Folding behaviors of lattice model proteins with three kinds of contact potentials

Meng Qin, Jun Wang, Yi Tang, and Wei Wang*
National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, Nanjing 210093, China
(Received 30 January 2003; published 18 June 2003)

The interaction potentials between the amino acids are very important in the study of protein folding and design. In this work, the folding behaviors of lattice model protein chains are studied using three kinds of contact potentials between the beads. For these three cases, a number of sequences are designed using the *Z*-score method, and then their folding behaviors are obtained via Monte Carlo simulations for different sizes of the chains. It is found that the proper weakening of hydrophobicity may speed up the folding and the elimination of the mixing interaction terms may deteriorate the foldability. The different features of the foldability are discussed by comparing the characteristics of the energy landscapes of these model chains. The formations of various contacts are also analyzed, which provide us with some microscopic information on the model systems and interaction potentials.

DOI: 10.1103/PhysRevE.67.061905 PACS number(s): 87.15.Cc, 87.15.Aa, 87.14.Ee

I. INTRODUCTION

Natural proteins are composed of 20 kinds of α-amino acids, and their folding to specific native structures is believed to be encoded by their nonrandom arrangements of the amino acids. Thus, the folding of proteins becomes a physical problem when all the interactions in the protein systems could be worked out. In general, a protein contains about thousands of atoms and interacts with a huge number of solvent molecules. Even by using the fastest computers, it is still quite difficult to simulate the folding processes for a reasonable size of amino acid sequence when all the interactions between the atoms are included. Therefore, protein folding is still a basic, important, and not completely understood issue in molecular biology and biophysics [1–3]. Previously, a number of simplified models are used for studying the folding process [4–8]. Among them, some models simplified the amino acids as beads, and the interactions between the beads were set based on the statistics of the residue pairs in crystal structures of the proteins [9–14]. These interactions are assumed to exist when the distances between the beads are less than a certain cutoff distance. The generally used statistical potential is the Miyazawa and Jernigan (MJ) potential, i.e., the so-called MJ potential or matrix [9,10]. Later, Thirumalai and co-worker modified the MJ potentials by setting a reference state from the solvent molecules in the MJ potential to threonine [15,16]. Li and his co-worker found some regularities that the MJ matrix can be reconstructed with its first two principal component vectors by decomposing the MJ matrix [11]. [These three kinds of potentials are referred as to the MJ, the modified MJ (MMJ), and the reconstructed MJ (RMJ) potentials in this work.] Therefore, an interesting question arises: which kind of potentials can be responsible for good folding features, namely, the fast kinetic accessibility and the high thermodynamic stability?

In this paper, folding behavior both in the thermodynamic and kinetic aspects for the model proteins with the above-mentioned three kinds of potentials are studied under lattice protocol. For a series of sequences designed by minimizing their *Z* scores, the simulations related to the folding and equilibrium are performed. The folding speed, characterized by the mean first passage time (MFPT) to the ground state, versus the reciprocal of the temperature is investigated in detail. We find that the sequences designed with the MMJ potential fold much faster than those with the MJ and with the RMJ potential. The collapse transition and the folding transition, and the related factor σ of foldability are evaluated. The sequences designed with the MMJ potentials show a better two-state transition behavior with the collapse transition and the folding transition almost concurrently at the same temperatures. On the other hand, the sequences designed with the MJ and with the RMJ potentials show much worse characteristics. The free-energy landscapes with respect to the coordinate *Q*, the similarity to the ground state, are investigated to elucidate the microscopic mechanism of the different folding behaviors. The formation of the hydrophobic contacts during the folding is also discussed.

The arrangement of this paper is as follows. In Sec. II, the model and the method employed in our study are described. In Sec. III, the results and some related discussions are presented. Finally, in Sec. IV, some conclusive remarks are given.

II. MODEL AND METHOD

A. Three potential matrices

By simplifying the naturally occurring 20 kinds of amino acids in proteins as 20 kinds of beads, the interactions between the amino acids are reduced to a 20×20 matrix, i.e., a contact potential or a contact matrix. Our study in this work is based on the following three kinds of contact matrices.

Miyazawa and Jernigan obtained effective inter-residue potentials using the quasichemical approximation [9]. That is, they assumed that the residues are in equilibrium with the solvents. Thus, the binding between residues i and j undergoes two steps. First, there is a desolvation that gives the reversible work required to expel a solvent molecule in contact with residue i (or j). Then, there is a mixing between residues i and j. The elements $M_{i,j}$ of the MJ contact matrix

*Author to whom correcpondence should be addressed. Electronic address: wangwei@nju.edu.cn

are defined as the excess energy due to the contact of residues i and j, i.e., the reversible work required to bring residues i and j into contact by

$$M_{i,j}=E_{i,j}+E_{0,0}-E_{i,0}-E_{j,0}, \tag{1}$$

where the subscript 0 refers to the solvent molecule. In this work, the data in Table III of Ref. [9] are employed in all the studies.

It was noted by Godzik *et al.* that the reference states have large effects in the calculation of the matrix [17]. As pointed out by Thirumalai *et al.*, the MJ scheme uses the random mixing approximation to calculate the element $M_{i,j}$ and it is difficult to estimate $E_{0,0}$ and $E_{i,0}$, since one should obtain both the average numbers of solvent-solvent contacts and residue-solvent contacts. Thirumalai and co-workers chose a different reference state within the MJ scheme. They used the residue threonine to replace the solvent molecule in Eq. (1), then they obtained the matrix of the modified MJ matrix as

$$X_{i,j}=M_{i,j}+M_{t,t}-M_{i,t}-M_{j,t}, \tag{2}$$

where t represents the residue threonine. Here, they assumed that the residue threonine is neutral in water environment, which results in the relations $X_{i,t}=X_{t,i}=0$, for $i=0,1,2,\ldots,20$ and the index 0 is the solvent molecule.

Beyond the MJ matrix, Li *et al.* found some simplicity in the MJ matrix, that is, the MJ matrix can be reconstructed from its first two related principal components by using the eigenvalue decomposition method [11]. Using this method, they rewrote $M_{i,j}$ is the following simple form:

$$M_{i,j}=C_0+C_1(q_i+q_j)+C_2q_iq_j, \tag{3}$$

where C_0, C_1, and C_2 are three constants, and q_i, with $i=1,2,\ldots,20$, are the components of the first eigenvector, which are correlated with the hydrophobicity of 20 kinds of amino acids. The reconstructed MJ matrix is similar to the original MJ matrix and the correlation coefficient between them is about 0.981.

The MMJ matrix is also analyzed by the decomposition method. It is found that there is no dominant eigenvalue for the MMJ matrix. The largest five eigenvalues are 5.8585, 2.4922, 1.9647, 1.8220, and 1.2028. The corresponding components of these eigenvectors do not have apparent correlations. This is quite different from the MJ matrix and the RMJ matrix. In other words, the MMJ matrix is not dominated by the hydrophobicity, which is the most important factor in the MJ matrix or the RMJ matrix [11]. In our study, the RMJ matrix and the MMJ matrix can be considered as two kinds of modifications starting from the widely used MJ matrix. The RMJ matrix keeps the only dominant eigencomponent, while the MMJ matrix weakens the effect of the dominant hydrophobicity to some degree. The comparison of the folding behaviors among the MJ matrix, the RMJ matrix, and the MMJ matrix may provide us with some insights into the role of various energetic ingredients and which would be essential for further modeling of folding.

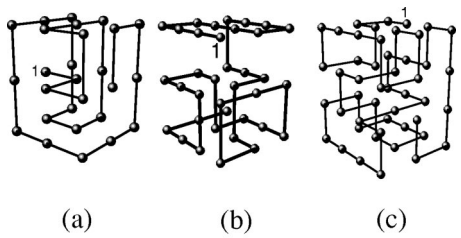

FIG. 1. Three native structures for the lattice model chains with the chain length (a) $L=27$, (b) 36, and (c) 48. The number of the native contacts are $Q_N=28$, 40, and 57, respectively.

B. Cubic lattice model

Lattice protein models have provided many insights for proteins and their folding despite some coarse approximations [4,6,7,13,18]. The most popular three-dimensional lattice proteins are modeled as self-avoided random walks on a cubic lattice with beads on vertices. Two beads, which are spatially neighboring with one lattice spacing, but not successive along the chain, can form a contact. The energy of the system is considered as the total contribution of the contacts,

$$E_s^{\Gamma}=\sum_{i\geqslant j+3} U_{i,j}\delta(r_{ij}-a), \tag{4}$$

where Γ indicates the conformation, $U_{i,j}$ is the contact potential between residues i and j, and $\delta(r_{ij}-a)$ characterizes the geometrical requirement of the contact between residues i and j with $\delta(0)=1$ for a contact or 0 otherwise. Here a is the lattice spacing. Three structures shown in Fig. 1 are chosen as our targets of sequence design. Figure 1(a) is a structure having the highest designability in Ref. [19], Fig. 1(b) is the native structure studied in Ref. [20], and Fig. 1(c) from Ref. [6]. They are often used for the studies of lattice protein modeling.

We perform the Monte Carlo (MC) folding simulations following the method described in Refs. [4,18]. Each simulation is started from a randomly coiled conformation. The arrival of the folded state in kinetics is assumed as the first visit of the native structure (as shown in Fig. 1). The parameter Q, i.e., the number of the native contacts in the current conformation, measures the structural similarity with the native state and is used to characterize the proceeding of the folding [21].

C. Thermodynamic characterization

Protein molecules usually experience two kinds of transitions during their folding [22,23]. One is the collapse transition, during which the chain shows a large change in its shape from a coil state to a compact form. The nonspecific attraction between residues proposes a large change in energy, which results in a peak in the heat capacity. The temperature of the peak is regularly marked as T_θ. The second transition indicates the establishment of the native structure. The characteristic temperature, marked as T_f, is generally described by the large change in the structural feature. Here we adopt the peak temperature of the fluctuation $\delta\chi$ of a

structural overlap function χ as defined in Ref. [24]. In this study, we use the histogram method to calculate the thermodynamic average for various physical quantities [4]. The heat capacity is defined as

$$C_v(T)=(\langle E^2\rangle-\langle E\rangle^2)/k_BT^2, \tag{5}$$

where E is the energy of the model chain and k_B is the Boltzmann constant (in this study, $k_B=1$). The overlap function χ is defined based on the data of the interbead distances,

$$\chi=1-\frac{1}{N^2-3N+2}\sum_{i\neq j,j\pm1}\delta(r_{i,j}-r_{ij}^N), \tag{6}$$

where r_{ij} and r_{ij}^N are the distances between the ith and the jth residues in a relevant conformation and that in the native structure, respectively. This function is physically similar to the order parameter Q, but includes some more details of the structural similarity. The quantity $\delta\chi$ is defined as the fluctuation of the function χ,

$$\delta\chi=\sqrt{\langle\chi^2\rangle-\langle\chi\rangle^2}. \tag{7}$$

A factor $\sigma=|T_\theta-T_f|/T_\theta$ is used to describe the foldability of the protein model chains [24,25]. The smaller the σ is, i.e., the more adjacent the collapse transition is from the folding transition, the better the foldability of the model chain is. Especially for the case that $\sigma\approx0$, i.e., the folding transition and the collapse transition occur at almost the same temperature, the model chain collapses and folds simultaneously. Consequently, the model chain may suffer less from the competition of other compact low-energy states. Thus, the chain reaches its native state quickly and the MFPT is quite short. On the other hand, when $\sigma\sim1$, the model chain may struggle in escaping from a lot of local minima, and the folding time is rather long.

D. *Z*-score-based sequence design method

Among the methods for the sequence design problem [26–30], the *Z*-score method [31] is popular and successful [32]. In this study, we use the *Z*-score method to design the sequences. The Z score is defined as

$$Z_{score}=\frac{E_{target}-E_{average}}{\sqrt{\langle E^2\rangle-\langle E\rangle^2}}, \tag{8}$$

where E_{target} is the energy in the target state and $E_{average}$ is the average energy of the unfolded states. Here $E_{average}$ is estimated from the contact averages as $E_{average}=N\langle e\rangle$, in which $\langle e\rangle$ is the average energy of all possible contacts and N is the number of contacts in the native state. By minimizing the Z scores, a series of model chains with a certain preassumed composition is designed for further simulational studies [26]. Regularly, about 100 sequences are selected for one of the three contact matrices and chain length $L=27$, 36, or 48. To establish a suitable comparison between different cases, the ratio of the hydrophobic residues to the polar ones is fixed for all designs. The detailed compositions of residues are chosen randomly to obtain a diverse distribution of sequences. In this work, the hydrophobic residues (denoted as H) include the ones such as *L*, *I*, *V*, *M*, *C*, *F*, *Y*, and *W*, and the polar type (denoted as P) corresponds to the rest residues as *R*, *S*, *T*, *Q*, *H*, *D*, *E*, *K*, *G*, *A*, *N*, and *P* [12]. The ratios of H:P used in this work are 13:14, 10:26, and 16:32 for the cases of $L=27$, 36, and 48, respectively. Other ratios of residues, H:P, and the corresponding designed sequences are also investigated. The results are similar and are not presented in this paper.

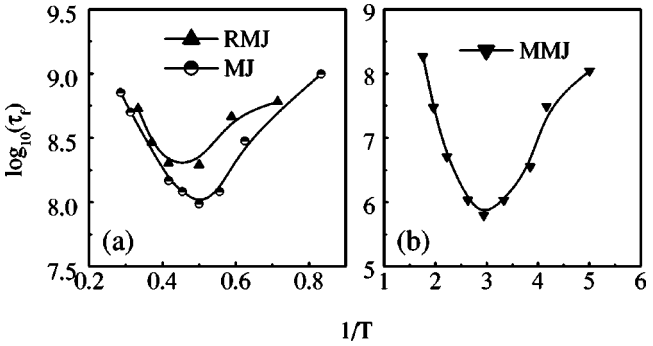

FIG. 2. The logarithm of folding time $[\log_{10}(\tau_f)]$ versus the reciprocal of the folding temperature $(1/T)$ with the chain length $L=27$. (a) For the sequences designed with the MJ and the RMJ matrices, (b) for the MMJ matrix.

III. RESULTS AND DISCUSSIONS

The kinetics of the model chains are characterized with the MFPT(τ_f) at various temperatures. At each temperature, about 100 runs for the folding starting from different initial conformations are carried out on an average. Figure 2 shows the logarithm of the folding time, $\log_{10}(\tau_f)$, of a model sequence versus the inversion of the temperature, $1/T$, for the three kinds of contact matrices, with $L=27$. In Fig. 2, the fast folding temperatures for the three cases are indicated. The fastest folding has the time of 6.25×10^5 MC steps at $T_{fast}\approx0.34$ for the cases using the MMJ matrix, and is 1×10^8 MC steps and 2×10^8 MC steps at $T_{fast}\approx2.0$ for the cases with the MJ matrix and with the RMJ potential, respectively. As a result, the folding of the sequences designed with the MMJ matrix is about 100 times faster than those designed with the MJ matrix or those with the RMJ matrix. The sequences designed with the RMJ potentials even fold more slowly than those with the MJ matrix. It is noted that the temperatures corresponding to the fastest folding are about the same for the MJ and the RMJ cases, which suggests that the sequences of these two matrices share some common features in their landscapes.

Some thermodynamic factors, such as the values of the temperatures T_θ and T_f and σ, relating to the foldability are investigated and the results related to three sequences which are randomly picked from the set of designed sequences are listed in Table I. The behaviors of other design sequences are similar (the results are not listed here). For the sequences designed with the MMJ matrix, the average value of σ is about 0.01. The repulsive elements of the MMJ matrix can

TABLE I. The folding chatacteristics for the three kinds of contact matrices for the chain length $L=27$. T_θ is the collapse transition temperature, T_f is the folding transition temperature, and σ_T is the foldability factor. T_{fast} is the fastest folding temperature and $\tau_f(T_{fast})$ is the folding time at T_{fast}.

Chain length: 27		T_θ	T_f	σ_t	T_m	τ_f (MC)
MJ	S_1	1.17	1.13	0.034	2.0	0.97×10^8
matrix	S_2	1.14	1.07	0.062	2.3	1.09×10^8
	S_3	1.36	1.0	0.26	2.2	0.99×10^8
RMJ	S_1	1.51	0.59	0.61	2.0	1.96×10^8
matrix	S_2	1.76	0.59	0.66	2.4	1.77×10^8
	S_3	1.46	0.70	0.60	2.2	2.04×10^8
MMJ	S_1	0.33	0.33	0	0.34	1.07×10^6
matrix	S_2	0.38	0.38	0	0.37	6.61×10^5
	S_3	0.30	0.29	0.033	0.32	1.59×10^6

decelerate the collapse of the model protein chains. The collapse transition is also its folding transition. For the MJ cases, the values of σ are in an intermediate range, and for the sequences designed with the RMJ matrix, the average σ over the three sequences is near 0.62. Obviously, the sequences designed with the MMJ matrix show a better foldability with smaller values of σ than those with the other two matrices. This coincides with our kinetic results above.

For longer chains with $L=36$ or 48, the results are similar (as shown in Fig. 3 and Table II). Studying the folding rate and some related thermodynamic factors as criteria for the folding kinetics lead to the same conclusion as for $L=27$.

Thus, it can be concluded that the sequences designed with the MMJ matrix show the fastest folding and the best foldability than those designed with the other two matrices, and the folding behavior of the sequences designed with the MJ matrix are more or less the same as those with the RMJ matrix. From this aspect, the MMJ matrix of the inter-residue interactions can provide model chains a good foldability and kinetic accessibility.

The stability of the model proteins, as characterized by the occurring probability P_f of the native state, is also studied. Figure 4(a) shows P_f versus the temperature T for $L=27$. We can see that as the temperature T decreases, the

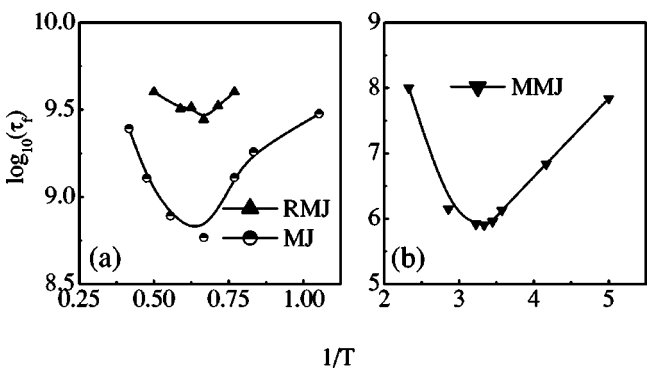

FIG. 3. The same as Fig. 2, but the chain length $L=36$.

TABLE II. The same as Table I, but the model chain length $L=36$.

Chain length: 36		T_θ	T_f	σ_t	T_m	τ_f (MC)
MJ	S_1	1.01	0.99	0.020	1.5	5.84×10^8
matrix	S_2	1.09	1.08	0.01	1.7	5.83×10^8
	S_3	1.09	1.09	0.00	1.6	5.91×10^8
RMJ	S_1	0.81	0.71	0.12	1.5	2.78×10^9
matrix	S_2	0.75	0.67	0.11	1.6	2.65×10^9
	S_3	0.72	0.62	0.14	1.5	2.89×10^9
MMJ	S_1	0.32	0.32	0	0.30	0.81×10^6
matrix	S_2	0.33	0.33	0	0.32	1.10×10^6
	S_3	0.33	0.33	0	0.30	1.61×10^6

values of P_f show a transition for the three matrices. However, the transition for the MMJ matrix is the sharpest transition than those for the other two matrices, which suggests that the transition for the case with the MMJ matrix is more cooperative than the cases with the other two matrices. In addition, the transition temperatures are different, and the lowest transition temperature is for the MMJ matrix among the three matrices. This implies that the unfolded state ensemble is more similar to the native structure for the MMJ case than those for the MJ case and for the RMJ case. Figure 4(a) also shows that the transition of the sequences designed with the MJ matrix is sharper than that of the sequences designed with the RMJ matrix. For the case of $L=36$, similar results are obtained as shown in Fig. 4(b). However, the transition of sequences designed with the RMJ matrix is not as sharp as those of the sequences designed with the MJ and the MMJ matrices.

To gain a better understanding on the effects of using different interactions, the free-energy landscapes are investigated in detail. The formula of the free energy is

$$F(Q,T)=U(Q,T)-TS(Q,T). \tag{9}$$

Here, S means the microcanonical entropy with $S(Q,T)=kT(E_{nat}-E_Q)\ln[h(Q)/h(N)]$, in which $h(Q)$ is the histogram of the conformations with Q ($\leqslant N$) in our simulations

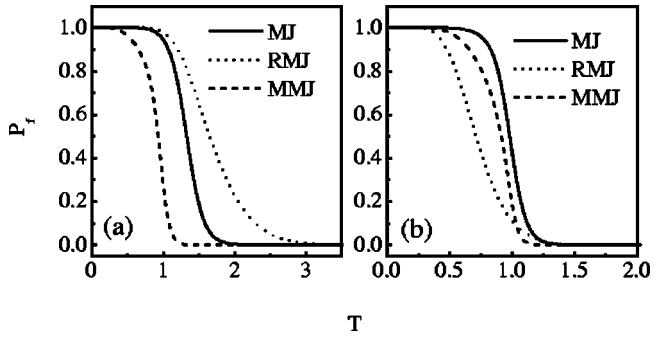

FIG. 4. The occurring probability of the native states P_f versus the folding temperature T: (a) for the chain length $L=27$, (b) for the chain length $L=36$.

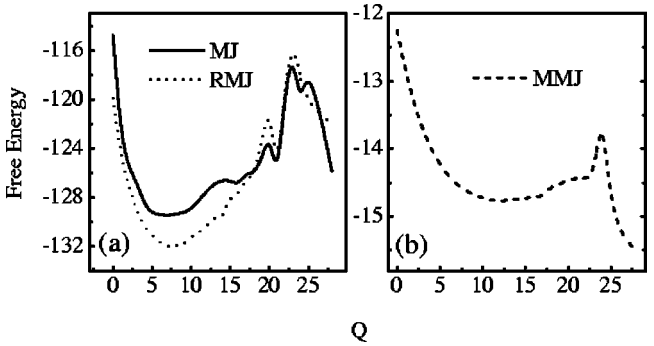

FIG. 5. The free energy versus the structural similarity Q near the folding temperatures for the chain length $L=27$: (a) for the MMJ and MJ matrices, (b) for the MMJ matrix.

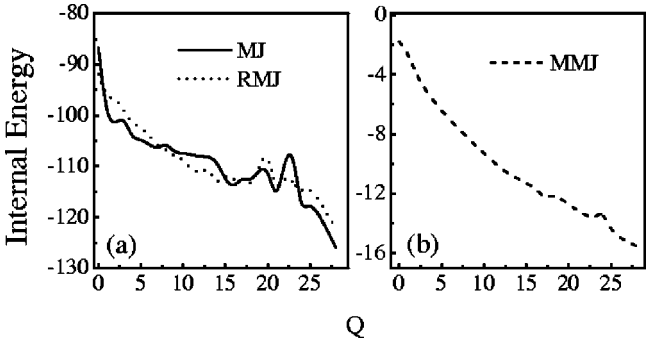

FIG. 7. The internal energy U versus the structural similarity Q near the folding temperature for the chain length $L=27$: (a) for the MJ and RMJ matrices, (b) for the MMJ matrix.

and $h(N)$ corresponds to the histogram of the native state. Figure 5 shows the typical profiles of the free-energy landscape with Q as the coordinate. The corresponding simulation temperatures are around the temperatures with fastest folding rates. In Fig. 5(a), the global minimum of free energy for the sequences designed with the MJ matrix does not locate at $Q=28$, which corresponds to the native state of the chains with $L=27$. The profile is rugged with many zigzags. Considering the regions with median Q from $Q=9$ to $Q=22$, there are some high barriers that increase the difficulty of diffusion on the landscape, thus affect the speed to search the folded structure in the unfolded state ensemble. For the cases with the RMJ matrix, the value of free energy is not the minimum, either. Especially, the peak in the transition state of the sequences with the RMJ potential is even larger than that of the sequences designed with the MJ matrix. This may contribute to the slower folding of sequences with the RMJ matrix.

However, the situation is quite different for the sequences designed with the MMJ matrix as shown in Fig. 5(b). The energy profile is smooth, and the barrier between the unfolded states and the native state is around $Q=24$ and the peak is also low. Along with this kind of landscapes, the folding of the model chains is very fast as expected, comparing with the folding of the sequences designed with the other two matrices.

For the cases of longer chain size of $L=36$ and 48, the profiles of free energy, energy, and entropy are obtained (as

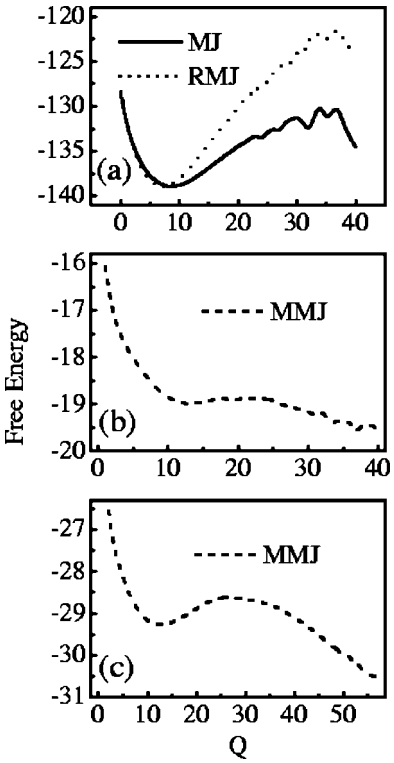

FIG. 6. The free energy versus the structural similarity Q near the folding temperatures for the chain length $L=36$ (a) and (b), $L=48$ (c).

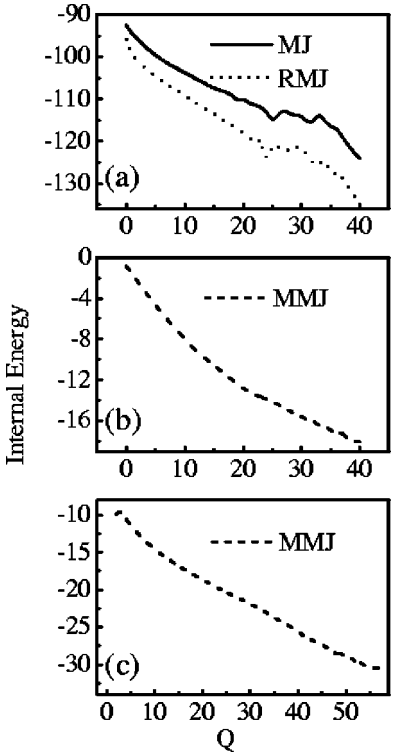

FIG. 8. The internal energy U versus the structural similarity Q near the folding temperature for the chain length $L=36$ (a) and (b), $L=48$ (c).

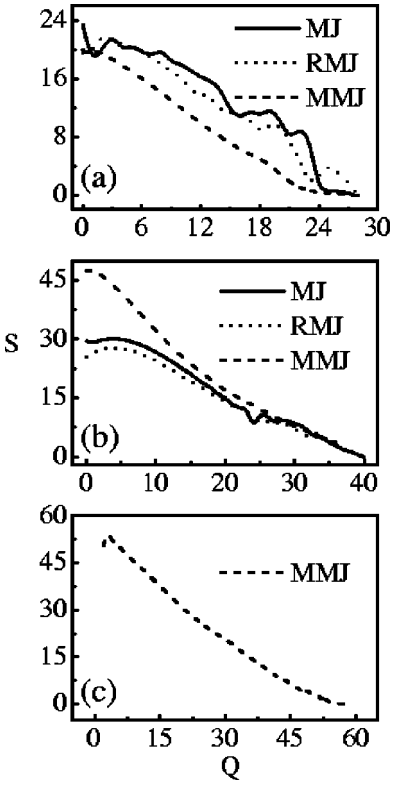

FIG. 9. The conformation entropy S versus the structural similarity Q at the folding temperature for chain length $L=27$ (a), $L=36$ (b), and $L=48$ (c).

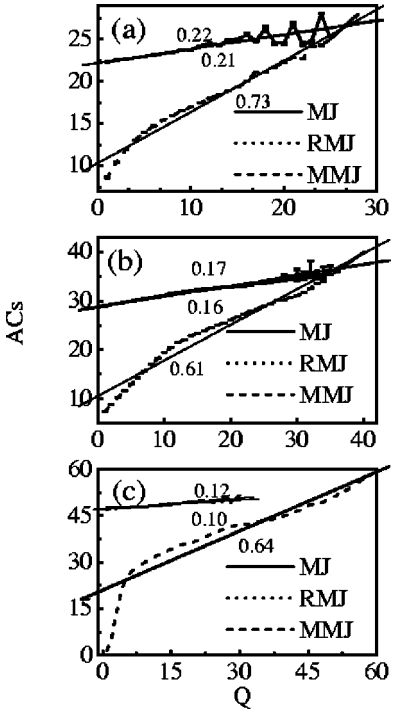

FIG. 10. The total number of contacts ACs versus the structural similarity Q for sequences at the folding temperature and the chain length $L=27$ (a), $L=36$ (b), and $L=48$ (c). The slopes are also marked, respectively. Note that in (c) for the chains with the MJ and the RMJ matrix, the curves are stopped at $Q\approx 30$ due to the limitation of the simulation time 1×10^{10}. The same reason for Fig. 11(c).

shown in Figs. 6–9). These results lead to a similar conclusion on the folding properties with various potentials as mentioned above for the three kinds of potentials.

To understand the origin of the differences among the three matrices, the processes of the formation of various contacts may provide some microscopic information. In this work, the contacts are generally classified into two categories, namely, native contacts and non-native contacts, according to their appearance in the native structure. Describing the model systems using the numbers of native contacts (Q) and of the non-native bonds (Q_{NN}) broadens our sight on the folding processes. Especially for cases with multiple transitions, the folding behaviors in the enlarged space would be more helpful. Figure 10 shows the total number of contacts C of the current conformations in the folding process versus that of the native contacts Q for the three contact matrices and different chain sizes. In Fig. 10(a), the data for the sequences designed with both the MJ and the RMJ matrices are presented. The total number of the contacts C reaches a high value of about 22 ($\approx$78% of the maximum of the number of contacts) within a rather short time. That is, the strongly attractive elements (HH contacts) of the MJ and RMJ potentials make the conformations basically compact after their initiative collapse. However, the number of the native contacts Q is quite small, even less than 5. It takes a very long time for the model chain to adjust its conformations to the native states by forming the native contacts gradually. The contacts of the current conformations follow a linear behavior with the increase of Q. Near the native conformation, there is a large fluctuation, since many non-native bonds are necessary to be broken to form some certain native contacts. This implies that the larger the fluctuation is, the more difficulty the folding may experience. This also indicates the existence of some energetic barriers in the energy landscape. The situation is quite different for sequences designed with the MMJ matrix. The total number of the contacts ACs varies linearly with the number of native contacts Q after some short initiations. The fluctuation near the native conformation is rather small as compared to the MJ and RMJ cases, implying that the energetic barriers are highly suppressed using the MMJ matrix, thus the folding is much faster since the chain need not compete with compact low-energy states. The slopes of ACs versus Q for three cases are marked in Fig. 10. The slope for the case of the MMJ matrix is obviously larger than the other two cases, indicating that the formation of the native contacts basically follows the formation of the total contacts. Long chains [as shown in Figs. 10(b) and 10(c)] show the similar behavior. For the chain size $L=48$, the points for both the MJ and the RMJ cases end around $Q=34$ due to the maximal running time of 5×10^9 MC in our simulations. Conclusively, the formation processes of various contacts reflect the features of the landscapes. The funnel-like landscape may make the folding direct towords the native state, and the ruggedness of the landscape introduces the fluctuation and provides some competitions in the folding processes. In the present cases, the sequences with the MMJ matrix generally have the landscapes more smooth

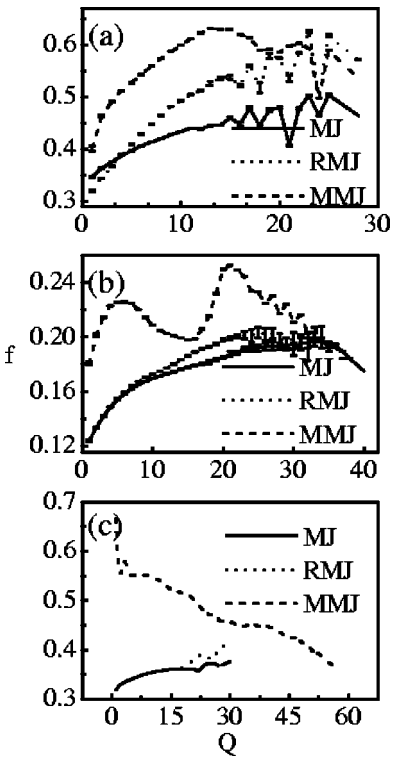

FIG. 11. The ratio (f) of the HH contacts to the total number of contact C versus the structural similarity Q for chains at the folding temperature for the chain length $L=27$ (a), $L=36$ (b), and $L=48$ (c).

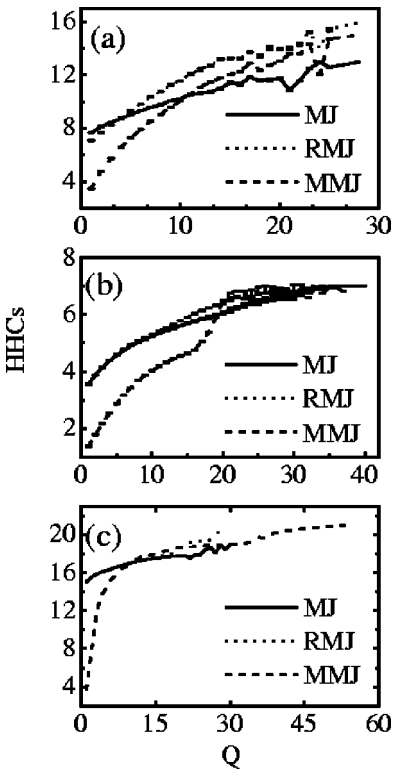

FIG. 12. The number of the HH contacts $HHCs$ versus the structural similarity Q for chains at the folding temperature for the chain length $L=27$ (a), $L=36$ (b), and $L=48$ (c).

and more funnel-like than those of the sequences with the MJ matrix, or with the RMJ one.

Actually, these three matrices act differently from each other in their hydrophobicity. An investigation on the formation of clusters of hydrophobic residues may illustrate the differences between them. Roughly, the contacts can be divided into three kinds: (1) HH contact (between hydrophobic residues), (2) PP contact (between hydrophilic residues), and (3) HP contact (between a hydrophobic residue and a hydrophilic one). The HH contacts, which regularly have larger energetic contribution, may be more important. Here we mainly concentrate on the formation of the HH contacts. We define f as the ratio of the HH contacts in all the contacts in a certain conformation. We make some simulations near the fastest folding temperatures to get a good distribution of conformations. Figure 11 shows the relationship of f versus Q for the three kinds of contact matrices and three system sizes $L=27$, 36, and 48, respectively. For all sequences, the values f for the MMJ case are much larger than those for the MJ and the RMJ cases in almost all the folding processes. It means that the HH contacts take a more important role in the corresponding folding processes. There is one or two oscillations of the value of f before the final saturation of f where the non-native HH contacts are all broken and all the native HH contacts are formed. For the $L=48$ sequences with the MMJ potential, the ratio f decreases during the full folding, but the number of HH contacts increase with the time (see Fig. 12). It indicates that the effect of the HH contacts in the large systems becomes weak due to the increase of entropic aspect, and the HH contacts may contribute more for the initial stage of the folding processes. Meanwhile, for the other two cases, the arrangement of the correct HH contacts seems to be the rate-limiting steps of the folding.

Looking from the separation of the hydrophobic and polar residues, the decreasing tendency of f for the case of the MMJ matrix suggests that the phase separation of the hydrophobic residues and the polar ones are established at the beginning, and the mixing effect takes more important role in the whole folding processes. Meanwhile, for the MJ matrix and the RMJ matrix, the increase of f implies that the phase separation is gradually finished during the whole folding processes. Especially, for the RMJ matrix, the factor f goes higher than that of the MJ cases. These phenomena are consistent with the features of these matrices. The RMJ matrix is created by climinating the mixing component of the MJ matrix, while the MMJ matrix conserves mainly the mixing term of the MJ matrix. From this information, it may be concluded that the mixing term is essential in building a proper potential for folding modeling.

IV. CONCLUSION

In this work, based on the lattice modeling, we study the kinetic and the thermodynamic characteristics of sequences designed with the three different matrices of the contact potentials. It is found that the sequences designed with the MMJ matrix show much better folding behaviors both in the kinetics and in the thermodynamics than the other two kinds of contact potentials.

Considering these three matrices, their features determine the foldability of the model proteins folding with one of them. The MMJ matrix is created by shifting the reference state. As a result, the residues with weaker hydrophobicity than residue Thr show a repulsion between each other. This feature pushes the energy landscape upward along the energy axis. Consequently, the accessible conformational space is limited, which makes the searching to some degree easier than the MJ matrix. Meanwhile, the MJ matrix and the RMJ matrix have strong attractions between residues. Therefore, the folding shows two steps, condensation and folding. The local energy penalty to break the non-native bonds acts as the main barrier during their folding, which is not suffered by the MMJ case. At the same time, the stability of the sequences prohibits high temperature for the MJ and the RMJ cases. As a result, the folding under the MJ and the RMJ matrices appears to be wandering on a rugged landscape, and behaves as a slow one. Especially, for the RMJ matrix, it is formulated to support the phase separation. The mixing between different types of residues is disadvantageous in their energetics. This may prevent the fine adjustment during the folding. So, the sequences designed with the RMJ matrix may fold even slower than those designed with the two other matrices. Analyzing the ratio of the *HH* contacts in all the contacts suggests that the controlling factor in the MMJ matrix is different from that in the MJ matrix or in the RMJ matrix. These provide us with some insights into the assessment of the interactions for folding modeling.

For more detailed models rather than lattice models, there are more complexities included. It is reasonable to have more competitive configurations during their modeling. To establish an efficient folding, it would be necessary to include more ingredients into the interactions, rather than a single hydrophobicity. Therefore, from this aspect, we could speculate that the interactions with more ingredients (as the MMJ matrix) may be more consistent for folding modeling than those with unique feature (such as in the MJ and the RMJ matrices). It is plausible to believe that further off-lattice simulations with this matrix of contact potentials may provide the same behaviors. Anyhow, detailed simulations on the off-lattice modeling would be valuable to judge the applicability of various potentials, which deserves further studies.

ACKNOWLEDGMENTS

This work was supported by the Foundation of the NNSF (Grant Nos. 90103031, 10074030, and 10204013) and the Nonlinear Science Project (973) of the NSM.

[1] C. Anfinsen, Science **181**, 223 (1973).
[2] C. Levinthal, J. Chem. Phys. **65**, 44 (1968).
[3] P.G. Wolynes, J.N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).
[4] N.D. Socci and J.N. Onuchic, J. Chem. Phys. **103**, 4732 (1995).
[5] T. Veitshans, D. Klimov, and D. Thirumalai, Folding Des. **2**, 1 (1996).
[6] E. Shakhnovich, V.I. Abkevich, and O. Ptitsyn, Nature (London) **379**, 96 (1996).
[7] H. Nymeyer, A.E. Garcia, and J.N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **95**, 5921 (1998).
[8] K. Fan, J. Wang, and W. Wang, Euro. J. Phys. B **30**, 381 (2002).
[9] S. Miyazawa and R.L. Jernigan, Macromolecules **18**, 534 (1985).
[10] S. Miyazawa and R.L. Jernigan, J. Mol. Biol. **256**, 623 (1996).
[11] H. Li, C. Tang, and N. Wingreen, Phys. Rev. Lett. **79**, 765 (1997).
[12] J. Wang and W. Wang, Nat. Struct. Biol. **6**, 1033 (1999).
[13] K.A. Dill *et al.*, Protein Sci. **4**, 561 (1995).
[14] A. Sali, E. Shakhnovich, and M. Karplus, J. Mol. Biol. **235**, 1614 (1994).
[15] M.R. Betancourt and D. Thirumalai, Protein Sci. **8**, 361 (1999).
[16] M.R. Betancoust and P. Thirumalai, J. Mol. Biol. **287**, 627 (1999).
[17] A. Godzik, A. Kolinski, and J. Skolnick, Protein Sci. **4**, 2107 (1995).
[18] A. R. Leach, *Molecular Modelling Principles and Applications* (Addison-Wesley, London, 1996).
[19] H. Li, R. Helling, C. Tang, and N. Wingreen, Science **273**, 666 (1996).
[20] R.A. Broglia, G. Tiana, H.E. Roman, and E. Vigezzi, Proc. Natl. Acad. Sci. U.S.A. **95**, 12 930 (1998).
[21] N.D. Socci, J.N. Onuchic, and P.G. Wolynes, J. Chem. Phys. **104**, 5860 (1996).
[22] C.J. Camacho and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **90**, 6369 (1993).
[23] F. Chahine *et al.*, Phys. Rev. Lett. **88**, 168101 (2002).
[24] D.K. Klimov and D. Thirumalai, Phys. Rev. Lett. **76**, 4070 (1996).
[25] D.K. Klimov and D. Thirumalai, J. Chem. Phys. **109**, 4119 (1998).
[26] E.I. Shakhnovich, Phys. Rev. Lett. **72**, 3907 (1994).
[27] F. Seno *et al.*, Phys. Rev. Lett. **77**, 1901 (1996).
[28] E.I. Shakhnovich, Folding Des. **3**, R45 (1998).
[29] A.M. Gutin, V.I. Abkevich, and E.I. Shaknovich, Proc. Natl. Acad. Sci. U.S.A. **92**, 1282 (1995).
[30] M. Qin, J. Wang, T.P. Li, and W. Wang, Int. J. Mod. Phys. B **16**, 631 (2002).
[31] J.U. Bowie, R. Luthy, and D. Eisenberg, Science **253**, 164 (1991).
[32] A.G. Street, D. Datta, D.B. Gordon, and S.L. Mayo, Phys. Rev. Lett. **84**, 5010 (2000).