A coarse-grained model and associated lattice Monte Carlo simulation of the coil–helix transition of a homopolypeptide

Yantao Chen, Qi Zhang, and Jiandong Ding[a)]
Department of Macromolecular Science, Key Laboratory of Molecular Engineering of Polymers, Fudan University, Shanghai 200433, China

(Received 12 May 2003; accepted 18 November 2003)

A new coarse-grained lattice model neglecting atomic details is proposed for the coil–helix transition and a new physical parameter is suggested to characterize a helical structure. In our model, each residue is represented by eight lattice sites, and side groups are not considered explicitly. Chirality and hydrogen bonding are taken into consideration in addition to chain connectivity and the excluded volume effect. Through a dynamic Monte Carlo simulation, the physical properties of the coil–helix transition of a single homopolypeptide have been produced successfully within a short computing time on the PC. We also examined the effects of the variation of chain configurations including chain size and chain shape, etc. A spatial correlation function has been introduced in order to characterize periodicity of a helical chain in a simple way. A propagation parameter and a nucleation parameter have also been calculated, which compares favorably with the results of the Zimm–Bragg theory for the coil–helix transition. © *2004 American Institute of Physics.* [DOI: 10.1063/1.1640667]

I. INTRODUCTION

Protein folding is a challenging problem in natural science.[1–5] Computer simulation or ''computer experiment'' constitutes one of the main research methodologies for folding studies. However, any computer simulation based upon all-atom details is, to our knowledge, rather time consuming to perform on present-day computers for long chains. It is thus strongly desirable to improve the modeling efficiency at the cost of neglecting atomic details to some extent. In such a model, the interactions are approximated to capture the important physical features[6–9] and the amino acid residues are coarse-grained by single monomers.

Molecular dynamics simulation based upon off-lattice models has been successfully used in studying protein folding, especially during the last decade.[10–19] The lattice Monte Carlo simulation based on a coarse-grained model ignoring atomic details is an alternative approach to model the protein folding process with very high efficiency, although off-lattice Monte Carlo simulations are also available for the coil–helix transition.[20–23] Lattice models have been criticized due to its artificial lattice restriction. Nevertheless, it is generally appreciated that a suitable lattice model may capture some key features of a protein, for instance, chain connectivity, excluded volume effect, etc.[24–26] A coarse-grained model may not be sufficient for the investigation of a specific protein, but is meaningful for revealing the universal behavior of protein, as protein is believed to exhibit simplicity in many aspects.[27,28] In fact, simple statistical models have succeeded in studies of folding-related problems reported recently.[20,24,29–34]

Protein is folded accompanied by a series of conformational transitions such as a coil–globule transition and a coil–helix transition. Presently, much attention has been paid to the coil–globule transition of a polymer chain via Monte Carlo simulations.[24,33,35,36] The coil–helix transition has also been investigated by off-lattice Monte Carlo simulations.[20,37] In contrast, lattice Monte Carlo simulations of the coil–helix transition are rarely reported.[38–41] In this paper, we propose a simple coarse-grained lattice model for the coil–helix transition of a homopolypeptide and examine the thermodynamic properties as the first step of our Monte Carlo simulation of protein folding.

The structure of an α helix is one of the most common secondary structures in native proteins, and plays a very important role for protein to perform a function. Thus, it is important to explore the physical properties of the formation of an α helix. Many factors stabilizing an α helix have been examined quantitatively due to recent advances in experimental technology.[42,43] The basic mechanism of α helix formation has also been investigated according to simplified theoretical models.[44–49] It is commonly known that an α helix is right handed with 3.6 residues per helical turn where a hydrogen bond forms between residue pairs separated by a four-residue interval in sequence.[50] Nevertheless, present characterization parameters of helices are still not sufficient in simulation.

The remaining parts of this article are organized as follows: In Sec. II we describe the coarse-grained model we proposed and the dynamic Monte Carlo simulation technique we employed. The results and some detailed discussions are presented in Sec. III, where a helix with regular periodicity can be obtained and a new statistical parameter, spatial correlation function, is introduced to characterize the helix. A comparison between the Zimm–Bragg theory and our simulations is made in terms of the propagation parameter

[a)]Corresponding author. Electronic mail: jdding1@fudan.edu.cn

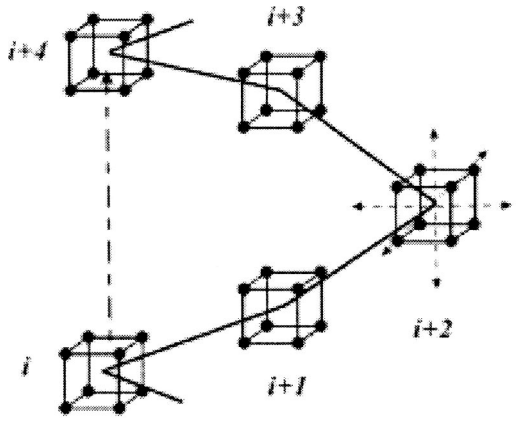


FIG. 1. Schematic presentation of the coarse-grained lattice model. Each residue occupies eight lattice sites leading to 87 permitted bond orientations. A site that has already been occupied by a residue cannot be occupied by another one. The elementary movement is set to allow a ''half-residue'' jump to four nearest-neighbor vacant sites along the six marked directions. A hydrogen bond might be formed between residues separated by a sequence interval of four.

and nucleation parameter. Some conclusions are drawn in Sec. IV.

II. MODEL AND SIMULATION APPROACH

A. Coarse-grained lattice model for helix

In our coarse-grained lattice model, a residue is modeled by a basic unit. A single homopolypeptide chain is considered here and the sequence dependence of an actual polypeptide chain is tentatively not included in this model.

Because the bond fluctuation model is employed, one residue can occupy eight sites in three dimensions, as shown in Fig. 1, and the bond length may fluctuate within a defined range.[51–54] This model was originally proposed to deal with commodity polymers with the advantage of being not as coarse as in the conventional single-site cubic lattice model and being easy to incorporate branching points. (For example, the disulfide bond as well as sequenced chains have been studied by us and the results will be published elsewhere.) The so-called ''sub-bead'', i.e., the occupation of a lattice site can only move to a nearest-neighboring vacant site in an elementary movement.

One of the main advantages of the present bond-fluctuation model over the conventional single-site lattice model is that the number of the permitted bond orientation in three dimensions is as large as 87 and thus the present lattice model can be considered quasicontinuous or quasi-off-lattice. Compared to studies of the coil–globule transition, this advantage is more important in studies of the coil–helix transition, because a ''realistic'' helix is hard to be formed from a small number of permitted bond angles unless the permitted angles are just the targeted ones. Furthermore, one bond is not allowed to cross another in the simulation for the microrelaxation modes defined by the bond fluctuation model.[52] On the other hand, as emphasized by Deutsch and Binder,[52] too many permitted bond angles (namely, more than 87 bond orientations in the bond-fluctuation model) or bond lengths would lead to the crossover of two bonds in some elementary movements or microrelaxation modes, which is incorrect to chain dynamics. Because we employed the microrelaxation modes defined by Deutsch and Binder,[52] the excluded volume effect has thus been taken into consideration in the present model both statically and kinetically.

In this work, we aim to construct an α helix. Thus, hydrophobicity is not included for simplicity. The hydrophobic interaction is prevailing in a coil–globule transition.[55] To describe protein's secondary structures, hydrogen bonds and chirality are more important, and are dealt with in a simplified way in our model. The energy function for our helix model is written as

$$E = E_l + E_\theta + E_{\mathrm{ch}} + E_{\mathrm{HB}}, \tag{1}$$

where the four terms on the right side of Eq. (1) represent fluctuations of the distance between consecutive residues, fluctuations of angle, chirality of residue, and hydrogen bonding between residues.

The potential for bond length fluctuations between two consecutive residues or α carbons is given by

$$E_l = \sum_{i=1}^{N-1} u_{l,i} = \sum_{i=1}^{N-1} \frac{1}{2} k_l (l_i - l_o)^2, \tag{2}$$

and that for bond angle fluctuations is given by

$$E_\theta = \sum_{i=2}^{N-1} u_{\theta,i} = \sum_{i=2}^{N-1} \frac{1}{2} k_\theta (\cos\theta_i - \cos\theta_o)^2. \tag{3}$$

Here N is the number of residues per polypeptide, and k_l and k_θ are the corresponding elastic constants for the associated harmonic potentials; l_i is the module of the bond vector $\mathbf{l}_i$ that denotes the vector connecting residue i to residue $i+1$; θ_i is the angle between $\mathbf{l}_{i-1}$ and $\mathbf{l}_i$; θ_o and l_o are equilibrium values. θ_o is chosen as 88.9°, according to the real value in the α helix. l_o is set to be $\sqrt{6}$, because the microrelaxation mode or basic local movement around this bond length is highly probable in the bond fluctuation model.[52]

While fluctuations of the bond length and angle are introduced mainly to let residue be movable, the other two terms in Eq. (2) are really important in the helix formation and description. In our coarse-grained model, the side-group effect is not reflected explicitly, but it is considered indirectly because chirality is necessary to stabilize a helix. Therefore, we assume the following potential to include chirality:

$$E_{\mathrm{ch}} = \sum_{i=1}^{N-3} u_{\mathrm{ch},i}, \tag{4}$$

where

$$u_{\mathrm{ch},i} = \begin{cases} \epsilon_{\mathrm{ch}} & (\mathbf{l}_i \times \mathbf{l}_{i+1} \cdot \mathbf{l}_{i+2} \leqslant 0), \\ 0 & (\mathbf{l}_i \times \mathbf{l}_{i+1} \cdot \mathbf{l}_{i+2} > 0). \end{cases} \tag{5}$$

According to this potential, a non-right-handed orientation for three consecutive bonds is penalized with a positive energy ϵ_{ch}.

A hydrogen bond is another key factor for a helix. Limited by our coarse-grained model, the hydrogen bond is artificially assumed to form between two residues with a four-residue interval in sequence and with suitable spatial distance. The hydrogen bond potential E_{HB} is given by

$$E_{\mathrm{HB}}=\sum_{i=1}^{N-4} u_{\mathrm{HB},i}, \tag{6}$$

where

$$u_{\mathrm{HB},i}=\begin{cases} 0 & (h_{i,i+4}<h_1), \\ -\epsilon_{\mathrm{HB}} & (h_1\leqslant h_{i,i+4}\leqslant h_2), \\ 0 & (h_{i,i+4}>h_2). \end{cases} \tag{7}$$

Here, we assume that only the $(i\pm4)$th residue can form a hydrogen bond with the ith residue, as shown in Fig. 1. $h_{i,i+4}$ is the module of the bond vector $\mathbf{h}_{i,i+4}$ that denotes the vector connecting residue i to residue $i+4$. The hydrogen bond interaction occurs within the range from h_1 to h_2. This range may, in principle, be estimated by the ratio of the hydrogen bond's length over the distance between two consecutive α carbons in a natural α helix. Further, by trial and error, we set $h_1=\sqrt{11.6}$ and $h_2=\sqrt{12.4}$ in this simulation. Once a hydrogen bond forms, the system gains an energy $-\epsilon_{\mathrm{HB}}$.

The other parameters of our energy function are determined mainly by trial and error. The values used in this paper are listed as follows: $\epsilon_{\mathrm{ch}}=\epsilon_{\mathrm{HB}}/2$, $k_l=\epsilon_{\mathrm{ch}}$, $k_\theta=\epsilon_{\mathrm{ch}}/2$.

B. Simulation approach

The dynamic Monte Carlo simulation was performed on a cubic lattice. A single polypeptide containing 32 residues was modeled in a box composed of $36\times36\times36$ lattice sites, unless otherwise indicated. A periodic boundary condition was set along each dimension. In each attempt, we first select a residue randomly. Then, one of the nearest neighbor ''site groups'' along six principal directions in a simple cubic lattice system is selected randomly,[51–54] as shown in Fig. 1. Every ''site group'' is composed of four sites in the three-dimensional eight-site bond fluctuation model. The coordination number is 6 for ''site groups'' and thus 24 for sites. The residue cannot be moved to the position of the selected site group, unless all sites in this group are vacant and the bond length restriction is obeyed (the number of the permitted bond orientation is 87). After the attempt of a move, Metropolis importance sampling[56] is employed as the criterion for acceptance or rejection in every attempt. The acceptance probability of the tried state is set as minimum of 1 and $\exp[-\Delta E/(k_{\mathrm{B}}T)]$, where ΔE is the energy difference between attempted state and old state, and $k_{\mathrm{B}}T$ is the Boltzmann constant multiplied by the absolute temperature. One Monte Carlo step (MCS) is defined as N attempts so that each residue in the polypeptide has a chance to move on average. Based upon the interpretation of Metropolis sampling via the master equation by Binder,[57] the dynamic Monte Carlo simulation is justified to study kinetics, because the number of MCS' is proportional to physical time. Different from multicanonical Monte Carlo simulation,[20,22,58] the present approach might be applied to the investigation of folding kinetics, although only thermodynamic behaviors have been examined in the first paper of our research in this series.

In addition to Metropolis sampling, thermal annealing is also employed to avoid local minima that are present in the energy landscape of protein folding. At every temperature in annealing, the chain was relaxed sufficiently to reach thermodynamic equilibrium before equilibrium statistics was collected. In this study, there were 40 temperatures simulated from an athermal state to the lowest simulated temperature for each trajectory. Each trajectory was initiated from different random-coil states. At each temperature, w MCS' are used for thermodynamic equilibration, and then another w MCS' are used for the collection of statistics. Because chains relax slowly at low temperature, the relaxation time and statistics time are elongated with the decrease of temperature: for the first 1/3 examined inverse temperatures (in a high-temperature regime), $w=w_o$; for the 1/3–2/3 inverse temperatures, $w=2w_o$; for the 2/3–5/6 inverse temperatures, $w=3w_o$; in the lowest temperature range, $w=4w_o$. $w_o=450\,000$ for $N=32$. The program was coded via FORTRAN by us. For $N=32$, only about 45 h was spent to run all of the 50 trajectories on a PC with 2.4 GHz CPU. For $N>32$, the relaxation time and statistics time will be elongated correspondingly and the number of lattice sites will be enlarged.

III. RESULTS AND DISCUSSION

A. Specific heat peak and thermodynamic transition

To quantitatively describe the coil–helix transition, we first study the reduced specific heat at different temperatures,

$$c_V^*=\frac{C_V}{Nk_{\mathrm{B}}}, \tag{8}$$

which was calculated from energy fluctuations,

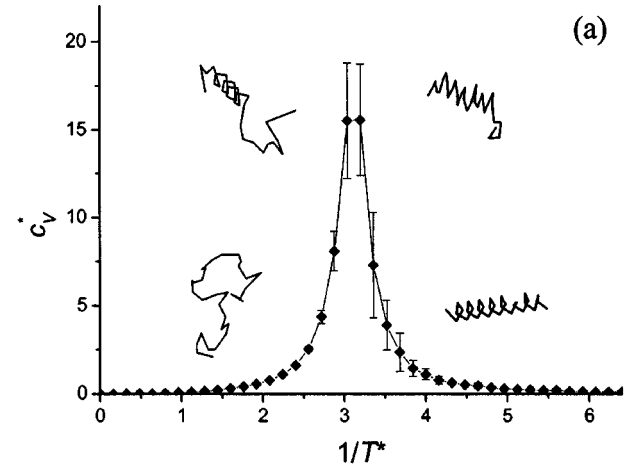
$$C_V=\frac{\langle E^2\rangle-\langle E\rangle^2}{k_{\mathrm{B}}T^2}. \tag{9}$$

Here, ''$\langle\ \rangle$'' denotes the configurational average.

The calculated reduced specific heat as a function of temperature is shown in Fig. 2, where we have used the reduced, dimensionless temperature $T^*(\equiv k_{\mathrm{B}}T/\epsilon_{\mathrm{HB}})$. A peak occurs around the transition energy $(1/T_c^*=3.12)$ with large fluctuations. Some snapshots are also shown in Fig. 2(a) as an inset, where both coil and helix states can be observed. The conformation of helix has been reproduced successfully and a structural transition exists in our simulated system. In contrast to fluctuation energy of bond length and that of bond angle, hydrogen bonding energy and chirality energy has a dominating effect in the coil–helix transition, as shown in Fig. 2(b).

B. Size and shape variations near the coil–helix transition

To describe chain configurations, one conventionally uses the mean square end-to-end distance $\langle R^2\rangle$ and mean square radius of gyration $\langle S^2\rangle$. The simulated results are displayed in Fig. 3. As the temperature is decreased, $\langle R^2\rangle$ and $\langle S^2\rangle$ decrease first, and increase below the coil–helix transition. It is well known that for a random coil, these two parameters should be proportional to each other. This is, however, not true in the case of the coil–helix transition. The



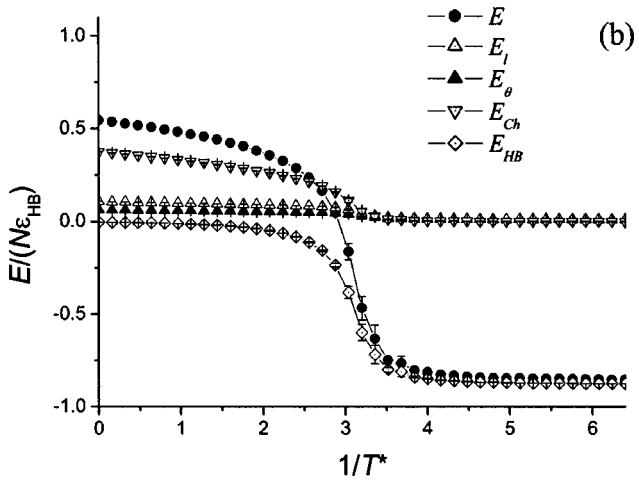


FIG. 2. Thermodynamic parameters as functions of the reduced inverse temperature $1/T^*$. (a) Reduced specific heat c_v^*. Insets are typical snapshots of the configurations at associated $1/T^*$; (b) Chain energy E and its four sub-terms in Eq. (1). The error bars come from the standard deviations of all trajectories.

two ends are separated in a rodlike helix, which accounts for the rapid increase of the end-to-end distance after the transition.

To describe the chain's shape extensively, we calculate the tensor of radius of gyration,[53,59]

$$\mathbf{R}=(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)=\begin{pmatrix} x_1 & x_2 & \cdots & x_N \\ y_1 & y_2 & \cdots & y_N \\ z_1 & z_2 & \cdots & z_N \end{pmatrix}. \tag{10}$$

Here, $\mathbf{R}$ is the position matrix for a chain at a specific configuration, and $\mathbf{r}_i$ is the position vector of the ith residue relative to the center of mass of the entire chain. Then, the tensor of radius of gyration $\mathbf{S}$ can be defined as

$$\mathbf{S}=\frac{1}{N}\mathbf{R}\mathbf{R}^T, \tag{11}$$

where $\mathbf{R}^T$ is the transposed matrix of $\mathbf{R}$. (The trace of the matrix $\mathbf{S}$ reduces to the square radius of gyration S^2.) After the diagonalization of $\mathbf{S}$, three principal axes of the radius-of-gyration tensor can be obtained. The ratios among the three principal axes $L_{\max}$, L_{mid}, and $L_{\min}$ may be used to evaluate the average shape of the simulated chain.

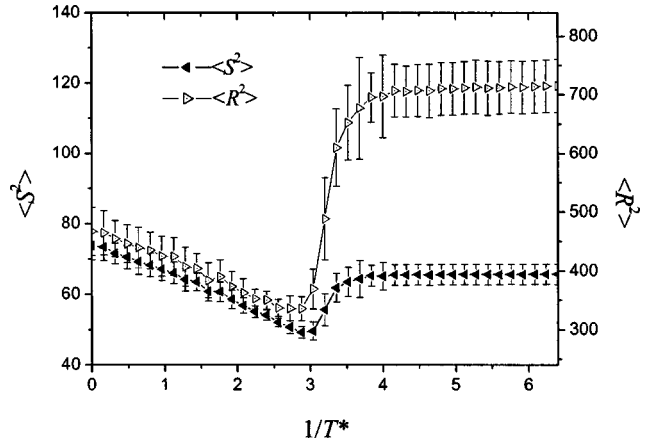


FIG. 3. Mean square radius of gyration $\langle S^2\rangle$ and mean square end-to-end distance $\langle R^2\rangle$ of the model as a function of the reduced inverse temperature $1/T^*$.

The results for the coil–helix transition are shown in Fig. 4. The increase of the ratio of $L_{\max}^2/L_{\min}^2$ indicates that the chain is elongated as T is lowered. In the helical state, the ratio of L_{mid}^2 over $L_{\min}^2$ is about 1.45, which implies that the eventual helix structure at an extremely low temperature is similar to a cylinder with a single-axis symmetry, for which the ideal $L_{\mathrm{mid}}^2/L_{\min}^2$ should have a value of unity.

C. Helix number and length

If the temperature is not sufficiently low, there might be multiple helical domains in the chain. According to the Zimm–Bragg method,[44] which is employed here as the criterion for recognizing a helical segment, the ith residue is in the helical state when it is bonded with the $(i-4)$th residue; consecutive, unbroken helical residues constitute one helix domain in a chain. The maximum number of a helical domain is thus $N-4$. Based on this criterion, we can calculate the average length of a helical domain L_h and the average number of helix in a chain N_h, which are shown in Fig. 5.

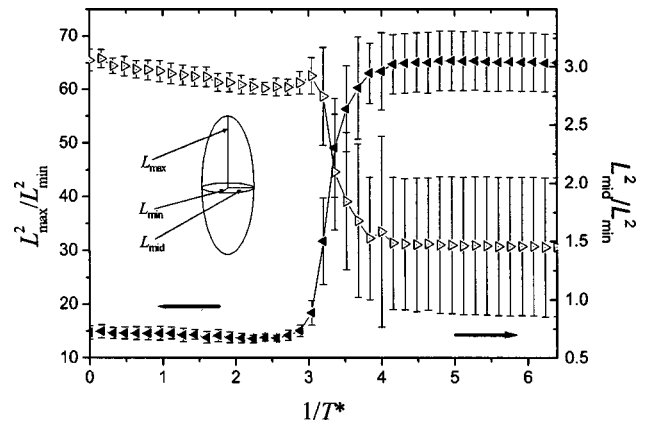


FIG. 4. Variations of the ratios of principal axes for the radius-of-gyration tensor as a function of the reduced inverse temperature $1/T^*$.

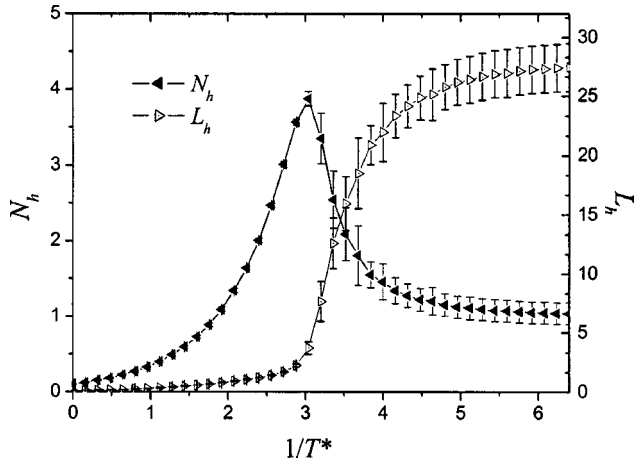


FIG. 5. Averaged number of the helix per chain N_h and averaged length of each helix L_h as functions of the reduced inverse temperature $1/T^*$.

As shown in Fig. 5, N_h first increases sharply as T is lowered, then decreases to about 1, and meanwhile L_h reaches its maximum, $32-4=28$ for chain length $N=32$ in Fig. 5. The peak value of N_h increases if the total number of residue is increased (data not shown). The peak of the curve for N_h and that of specific heat appear at very similar temperatures. Therefore, the process of helix formation can be deduced from these measurements. At the first step, many partial or short helices are formed, and at the transition temperature the number of helices reaches a maximum; as the second step, those partially formed helices merge to each other leading to a single helical structure (from multidomain to monodomain). Thus, helix formation is accompanied by multiple nuclei.

D. Periodicity of the resultant helix

At a finite temperature the stable helical structure may not be perfect. As an example, three snapshots of configurations of different chain lengths at the same reduced temperature are shown in Fig. 6. We are then faced with a problem of how to calculate the period of a helix and estimate the regularity of a helix. To resolve this problem, we employ the concept of a correlation function and suggest the following form to describe the periodicity of a helix.

The orientational correlation function is defined by us as

$$G(m)=\frac{1}{N-3}\sum_{i=2}^{N-2} g(m,i), \tag{12}$$

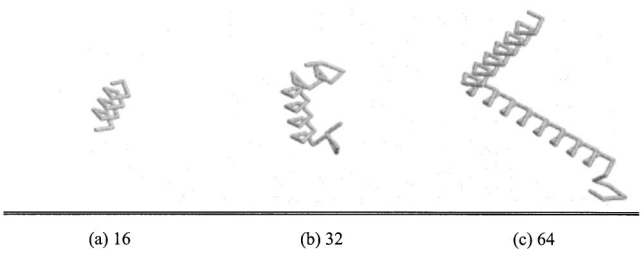


FIG. 6. Typical helix snapshots of three chains with different chain lengths at a fairly low temperature.

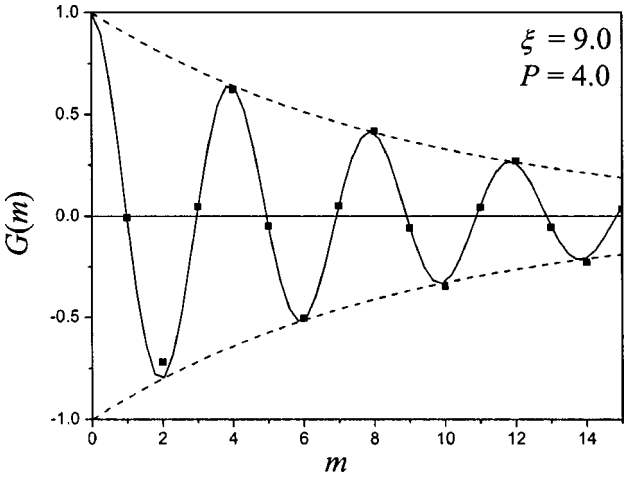


FIG. 7. Spatial correlation function $G(m)$ of orientation versus m, residue interval in sequence. The solid line is the fitted curve from Eqs. (15) with marked correlation length ξ and period of helix P, while the dashed lines indicate the damping contour described by Eq. (16).

where

$$g(m,i)=\frac{\sum_{j=1}^{N-m-1}(\cos\theta_{i,j}-\overline{\cos\theta_{i,j}})(\cos\theta_{i,j+m}-\overline{\cos\theta_{i,j}})}{\sum_{j=1}^{N-1}(\cos\theta_{i,j}-\overline{\cos\theta_{i,j}})^2} \tag{13}$$

and

$$\cos\theta_{i,j}=\frac{\mathbf{l}_i\cdot\mathbf{l}_j}{|\mathbf{l}_i\cdot\mathbf{l}_j|}. \tag{14}$$

In this function, m means the sequence interval, which ranges from 0 to $N/2-1$ in our calculation. $\overline{\cos\theta_{i,j}}$ denotes the average of $\cos\theta_{i,j}$ over j from 1 to $N-1$.

A typical curve for $G(m)$ is shown in Fig. 7 for a helix at low temperature. The oscillatory behavior of this curve reflects the periodic variation of monomer positions in a helix. Then we introduced the following equation to describe the spatial correlation function of a helix,

$$G(m)=\exp(-m/\xi)\cos(2\pi m/P). \tag{15}$$

Here, two parameters are needed to fit the data, ξ and P; these parameters can be interpreted as the orientational correlation length and period of helix, respectively. The simulation data is fitted by this equation very satisfactorily. The contour of the damping part in Fig. 7 closely follows the equation

$$G(m)=\pm\exp(-m/\xi). \tag{16}$$

In contrast to Eq. (15), the correlation function at an athermal state corresponding to a random coil can only be described as $G(m)=\exp(-m/\xi)$ (data not shown).

The calculated periods and correlation lengths vary with temperature and are shown in Fig. 8. The helix period P is about 4 at low temperature, which is close to 3.6 in a real α helix. ξ increases as the temperature is decreased. The correlation length defined above has the physical meaning of the persistent length defined in a wormlike chain. Figure 8 demonstrates that a chain in the helical state is persistent for a

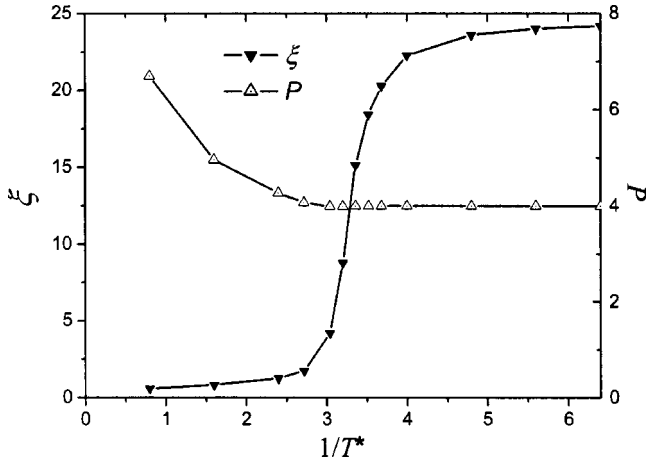


FIG. 8. Averaged correlation length ξ and period of helix P as functions of the reduced inverse temperature $1/T^*$. The lines are connected between the two adjacent data points just for the guidance of eyes. Chain length $N=32$.

longer distance as the temperature is decreased, as the result of the increase of length of each helical domain and also the increase of the regularity of the helix.

Figure 9 indicates that for long chains, ξ is not sensitive to the chain length at a given temperature, although for short chains, ξ seems to be proportional to chain length N (or $N-4$) because just one helix segment exists in a short polypeptide at low temperature. Thus, the rigidity of the polypeptide helix can be quantitatively calibrated by its correlation length. Equations (12)–(16) are also applicable for multihelices in a single chain or helix in multiple chains.

E. Comparison with the Zimm–Bragg theory

The classical Zimm–Bragg theory[44] provides a convenient means for comparing propensities of the coil–helix transition based on different models. In this theory, two parameters have been proposed: the nucleation parameter σ reflecting the difficulty level of nucleating a helical domain among coil segments and propagation parameters s measuring the propagating ability at the helix–coil interface. These two parameters are often considered as a standard for the characterization of the coil–helix transition. Since 1970, values of s and σ have been determined experimentally by different groups.[60–64] In recent years, some groups also attempted to determine these two parameters through computer simulations, and have obtained some important insights.[22,65–69] In this paper, the values of σ and s are directly calculated from the dynamic process of helix formation. The calculation method will be discussed in detail in a later paper. The main idea is to obtain the values of σ and s from the equilibrium constant of the conversion between coil and helix segments.

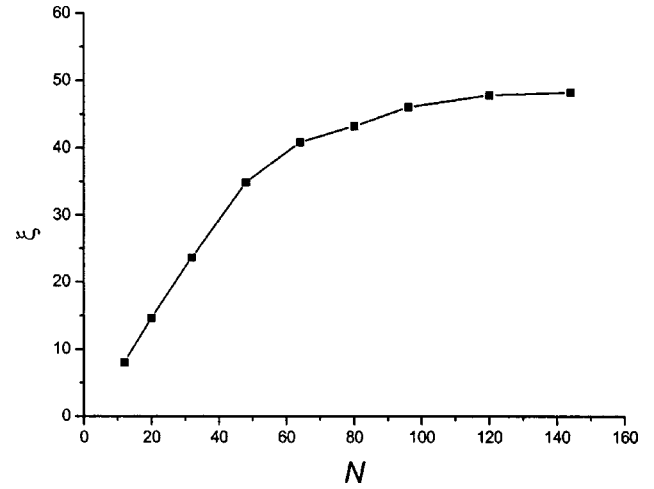


FIG. 9. Averaged correlation length ξ as a function of chain length N. The corresponded reduced inverse temperature was set as $2/T_c^*$. Here, $1/T_c^*$ refers to the transition point for associated chain length determined from the peak position in the curve of reduced specific heat versus reduced inverse temperature.

In the following we use "0" and "1" to denote a coil residue and a helical residue. The ith residue is marked as "1," only if a hydrogen bond is formed between the ith residue and the $(i-4)$th residue. The probability of nucleation can be calculated from

$$p_{000\to010}=\frac{\langle \mathrm{Num}_{000\to010,\mathrm{suc}}\rangle}{\langle \mathrm{Num}_{000\to010,\mathrm{try}}\rangle}. \tag{17}$$

Here, 000→010 denotes the conversion of the middle residue from a coil state to the helical state because of hydrogen bonding. Note that this process is equivalent to the nucleation step originally defined in the Zimm–Bragg model.[44] In Eq. (17), $\mathrm{Num}_{000\to010,\mathrm{try}}$ represents the Monte Carlo time (or try number) during which a residue is attempting to change its state, and $\mathrm{Num}_{000\to010,\mathrm{suc}}$ the times of successful conversion for the tried residue; the average is taken over all $N-4$ residues that are possible to be in the helical state; $p_{000\to010}$ denotes the probability of successful attempts for nucleus formation. In a similar definition, the probability of successful attempts for nucleus annihilation for a pertinent attempt ($p_{010\to000}$) can also be calculated. Thus, we can obtain the value of equilibrium constant $K_{000\leftrightarrow010}$ by

$$K_{000\leftrightarrow010}=\frac{p_{000\to010}}{p_{010\to000}}, \tag{18}$$

which must be equal to the statistical weight σs of the nucleation step in the Zimm–Bragg theory. The values of $p_{000\to010}$, $p_{010\to000}$, and $K_{000\leftrightarrow010}$ at different temperatures are shown in Fig. 10.

The equilibrium constant $K_{100\leftrightarrow110}$ leads to the calculation of the statistical weight s of the propagation step straightforwardly. Thus, we can determine σ and s based upon a dynamic computer simulation of the coil–helix transition. The results are shown in Fig. 11. The nucleation parameter has the same order of magnitude among the simulated temperature range and seems not very sensitive to temperature. In contrast to it, the value of the propagation parameter varies among several orders of magnitude, and s is nearly 1 at the coil–helix transition point. The dependence of σ and s on temperature agree with the Zimm–Bragg theory[44] quite well.

Based upon the calculated σ and s, helix ratio θ can be calculated[22,68] as the result of the Zimm–Bragg theory,[44]

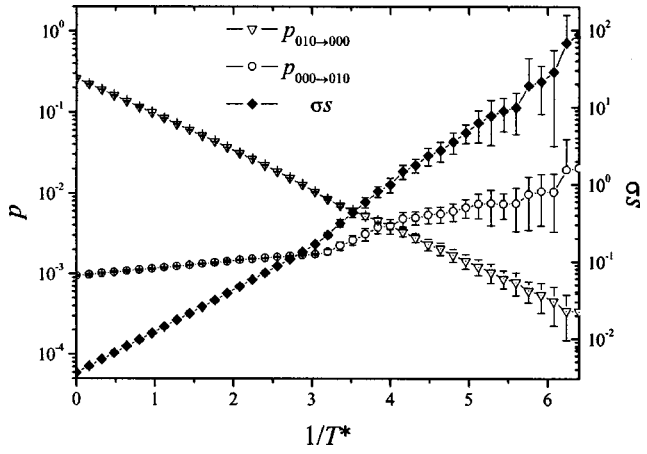


FIG. 10. Conversion probability in the simulation p and calculated statistical weight σs as functions of the reduced inverse temperature $1/T^*$. Open circles represent the probability of forming a helix nucleus in a full coil segment; open triangles represent the probability of dissolving a helix nucleus into a full coil segment; filled diamonds represent the statistical weight σs.

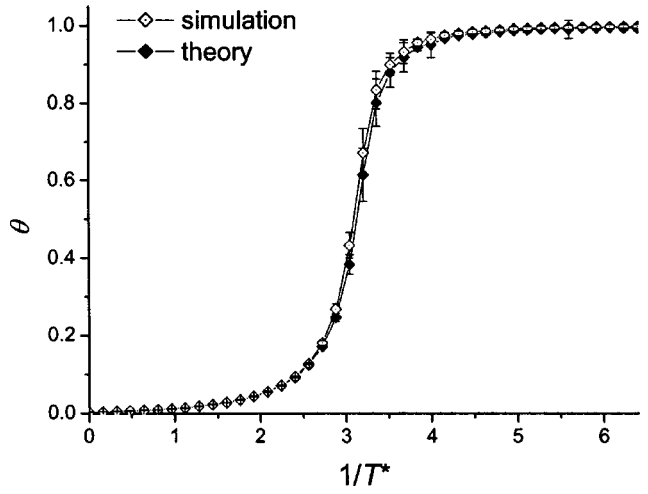


FIG. 12. A comparison of the helix ratio obtained directly from simulations and that from the Zimm–Bragg theory at given reduced inverse temperatures $1/T^*$.

$$\theta = \frac{\langle n \rangle}{N-4} = \frac{1}{2} - \frac{1-s}{2\sqrt{(1-s)^2+4s\sigma}}, \tag{19}$$

where n refers to the whole number of helical residues in a chain. Of course, n can also be directly obtained from simulation. Figure 12 demonstrates that the direct simulation result satisfactorily agree with the theoretical estimate based on Eq. (19) with the ''measured'' nucleation parameter and propagation parameter.

IV. CONCLUDING REMARKS

In this work, we have constructed a new coarse-grained lattice model for protein folding and have reproduced a homopolypeptide's coil–helix transition by dynamic Monte Carlo simulations. Physical properties such as chain size and shape have been examined. An orientational correlation function has been introduced to characterize the periodicity and persistency of the helix state. Propagation and nucleation parameters have also been determined, and by comparison we found that our coarse-grained model is consistent with the Zimm–Bragg theory.

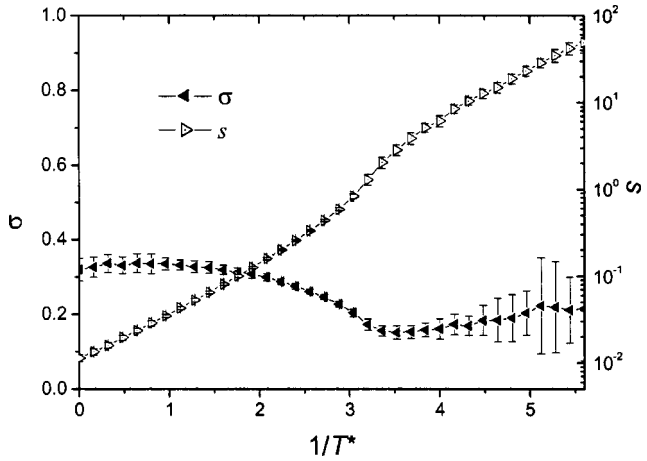


FIG. 11. Nucleation parameter σ and propagation parameter s as functions of the reduced inverse temperature $1/T^*$.

The simulation of helix based on a simplified model is, generally speaking, not new. The main novelty of this paper may be briefly summarized as follows: (1) the new model for reproducing a helical structure. Lattice computer simulations have been performed in the present work while most related works are based upon an off-lattice model.[11,12,20,21] Furthermore, an eight-site lattice model is employed in this paper in contrast to a single-site, or modified lattice model in other work.[24,38] The advantage of the eight-site lattice model over the conventional single-site model is that more bond lengths and bond angles are permitted. It is not difficult to understand that a conventional lattice model is hard to reproduce a helix. The eight-site model may be regarded as a quasi-off-lattice model and meanwhile it keeps the features of lattice simulations such as being easy to incorporate the excluded volume interaction, nearest-neighbor interaction, etc. The interactions such as chirality can be introduced in a very simple way. The model is, albeit not limited to, especially suitable for dynamic Monte Carlo simulation. (2) The new method for characterizing a helix. It is very important to obtain the period and the persistent length for a helix in light of Statistical Physics. The spatial correlation function has been proposed, which might not be limited to our lattice model and will be very useful in various kinds of simulation of helices.

Our simulations agree with the Zimm–Bragg theory to a certain degree. The new model has captured the key physical properties of the coil–helix transition as a function of temperature. The model and characterization methods will be useful, unless fine spatial structures of proteins are expected. Folding kinetics and the protein–protein interaction might also be investigated, since the present lattice model and the associated lattice Monte Carlo simulation approach are very efficient in computation.

ACKNOWLEDGMENTS

We are grateful for the financial support from the National Science Foundation (NSF) of China (No. 29825109, No. 20174006, and the Special Fund for Excellent Originality Research Group), the Award Foundation for Young Teachers from the Ministry of Education, 973 project (No. G1999054306-03), 863 project (No. 2001AA215451), Science and Technology Developing Foundation of Shanghai (No. 02DZ11010, No. 00DJ14001-5). A critical reading of the manuscript by Professor Jeff Z. Y. Chen and helpful comments by Professor Yaoqi Zhou are appreciated.

[1] C. Levinthal, J. Chem. Phys. **65**, 44 (1968).
[2] C. B. Anfinsen, Science **181**, 223 (1973).
[3] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[4] K. A. Dill and H. S. Chan, Nat. Struct. Biol. **4**, 10 (1997).
[5] M. Gross and K. W. Plaxco, Nature (London) **388**, 419 (1997).
[6] N. Gō and H. Taketomi, Proc. Natl. Acad. Sci. U.S.A. **75**, 559 (1978).
[7] N. Gō and H. Taketomi, Int. J. Protein Res. **13**, 447 (1979).
[8] H. Taketomi, Y. Ueda, and N. Gō, Int. J. Pept. Protein Res. **7**, 445 (1975).
[9] Y. Ueda, H. Taketomi, and N. Gō, Biopolymers **17**, 1531 (1978).
[10] A. V. Smith and C. K. Hall, J. Chem. Phys. **113**, 9331 (2000).
[11] A. V. Smith and C. K. Hall, Proteins **44**, 344 (2001).
[12] A. V. Smith and C. K. Hall, Proteins **44**, 376 (2001).
[13] Y. Zhou and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **94**, 14429 (1997).
[14] Y. Zhou and M. Karplus, Nature (London) **401**, 400 (1999).
[15] Y. Zhou and M. Karplus, J. Mol. Spectrosc. **293**, 917 (1999).
[16] Y. Zhou, M. Karplus, J. M. Wichert, and C. K. Hall, J. Chem. Phys. **107**, 10691 (1997).
[17] Y. Zhou and A. Linhananta, J. Phys. Chem. B **106**, 1481 (2002).
[18] H. Jang, C. K. Hall, and Y. Zhou, Biophys. J. **82**, 646 (2002).
[19] A. Linhananta and Y. Zhou, J. Chem. Phys. **117**, 8983 (2002).
[20] J. Kemp and Z. Y. Chen, Phys. Rev. Lett. **81**, 3880 (1998).
[21] J. Kemp and J. Z. Y. Chen, Europhys. Lett. **59**, 721 (2002).
[22] U. H. E. Hansmann and Y. Okamoto, J. Chem. Phys. **110**, 1267 (1999).
[23] J. Shimada, E. L. Kussell, and E. I. Shakhnovich, J. Mol. Spectrosc. **308**, 79 (2001).
[24] K. F. Lau and K. A. Dill, Macromolecules **22**, 3986 (1989).
[25] A. Säli, E. I. Shakhnovich, and M. Karplus, J. Mol. Spectrosc. **235**, 1614 (1994).
[26] D. K. Klimov and D. Thirumalai, J. Chem. Phys. **109**, 4119 (1998).
[27] E. I. Shakhnovich, Folding Des. **1**, R50 (1996).
[28] D. Baker, Nature (London) **405**, 39 (2000).
[29] Y. Zhou, C. K. Hall, and M. Karplus, Phys. Rev. Lett. **77**, 2822 (1996).
[30] M. Muthukumar, J. Chem. Phys. **104**, 691 (1995).
[31] M. Muthukumar, C. K. Ober, and E. L. Thomas, Science **277**, 1225 (1997).
[32] T. B. Liverpool, R. Golestanian, and K. Kremer, Phys. Rev. Lett. **80**, 405 (1998).
[33] V. A. Ivanov, W. Paul, and K. Binder, J. Chem. Phys. **109**, 5659 (1998).
[34] D. Hu, J. Yu, K. Wong, B. Bagchi, P. J. Rossky, and P. F. Barbara, Nature (London) **405**, 1030 (2000).
[35] L. Zhang and D. Zhao, Macromol. Theory Simul. **10**, 518 (2001).
[36] A. V. Chertovich, V. A. Ivanov, B. G. Zavin, and A. R. Khokhlov, Macromol. Theory Simul. **11**, 751 (2002).
[37] Y. Okamoto and U. H. E. Hansmann, J. Phys. Chem. **99**, 11276 (1995).
[38] J. Skolnick and A. Kolinski, Science **250**, 1121 (1990).
[39] A. Sikorski, A. Kolinski, and J. Skolnick, Biophys. J. **75**, 92 (1998).
[40] B. Ilkowski, J. Skolnick, and A. Kolinski, Macromol. Theory Simul. **9**, 523 (2000).
[41] C. M. Chen, Phys. Rev. E **63**, 010901 (2000).
[42] A. Chakrabartty and R. L. Baldwin, Adv. Protein Chem. **46**, 141 (1995).
[43] T. Kiefhaber, A. M. Labhardt, and R. L. Baldwin, Nature (London) **375**, 513 (1995).
[44] B. H. Zimm and L. K. Bragg, J. Chem. Phys. **31**, 526 (1959).
[45] R. Lifson and A. Roig, J. Chem. Phys. **34**, 1963 (1961).
[46] M. Fixman and D. Zeroka, J. Chem. Phys. **48**, 5223 (1968).
[47] H. Qian and J. A. Schellman, J. Phys. Chem. **96**, 3987 (1992).
[48] T. Wieprecht and J. Seelig, Biophys. J. **78**, 959 (2000).
[49] M. N. Tamashiro and P. Pincus, Phys. Rev. E **63**, 021909 (2000).
[50] T. L. Blundell and L. N. Johnson, *Protein Crystallography* (Academic, New York, 1976), Vol. 24.
[51] I. Carmesin and K. Kremer, Macromolecules **21**, 2819 (1988).
[52] H. P. Deutsch and K. Binder, J. Chem. Phys. **94**, 2294 (1991).
[53] G. Xu, J. Ding, and Y. Yang, J. Chem. Phys. **107**, 4070 (1997).
[54] G. Xu, J. Ding, and Y. Yang, Rheol. Acta **40**, 60 (2001).
[55] K. A. Dill, Biochemistry **29**, 7133 (1990).
[56] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[57] K. Binder and D. W. Heermann, *Monte Carlo Simulation in Statistical Physics* (Springer-Verlag, Heidelberg, 1988).
[58] U. H. E. Hansmann and Y. Okamoto, J. Comput. Chem. **14**, 1333 (1993).
[59] Y. Yang and T. Yu, Sci. Sin., Ser. A **29**, 1096 (1986).
[60] P. H. von Dreele, D. Poland, and H. A. Scheraga, Macromolecules **4**, 396 (1971).
[61] P. H. von Dreele, N. Lotan, V. S. Ananthanarayanan, R. H. Andreatta, D. Poland, and H. A. Scheraga, Macromolecules **4**, 408 (1971).
[62] J. Wōjcik, K. H. Altmann, and H. A. Sheraga, Biopolymers **30**, 121 (1990).
[63] J. M. Scholtz, H. Qian, E. J. York, J. M. Stewart, and R. L. Baldwin, Biopolymers **31**, 1463 (1991).
[64] A. Chakrabartty, T. Kortemme, and R. L. Baldwin, Protein Sci. **3**, 843 (1994).
[65] V. Daggett, P. A. Kollman, and I. D. Kuntz, Biopolymers **31**, 1115 (1991).
[66] A. S. Yang and B. Honig, J. Mol. Bio. **252**, 351 (1995).
[67] L. Wang, T. O'Connell, A. Tropsha, and J. Hermans, Proc. Natl. Acad. Sci. U.S.A. **92**, 10924 (1995).
[68] Y. Okamoto and U. H. E. Hansmann, J. Phys. Chem. **99**, 11276 (1995).
[69] A. V. Smith and C. K. Hall, Proteins: Struct., Funct., Genet. **44**, 344 (2001).