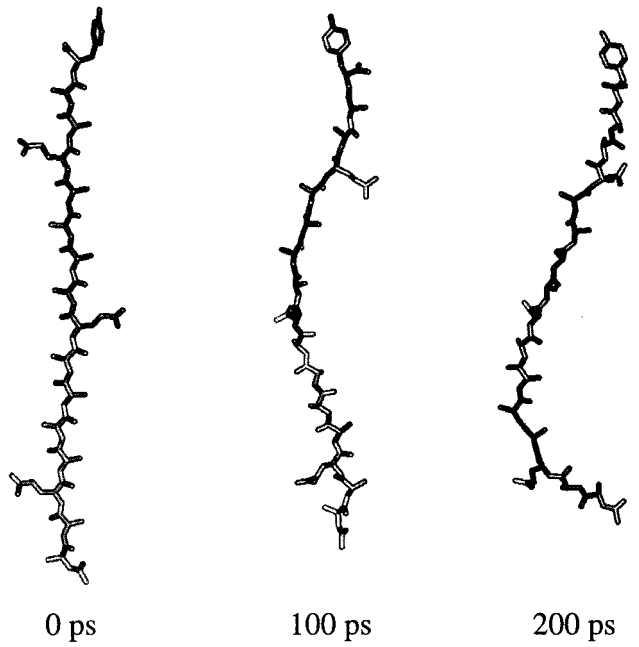

Figure 12. Snapshots of the conformations of the 16-residue synthetic peptide during a 200 ps conventional MD simulation in explicit water at 300 K using the CHARMM force field..

C-terminal, and at 40 ps the helix segment extended to residues from 8 to 16. This long helix segment began to break at 44 ps from the C-terminal, and the helix propagated toward N-terminal at 50 ps. At 66 ps, a hydrogen bond was formed between the CO of A1 and the NH of A5. At 70 ps, the helix segment shifted to residues from 5 to 11. At 74 ps, a 3_{10} helix was formed between residues 3 and 6 and another helix segment was formed from residues 6 to 12. At 80 ps, a long helix segment was formed from residues 1 to 13. At 86 ps, a helix was formed throughout the entire peptide with a slight distortion at its C-terminal. The helix structure was broken shortly after 100 ps into two helix segments, one from residues 1 to 7 and another one from residues 8 to 16. The conformational changes of the peptide continued to the end of the 200 ps simulation. A number of snapshots are shown in Figure 13. The results obtained from the 200 ps self-guided MD simulation using the CHARMM force field[37,38] are consistent with the experimental results that showed that this peptide has ca. 50% helix structure in aqueous solution.[36]

During the 200 ps self-guided MD simulation using the AMBER[39,40] force field, the 16-residue peptide also folded into variety of helical segments, but the structural details differed from that obtained using the CHARMM force field. A number of snapshots of the structure of the peptide during the self-guided MD simulation using the AMBER force field were shown in Figure 14. The analysis of the conformations of the peptide during the simulation using the AMBER force field showed that the helix form is the dominant structural elements of the peptide, again consistent with the experimental results.[36] The detailed analysis of the helix folding of this peptide is beyond the scope of this paper and will be the subject of a separate publication.

Taken together, our comparative simulations on the 16-mer peptide in aqueous solution using the conventional MD and the self-guided MD methods clearly showed that the self-guided MD simulation method has achieved a dramatic improvement in conformational searching efficiency. It is important to point out that the improvement in conformational searching efficiency does not depend on the force field used in the simulations,

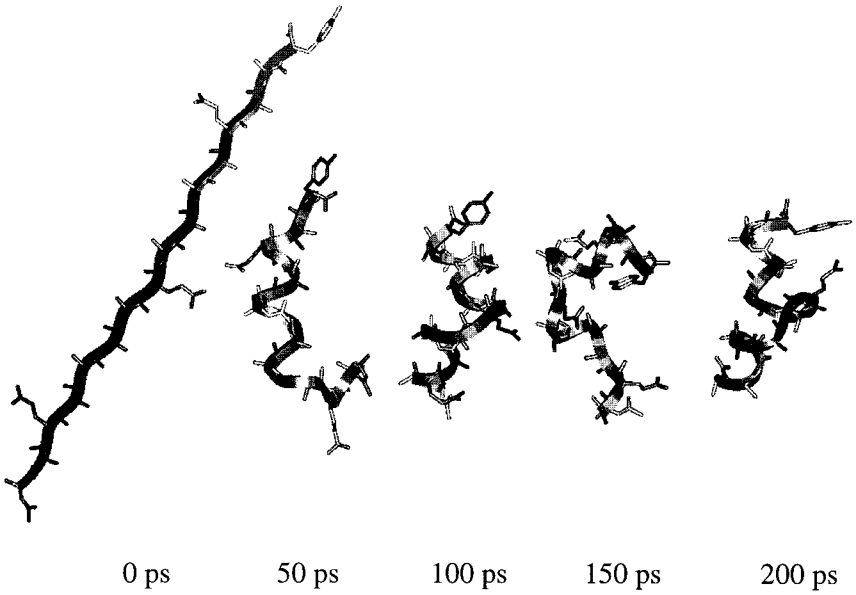

Figure 13. Snapshots of the conformations of the 16-residue synthetic peptide during a 200 ps self-guided MD simulation in explicit water at 300 K using the CHARMM force field. For clarity, hydrogen atoms were not shown. In the simulation, ($\lambda = 0.5$ and $t_l = 2$ ps.

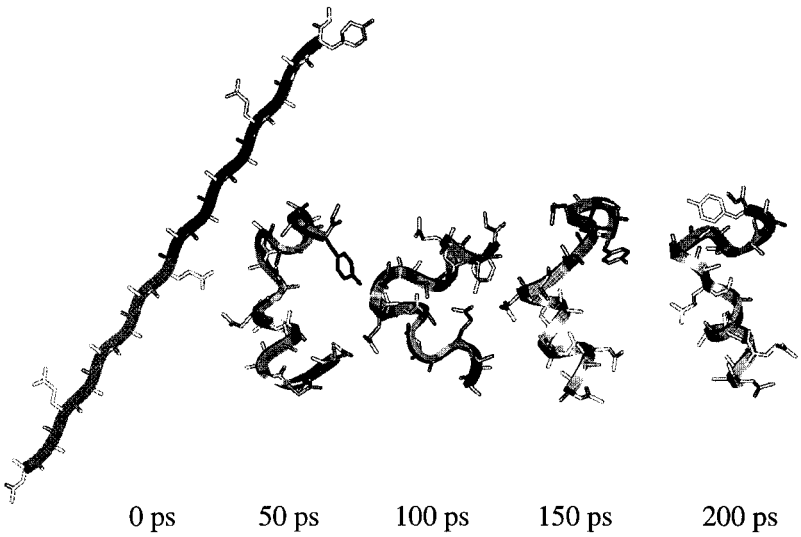

Figure 14. Snapshots of the conformations of the 16-residue synthetic peptide during a 200 ps self-guided MD simulation in explicit water at 300 K using the AMBER force field. For clarity, hydrogen atoms were not shown. In the simulation, ($\lambda = 0.5$ and $t_l = 2$ ps.

although the structural details clearly depend on the force field used in the simulation. Our results indicate that it is feasible to study the folding dynamics of relatively large peptide systems in explicit water using the self-guided MD simulation method.

Summary and Conclusions

We developed the self-guided MD simulation method primarily to improve the conformational searching efficiency in MD simulations. This method was developed on the basis of our understanding that the major reason that limits the conformational searching efficiency is the extremely slow inter- and intramolecular systematic motion. To enhance these systematic motions, we introduced a guiding force, termed the self-guiding force, into the equation of motion, in addition to the force derived from the force field. To enhance effectively both the inter- and the intramolecular systematic motions for molecular systems, bonded substructures were defined within a molecule.

The self-guided MD simulation method was applied to two peptides, an alanine dipeptide and a 16-mer synthetic peptide. The simulation studies on the alanine dipeptide system in vacuum and in explicit water showed that the self-guided MD and conventional MD simulations resulted in almost identical ensemble averages and similar conformational distributions, although the self-guided MD simulation seemed somewhat emphasized on the low-energy conformations. These results suggested that the guiding force did not alter significantly the thermodynamics properties of the peptide. In simulations both in vacuum and in explicit water, the conformational searching efficiency was improved by the guiding force, as evident from the more frequent transitions between local minima.

The simulations of the 16-mer synthetic peptide in vacuum clearly demonstrated an extraordinary ability of the self-guided MD simulation method to overcome energy barriers. Starting from the fully extended conformation, during the 10 000 ps conventional MD simulation, the peptide came to a compact structure with potential energy of ca. −505 kcal/mol at 100 ps and did not have much conformational changes thereafter during the remaining period of the 10 000 ps simulation. During the 10 000 ps self-guided MD simulation, starting from the same extended conformation, within 100 ps, the peptide folded into

a complete helix structure with a potential energy of ca. −525 kcal/mol, significantly lower than that obtained from the 10 000 ps conventional MD simulation. The peptide continued to have significant conformational changes thereafter, as evident from the frequent changes in potential energy. To further investigate the ability of the self-guided MD simulation method to overcome energy barriers of the system to reach lower energy states, another 10 000 ps self-guided MD simulation was performed starting from the "trapped structure" obtained at the end of the 10 000 ps conventional MD simulation. At the first 1000 ps of the self-guided MD simulation, two large energy barriers were overcome and the system reached conformational states with significant lower potential energy. The helix structure was formed at ca. 1000 ps. This example clearly showed an extraordinary ability of the self-guided MD simulation to overcome energy barriers and to reach conformational states with potential energy significantly lower than that obtained from a much longer period of a conventional MD simulation at the same conditions.

The simulations of the 16-residue synthetic peptide in explicit water using both the CHARMM and AMBER force fields demonstrated the dramatic improvement in conformational searching efficiency during the self-guided MD simulations. The peptide did not have significant conformational change during the 200 ps simulations using the conventional MD method. In contrast, the peptide underwent frequent and significant conformational changes during the 200 ps simulations using the self-guided MD method and folded into conformations with a variety of helical segments, consistent with experimental results.[36] These results indicated clearly that the guiding force dramatically speeds up the systematic conformational changes of the system. It is important to point out that the improvement of conformational searching efficiency does not depend on the force field used in the simulation.

The guiding effect can be adjusted through two parameters λ and t_l. A larger value of λ will speed up more the systematic conformational changes and improve more the overall conformational searching efficiency. However, a too large value of λ might result in more alteration in ensemble averages and the conformational distribution. A proper value of λ will significantly improve the overall conformational searching efficiency and, at the same time, does not significantly alter the thermodynamics properties of the system. Since t_l represents the averaging time, a relatively small value should be chosen for systems with fast conformational changes and a relatively large value should be chosen for systems with slow conformational changes to achieve the best guiding effect.

Since this new MD simulation method can dramatically improve the conformational searching efficiency, especially for large molecular systems, it may be used as a powerful tool to study many important yet slow processes or events that are difficult to study using the conventional MD method with current available computing power.

Acknowledgment. The authors thank Dr. Bill Milne of the National Institutes of Health, Dr. Bob Henderson of the Lakeland Community College, Dr. Dusanka Janezic of the National Institute of Chemistry, Slovenia, and Dr. Marc Nicklous of the National Institutes of Health for critical reading of the manuscript and many useful suggestions. Dr. Ming Liu of the Georgetown University Medical Center is acknowledged for many stimulating discussions. The financial support from the Department of Defense (DOD DAMD17-93-V-3018) to the Georgetown Institute for Cognitive and Computational Sciences and from the Lombardi Cancer Center is greatly appreciated. We thank the Georgetown University Molecular Modeling Center for the use of its computer facility funded from the National Science Foundation (NSF-CHE-9601976).

References and Notes

(1) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Clarendon Press: Oxford, 1987.
(2) Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631.
(3) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395.
(4) McCammon, J. A.; Harvey, S. C. *Dynamics of proteins and nucleic acids*; Cambridge University Press: Cambridge, 1987.
(5) Rapaport, D. C. *Comput. Phys. Rep.* **1988**, *9*, 1.
(6) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(7) Jain A.; Vaidehi, N.; G. Rodriguez, G. *J. Comput. Phys.* **1993**, *106*, 258.
(8) Mazur, A. K.; Dorofeev, V. E.; Abagyan, R. A. *J. Comput. Phys.* **1991**, *92*, 261.
(9) Swindoll, R. D.; Haile, J. M. *J. Comput. Phys.* **1984**, *53*, 289.
(10) Teleman, O.; Jonsson, B. *J. Comput. Chem.* **1986**, *7*, 58.
(11) Tuckerman, M. E.; Martyna, G. J.; Berne, B. J. *J. Chem. Phys.* **1990**, *93*, 1287.
(12) Tuckerman, M. E.; Berne, B. J.; Rossi, A. *J. Chem. Phys.* **1991**, *94*, 1465.
(13) Tuckerman, M. E.; Berne, B. J.; Martyna, G. J. A. *J. Chem. Phys.* **1991**, *94*, 6811.
(14) Janezic, D.; Orel, B. *Int. J. Quantum Chem.* **1994**, *51*, 407.
(15) Wesson, L.; Eisenburg, D. *Protein Sci.* **1992**, *1*, 227.
(16) Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. *Science* **1983**, *220*, 671.
(17) Johnson, M. W. *Simulated annealing and optimization*; American Science Press: Syracuse, NY, 1988.
(18) Wilson, C.; Doniach, S. *Proteins: Struct., Funct., Genet.* **1989**, *6*, 193.
(19) Hockney, R. W. *Methods Comput. Phys.* **1970**, *9*, 136.
(20) Potter, D. *Computational physics*; Wiley: New York, 1972; Chapter 5.
(21) Benrendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(22) Hoover, W. G. *Annu. Rev. Phys. Chem.* **1983**, *34*, 103.
(23) Hoover, W. G.; Ladd, A. J. C.; Moran, B. *Phys. Rev. Lett.* **1982**, *48*, 1818.
(24) Evans, D. J. *J. Chem. Phys.* **1983**, *78*, 3297.
(25) Evans D. J.; Morriss, G. P. *Comput. Phys. Rep.* 1984, **1**, 297.
(26) Smart, L.; Marrone, T. J.; McCammon, J. A. *J. Comput. Chem.* **1997**, *18*, 1750.
(27) Marrone, T. J.; Gilbson, M. K.; McCammon, J. A. *J. Phys. Chem.* **1996**, *100*, 1439.
(28) Pellegrini, M.; Gronbech-Jensen, N.; Doniach, S. *J. Chem. Phys.* **1996**, *104*, 8639.
(29) Kumar, S.; Payne, P. W.; Vasquez, M. *J. Comput. Chem.* **1996**, *17*, 1269.
(30) Schmidt, A. B.; Fine, R. M. *Mol. Simul.* **1994**, *13*, 347.
(31) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *101*, 5032.
(32) Smith, P. E.; Pettitt, B. M.; Karplus, M. *J. Phys. Chem.* **1993**, *97*, 6907.
(33) Schiffer, C. A.; Caldwell, J. W.; Stroud, R. M.; Kollman, P. A. *Protein Sci.* **1992**, *1*, 396.
(34) Tobias, D. J.; Brooks, C. L. *J. Phys. Chem.* **1992**, *96*, 3864.
(35) Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *J. Am. Chem. Soc*. **1991**, *113*, 5102.
(36) Scholtz, J. M.; Qian, H.; Robins, V. H.; Baldwin, R. L. *Biochemistry* **1993**, *32*, 9668.
(37) Brooks B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(38) MacKerell Jr., A. D.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc*. **1995**, *117*, 11946.
(39) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(40) Weiner, S.J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(42) Scholtz, J. M.; Baldwin, R. L. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 95.
(43) Tobias, D.; Mertz, J. E.; Brooks, C. L., III. *Biochemistry* **1991**, *30*, 6054.
(44) Tirado-Rives, J.; Maxwell, D. S.; Jorgensen, W. L. *J. Am. Chem. Soc*. **1993**, *115*, 11590.

(45) Sung S.-S.; Wu, X.-W. *Proteins: Struct., Funct., Genetics* **1996**, *25***,** 202.
(46) Sung S.-S.; Wu, X.-W. *Biopolymers* **1997**, *42***,** 633.
(47) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. *Proteins: Struct., Funct., Genetics* **1988**, *4*, 31.
(48) Daggett, V.; Levitt, M. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 353.
(49) Kolinski, A.; Skolnick, J. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 338.
(50) Kolinski, A.; Skolnick, J. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 353.
(51) Lattman, E. E.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 439.
(52) Tirado-Rives, J.; Jorgensen, W. L. *Biochemistry* **1993**, *32*, 4175.