Vol 13 No 9, September 2004
1009-1963/2004/13(09)/1574-08

Chinese Physics



# Folding of lattice protein model chains with fixed ends*

Ji Gao-Feng(吉高峰), Xue Bin(薛 彬), and Wang Wei(王 炜)†

*National Laboratory of Solid State Microstructures, Department of Physics, Nanjing University, Nanjing 210093, China*

(Received 19 February 2004; revised manuscript received 18 May 2004)

Using Monte Carlo simulations, we have studied the folding dynamics and thermodynamics of geometrically constrained lattice protein model chains. The constraints are realized by fixing one or both terminals of the chains. By comparing the results with that of the free-end chains, we find that the folding behaviours of the end-constrained chains are not completely similar to that of the free-end chains. Both kinds of constraints on the chain ends affect the folding dynamics of the chains: i.e., the folding rate, but not the thermodynamics. The thermodynamic behaviour of the one-end-fixed chains shows less difference from that of the free-end chains, while the thermodynamic behaviour of the two-end-fixed chains has obvious difference from that of the free-end chains. The origin of these differences comes from the differences of the ergodicity of the chains in the conformational space.

**Keywords:** lattice protein model, Monte Carlo simulation, folding behaviour, constraints on the ends of chains

**PACC:** 8715, 7115Q

## 1. Introduction

The study of protein folding, i.e. finding the three-dimensional native structure of a protein from its primary amino acid sequence, is still a very interesting and unsolved problem. The necessity for studying this problem lies on the dependence of the protein's function on its native structure.[1−4] Protein chains can fold rapidly to their own native conformations under appropriate biological conditions. Many experimental methods and theoretical models have been developed to study and simulate the folding behaviours of various proteins during the past few decades. Experimentally, it is still impossible to observe the time course of the folding. The all-atom simulation, which takes into account many complex details of proteins, is also unrealistic due to the speed of computers at the present time. For example, it is difficult to perform a folding simulation even for a small part of protein with only 36 amino acids.[5] Hence many simplified models have been proposed. These simple models usually neglect some less-important factors but reserve the essential aspects of proteins. Thus these models mimic roughly many natures of real proteins and provide some qualitative description or even some quantitative characteristics of proteins.[1−21] The most commonly used $3 \times 3 \times 3$ lattice protein model considers the protein as a heteropolymer chain of 27 residues arranged in a simple cubic lattice. Although this model is simple, it may correspond to a helical protein with 60 residues.[8,9] Many researchers have used this lattice protein model to study proteins and protein folding, and have obtained many meaningful results.[6,10−19]

Previously, the lattice protein chains are free of geometrical constraints. Here in this paper, we study the influence of geometrical constraints on the folding behaviour and compare this behaviour with that of the unconstrained chains. The geometrical constraints under our consideration are of the one-end-fixed and the two-end-fixed constraints. Both kinds of constraints have relevant biological interest. It has been presumed that under biological conditions, the nascent polypeptide chain folds during its synthesis.[20] Thus, in the folding process, one end of the nascent polypeptide chain could be always considered as sticking on the ribosome. Due to the huge mass of the ribosome compared with the nascent polypeptide chain, the nascent polypeptide chain has a relatively slow velocity of motion and can be approximately modelled as a one-end-

*Project supported by the National Natural Science Foundation of China (Grant Nos 90103031, 10074030, 10021001 and 10204013), and the Nonlinear Project of the State Key Development Program for Basic Research of the NSM of China.

†Author to whom correspondence should be addressed. E-mail:wangwei@nju.edu.cn

fixed chain. It is worth noting that the dimension of ribosome could be much bigger than the polypeptide chain studied. Anyway, in a rough approximation, especially in the current lattice model in which only the nearest-neighbour interaction is taken into consideration, the effect of the dimension of the ribosome could be neglected. In addition, for a two-domain protein system, the folding of each domain can also be modelled by a one-end-fixed protein chain since it is argued that all the domains fold separately.[20,21] Furthermore, in a multi-domain protein system, the two terminals of the domain in the middle part are bounded to other two neighbouring domains. Then the domain in the middle part has a two-end-fixed constraint. Thus, the model chain which has constrained end or ends is close to the real situations in the biological environments. From the viewpoint of physics, the constraints on the end (or the ends) of a protein chain greatly reduce its conformational space. Due to the constraints, some conformations may be unreachable and the chain could not follow some folding pathways, thus the folding behaviour of the constrained chains may be different from that of the chains without the constraints. To the best of our knowledge, this has not been studied theoretically. Hence it is important to know whether the modification by taking into consideration the constraints affects the folding behaviour or not. If the constraints on the protein chains bring some difference to the folding behaviour from that of the free-end chains, we may need to reconsider the modelling of protein chains with small sizes. Nevertheless, if not, we can use the models of single domain to interpret the folding behaviour of large protein correspondingly. These are the main purposes of this work. The paper is arranged as follows. In Section 2, we briefly introduce the model and methods. In Section 3, we present the results and the discussions. Then we give a summary in the last section.

## 2. Model and methods

The model for a protein chain in this paper is a three-dimensional simple cubic lattice model which is the same as that used in Refs.[6,10]. The protein-like heteropolymer is modelled as a chain of 27 monomers which are divided into two types, i.e. $A$ and $B$. The arrangement of the monomers along the chain is the same as that with sequence 002 in Refs.[6,10]. The native conformation of this model chain is shown in Fig.1. The numbers in the figure are the indices for the monomers. The potential energy of the chain system is the sum of the contact interactions between every nearest-neighbour pairs of monomers except the covalently linked pairs. The interaction energy between monomers of the same type is $E_l = -3$, whereas that between monomers of different types is $E_u = -1$. Here the reduced units are used, i.e. the interaction energy is a multiple of unit energy $\varepsilon$. In addition, we choose $k_B = 1$, then the simulation temperature is also in reduced units of $\varepsilon/k_B$. The Metropolis Monte Carlo algorithm (MC) is adopted to simulate the folding processes. The move sets in the MC simulations are the end moves, the corner moves and the crankshaft moves.[6,10] We perform MC simulations at various temperatures in different end-constrained situations: 1) free-end, 2) one-end-fixed and 3) two-end-fixed. In all the three cases, the simulation steps (MC steps) are $10^9$ or $10^{11}$ for each run, and the total number of runs for each case is up to 500. In the simulations for the one-end-fixed chain, one terminal of the chain is bounded to a fixed point, while for the two-end-fixed chain, two terminals of the chain are set at the diagonal positions of a $3 \times 3$ two-dimensional simple lattice (see Fig.1).

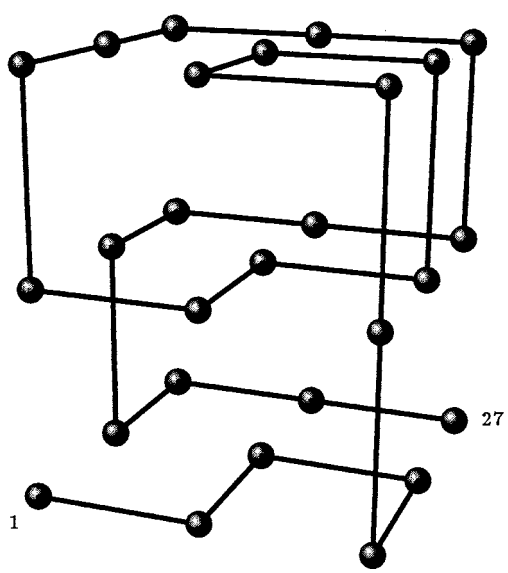


**Fig.1.** The native conformation of the lattice protein model chain studied. Each lattice site is denoted by a monomer index.

To calculate the thermodynamic quantities at various temperatures, the histogram method is used. The histogram method has been developed by many researchers [10,22–26] and is able to extract more information from one single simulation run. In this method, during a simulation at temperature $T$, the probability for every specific energy is recorded in a histogram $h(E,T)$. Meanwhile, $h(E,T)$ equals to the thermal

average of the density of states,

$$h(E,T) = \frac{n(E)\mathrm{e}^{-E/T}}{Z(T)}, \tag{1}$$

where $n(E)$ is the density of states for energy $E$, and $Z(T) = \sum_E n(E)\mathrm{e}^{-E/T}$. Hence the histogram method is also known as a method of density of states. Then at any given temperature $T'$, the thermal average of a certain quantity $A$, which is a function of $E$, can be calculated by

$$\langle A(E)\rangle = \frac{\sum_E A(E)n(E)\mathrm{e}^{-E/T'}}{\sum_E n(E)\mathrm{e}^{-E/T'}}, \tag{2}$$

where $\langle\cdots\rangle$ indicates the thermal average, $n(E)$ is resolved from Eq.(1) as $n(E) = h(E,T)\mathrm{e}^{E/T}Z(T)$. Then the specific heat at any temperature $T$ is given by $c_\mathrm{v}(T) = (\langle E(T)^2\rangle - \langle E(T)\rangle^2)/T^2$. Also, the free energy $F$ as a function of native contact number $Q$ is defined as $F(Q) = E(Q) - TS(Q)$, where $Q$ describes how much the chain system is similar to the native state, $S(Q)$ is the entropy and $n(Q)$ is the configurational density. All the details of these quantities could be found in the corresponding references.[10,11,27]

We define a 'transition region' which is a continuous range of native contact number $Q$, where the free energy shows a peak. For the model used in this work, the values of $Q$ in the 'transition region' range from 16 to 22. The reason why we choose this specific range can be found in Ref.[11]. We also use the symbol $P_\mathrm{nat}(T)$ which is the probability that the chain system is in the native state at temperature $T$, $P_\mathrm{nat}(T) = \exp(-E_\mathrm{nat}/T)/Z$, where $E_\mathrm{nat}$ is the energy of the native state and $Z$ is the partition function at temperature $T$. At the folding temperature $T_\mathrm{f}$, we have $P_\mathrm{nat}(T_\mathrm{f}) = 0.5$.[6,10] Hence the folding temperature can be obtained from the graph of $P_\mathrm{nat}(T)$ as a function of $T$.

# 3. Results and discussion

## 3.1. Dynamic analysis

In Fig.2 are shown the mean first passage times (MFPTs), i.e. the folding times, at various temperatures under different end-constrained conditions. The MFPTs are the averaged folding times (MC steps) needed to reach the native state from the initial extended states for the model protein chain. If the chain cannot find its native state within $10^9$ MC steps, we terminate the simulation and take $10^9$ MC steps as the folding time for this simulation. That is, $10^9$ MC steps is set to be the maximal folding time.[6] All the MFPTs are averaged over 500 independent runs with different initial conformations of the chain. When temperature $T > 3.1$, all the MFPTs are very long and most of them exceed the value of $10^9$ MC steps. Thus, all the chains do not fold at the high temperatures. When the temperature decreases from $T = 3.1$ to $T = 1.5$, all the MFPTs drop quickly and considerably to the order of $10^7$ or even lower. Further decreasing the temperature to $T = 1.0$, all the MFPTs rise quickly again. The general tendency in the MFPTs shows a 'V'-shape feature as the temperature changes, indicating that at both low and high temperatures, all the chains cannot fold due to the trap or the thermal unfolding, respectively. The slopes of MFPTs at low temperatures are larger than those at high temperatures. Thus all the chains are easy to be trapped at low temperatures. It is interesting to note that for the three cases with and without the constraints, the temperatures within the range of $1.3 < T < 1.75$ are most favourable for folding, i.e. the MFPTs are smaller at these temperatures than at others for the same chain. Meanwhile, the MFPTs for the one-end-fixed case are always longer than those for the free-end and the two-end-fixed cases. However, the MFPTs for the free-end case are longer than that for the two-end-fixed case when $T > 2.0$, and are shorter than that for the two-end-fixed case when $T < 2.0$. These variances in the accessibility to the native state reflect the differences in dynamics and the stabilities of the chains under different constraints.

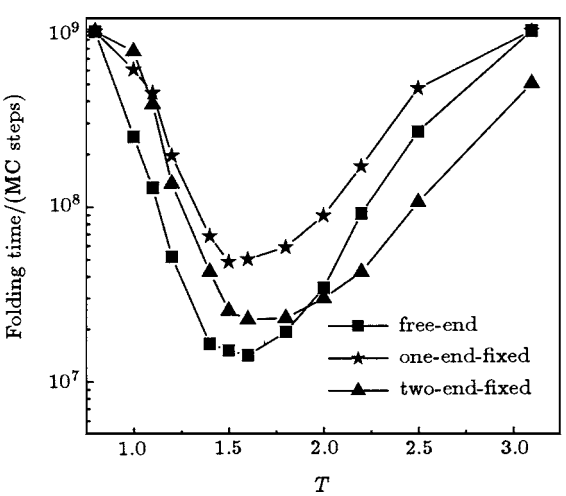


**Fig.2.** The MFPTs versus temperature for the sequence 002 under different constraints: i.e., the free-end, the one-end-fixed and the two-end-fixed cases. The vertical axis is in log base while the horizontal is linear.

It is well-known that the folding behaviours of protein chains are controlled by both the energy and the entropy of the chain systems. At low temperatures, the effect of energy dominates and traps the system in local minima easily, while at high temperatures, the effect of entropy takes over the role and leads the system to disordered phase. Thus, only in the medium temperature range, both the energy and the entropy balance each other. The systems can overcome the local minima and reach the native states within a short time.[6] In the case of one-end-fixed chains, the possible conformational space is reduced, making many pathways on the folding funnel unreachable. The chain may not find the fast folding pathways and the MFPTs in this case are long. However, the case of two-end-fixed chains is different although the conformational space is also reduced greatly. Since the two terminals are fixed at their native positions, the chains only experience a small number of conformations starting from their initial conformations. The chains fold faster than those with one-end-fixed do. It is noted that above temperature $T = 2.0$, the folding of the two-end-fixed chains is even faster than that of the free-end chains, while below $T = 2.0$ the folding is slower than that of the free-end chains. This is because at high temperatures the trapping effect on the chains in the local minima becomes weak.

In Fig.3, we show several time evolutions of the number of native contacts $Q(t)$ for three cases near the folding transition. The value of $Q(t)$ indicates the similarity of the conformation of the chain at time $t$ to the native conformation. From Fig.3(a) to (c), the simulation temperatures for the three cases are $T = 1.28$, 1.30 and 1.50, respectively, which are almost the same as their folding temperatures. It is clear that at the beginning stage, the values of $Q$ are small, then there are some switches between $Q \simeq 7$, i.e. the unfolded state, and $Q \simeq 28$, i.e. the native conformation. This indicates that there is a high folding cooperativity for the folding transition.

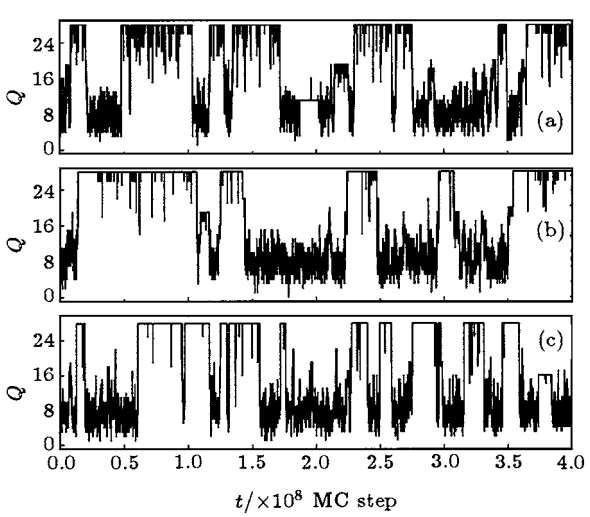


**Fig.3.** The time evolutions of $Q$ for the three kinds of chains. From Fig.3(a) to 3c) the figures are for the free-end, the one-end-fixed and the two-end-fixed chains, and the simulation temperatures are $T = 1.28$, 1.30 and 1.50, respectively.

To show the influence of native contacts on the conformation, we present the statistics of native contacts in the 'transition region' in Fig.4 for the three cases. The horizontal axis is the index for the native contacts: e.g., index 1 refers to the 1st native contact which represents the contact between the 1st and the 14th monomers in the native conformation, and so on. The maximal value of the index is 28, which is the 28th native contact for a $3 \times 3 \times 3$ simple cubic lattice. The vertical axis is the frequency of each native contact while the conformation is in the 'transition region'. During each simulation running up to $10^9$ MC steps, so long as the number of native contacts $Q$ is in the 'transition region', i.e. with a value in the range of $16 \leq Q \leq 22$, the index of the related native contact is recorded for calculating the frequency. The final results for the frequencies are averaged over 500 independent simulations. The frequencies reflect the importance and the stability of the native contacts during the folding, especially in the transition state. From Fig.4, we can see that in general the frequencies for the 1st, the 2nd and the 6th native contacts are smaller than other contacts.

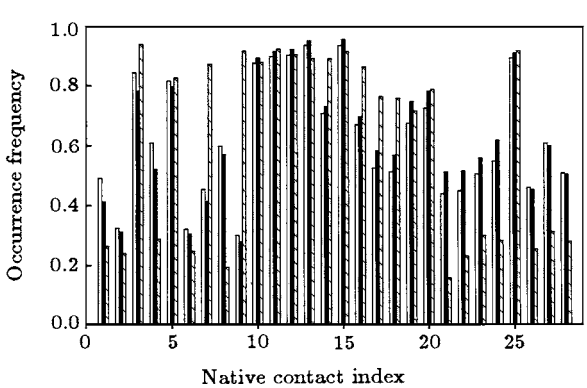


**Fig.4.** The occurrence frequency for each native contact in the 'transition region'. The simulation temperature is $T = 1.6$. The abscissa is the index of native contact which ranges from 1 to 28. Note that the blank, the grey and the slashed bars are for the free-end, the one-end-fixed and the two-end-fixed chains, respectively.

Hence, the monomers forming these contacts are 'active' in the folding process, whereas there are a lot of 'conservative' monomers such as those of the 3rd, the 5th and many other contacts of which frequencies are always higher than 0.8.

We can also see that the frequencies of the native contacts for the one-end-fixed chains are quite similar to those for the free-end chains. This means that the folding mechanisms of these two kinds of chains may be the same. However, the frequencies of the native contacts for the two-end-fixed chains show large differences from those for both the free-end and the one-end-fixed chains. For example, the frequencies of the 7th and the 9th contacts are much larger, and at the same time the frequencies of the last several native contacts with index larger than 20, except the 25th contact, are obviously small. This is due to the situation that the 7th and the 9th native contacts are formed between the 3rd and the 26th, the 4th and the 27th monomers, respectively. Because the 1st and the 27th monomers are fixed at the diagonal positions of the $3 \times 3$ two-dimensional plane of the cubic lattice, the 3rd and the 4th monomers can only take the crankshaft move sets after reaching equilibrium. These moves are obstructed by the very stable native contacts between the 2nd and the 13th monomers. Thus, the 7th and the 9th native contacts are very stable. In contrast, the low frequencies for the last several native contacts are due to the fact that they are spatially far away from the two constrained ends so that they cannot be formed easily.

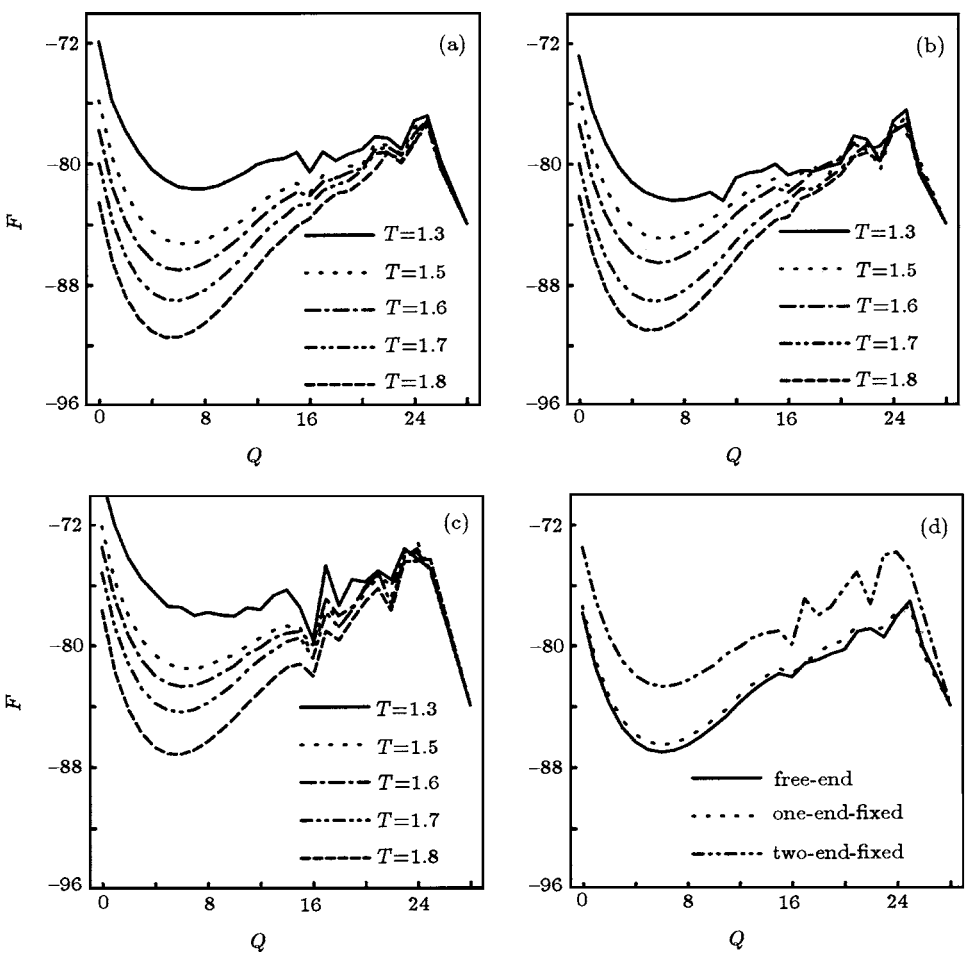


**Fig.5.** The plots of free energy $F(Q)$ against the reaction coordinate $Q$ at various temperatures for (a) the free-end chain, (b) the one-end-fixed chain, (c) the two-end-fixed chain. (d) A comparison for the above three cases at $T = 1.6$.

### 3.2. Thermodynamic analysis

In this section we discuss the thermodynamic properties of the three kinds of chains. The histogram method is used to calculate the free energy $F(Q)$ and the specific heat $c_v$. Preliminary simulations show that the critical transition temperatures of the three model systems are around 1.3–1.5, hence, the histogram samplings are taken at $T = 1.4$. In order to search the conformational space ergodically, the simulation steps in this section are set to $10^{11}$ MC steps. During such long time simulations the histograms for both $E$ and $Q$ are recorded for further analysis.[10,11,27]

We show the free energies as functions of the reaction coordinate $Q$ at different temperatures for the three kinds of chains in Fig.5(a)–(c), respectively. There are similar tendencies in all the curves. From Fig.5(a)–(c) we can see that for each kind of chain, when the temperature is low, the more the native contacts, the lower the free energy. Hence, the globally stable conformation is the native state. There are many local minima in the free energy curves, which can trap the chains, and there is a transition region where the free-energy shows a high barrier around $16 < Q < 26$. With raising temperature, the global minima of the free energies move to the left side where the values of $Q$ are around $Q \simeq 6$. That is, at high temperatures, conformations with less native contacts are preferred, and the chains no longer fold into their native conformations. Figure 5(d) shows a comparison among the three kinds of chains at $T = 1.6$. Clearly, the free energy curve for the two-end-fixed chains is rougher than those for the free-end and the one-end-fixed chains, and the values of the free energy are always larger than those for the other two kinds of chains. The local minima of the two-end-fixed chains are much deeper. Whereas, the free energy curves for the free-end and the one-end-fixed chains are quite close to each other. Hence, the thermodynamic behaviours of both the free-end and the one-end-fixed chains are similar, but are different from that of the two-end-fixed chains.

In Fig.6, the probabilities of the native state $P_{\rm nat}(T)$ for the three kinds of chains are plotted against temperature $T$. At low temperatures, all the values of $P_{\rm nat}(T)$ equal to 1, and as the temperature increases, these values decrease steeply from 1 to 0. The curves for the free-end and the one-end-fixed chains are almost coincident. The curve for the two-end-fixed chains shifts to the high temperature side. By applying $P_{\rm nat}(T_{\rm f}) = 0.5$, we obtain that the folding temperatures for these three kinds of chains are about $T_{\rm f} = 1.26$, 1.28 and 1.48, respectively. The two-end-fixed chain has a higher value of $T_{\rm f}$ than others. It is noteworthy that the folding transition for the three kinds of chains is very cooperative.

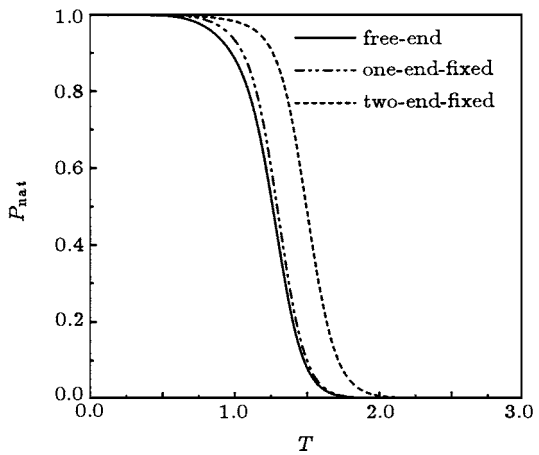


**Fig.6.** The probability of being in the native state $P_{\rm nat}$ versus temperature $T$ for the three kinds of chains. The data for this figure are obtained by extrapolation from the simulations at $T = 1.4$ using the histogram method.

In Fig.7 we show the specific heats for the three kinds of chains. In each case there appears a single peak in $c_v$ versus the temperature $T$. The curves of $c_v$ for the free-end and the one-end-fixed chains show basically no difference, but they are obviously different from the curve for the two-end-fixed chains. The temperatures of the specific heat peaks define the collapse temperatures $T_\theta$. The values of $T_\theta$ for the three kinds of chains are 1.33, 1.35 and 1.51, respectively. The highest value of $T_\theta$ is for the two-end-fixed chains. We find that the values of $\sigma = (T_\theta - T_{\rm f})/T_\theta$[28] for the three kinds of chains are 0.05, 0.05 and 0.02, respectively, which are basically the same within the error-bar. Hence, all the three kinds of chains under different constraints are good folders.[28] The physical origin for the similarities between the free-end chains and the one-end-fixed chains is that the conformational space of the one-end-fixed chains is almost the same as that of the free-end chains. Nevertheless, the conformational space of the two-end-fixed chains is greatly reduced. Consequently, the two-end-fixed case is different from the other two cases.

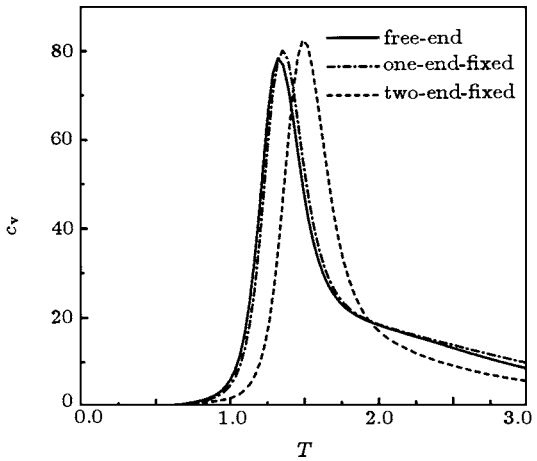


**Fig.7.** The specific heat versus temperature for the three kinds of chains. The data for this figure are obtained by extrapolation from the simulations at $T = 1.4$ using the histogram method.

## 4. Conclusion

We have studied the dynamics and thermodynamics of the geometrically constrained protein model chains by using a simple cubic lattice model. The geometrical constraints are realized by fixing one or both of the two terminals of a chain. We find that both of these two kinds of constraints affect the dynamic behaviours, i.e. the folding rate. The move sets chosen may influence the simulation results. To check the effectiveness of the currently used move sets, we introduce into these move sets a new move set of rigid rotation [29] to test the influence on dynamics of the move sets. The simulations give similar results. Hence, the dynamic results are reliable. Anyhow, it is interesting in the future to investigate thoroughly the influence on the dynamics of other move sets. In thermodynamics, the one-end-fixed chains show less difference from the free-end chains, but the two-end-fixed chains are obviously different from the free-end chains. The possible searching for the conformational space plays an important role in distinguishing the differences. It is found that using the one-end-fixed chains can search a partial conformational space as using the free-end chains can do, but some fast folding pathways cannot be reached. Hence, their dynamic behaviours may behave differently, but their thermodynamic behaviours are less different. The two-end-fixed chains have a greatly reduced conformational space, thus both the dynamics and thermodynamics are different from those of the free-end chains. As a result, for a two-domain protein system we could consider that each domain folds separately by using the previous model of single domain; yet for a multi-domain protein system, to understand the folding behaviour more clearly we should probably make a more detailed study.

The lattice model chains used in this paper have 27 monomers, which reflects some secondary structures of one single domain of proteins. The potentials used in this work have only two types. For more detailed description of the interactions among monomers, the potentials can be replaced by many other kinds of choices, such as those in Refs.[30–32] and so on. The chain length can also be prolonged to mimic more realistic proteins. We expect that the longer chains may have more complicated behaviours. Due to the complexity in the biological environment, the constraints used here can also be developed to many other kinds to study the influence of environments on the folding of chains.[33,34]

## References

[1] Elber R 1996 *Recent Developments in Theoretical Studies of Proteins* (Singapore: World Scientific) p 1
[2] Wolynes P G, Onuchic J N and Thirumalai D 1995 *Science* **268** 960
[3] Onuchic J N, Luthey-Schulten Z and Wolynes P G 1997 *Annu. Rev. Phys. Chem.* **48** 545
[4] Eaton W A, Munoz V, Hagen S J, Jas G S, Lapidus L J, Henry E R and Hofrichter J 2000 *Annu. Rev. Biophys. Biomol. Struct.* **29** 327
[5] Duan Y and Kollman P A 1998 *Science* **282** 740
[6] Socci N D and Onuchic J N 1994 *J. Chem. Phys.* **101** 1519
[7] Pande V S and Rokhsar D S 1999 *Proc. Natl. Acad. Sci.* USA **96** 1273
[8] Onuchic J N 1997 *Proc. Natl. Acad. Sci.* USA **94** 7129
[9] Onuchic J N, Wolynes P G, Luthey-Schulten Z and Socci N D 1995 *Proc. Natl. Acad. Sci.* USA **92** 3626
[10] Socci N D and Onuchic J N 1995 *J. Chem. Phys.* **103** 4732
[11] Socci N D, Onuchic J N and Wolynes P G 1996 *J. Chem. Phys.* **104** 5860
[12] Chan H S and Dill K A 1991 *Annu. Rev. Biophys. Biophys. Chem.* **20** 447
[13] Miller R, Danko C A, Fasolka M J, Balazs A C, Chan H S and Dill K A 1992 *J. Chem. Phys.* **96** 768
[14] Chan H S and Dill K A 1993 *J. Chem. Phys.* **99** 2116
[15] Fiebig K M and Dill K A 1993 *J. Chem. Phys.* **98** 3475
[16] Shakhnovich E, Farztdinov G, Gutin A M and Karplus M 1991 *Phys. Rev. Lett.* **67** 1665
[17] Shakhnovich E I and Gutin A M 1990 *Nature* **346** 773
[18] Shakhnovich E I and Gutin A M 1993 *Proc. Natl. Acad. Sci.* USA **90** 7195

[19] Camacho C J and Thirumalai D 1993 *Proc. Natl. Acad. Sci.* USA **90** 6369
[20] Netzer W J and Hartl F U 1997 *Nature* **388** 343
[21] Orengo C A, Jones D T and Thornton J M 1994 *Nature* **372** 631
[22] Ferrenberg A M and Swendsen R H 1988 *Phys. Rev. Lett.* **61** 2635
[23] Ferrenberg A M and Swendsen R H 1989 *Phys. Rev. Lett.* **63** 1195
[24] Alves N A, Berg B A and Villanova R 1990 *Phys. Rev.* B **41** 383
[25] Bouzida D, Kumar S and Swendsen R H 1992 *Phys. Rev.* A **45** 8894
[26] Humar S, Bouzida D, Swendsen R H, Kollman P A and Rosenberg J M 1992 *J. Comput. Chem.* **13** 1011
[27] Clementi C, Nymeyer H and Onuchic J N 2000 *J. Mol. Biol.* **298** 937
[28] Veitshans T, Klimov D and Thirumalai D 1997 *Folding Des.* **2** 1
[29] Chan H S and Dill K A 1994 *J. Chem. Phys.* **100** 9238
[30] Miyazawa S and Jernigan R L 1985 *Macromolecules* **18** 534
[31] Kolinski A, Godzik A and Skolnick J 1993 *J. Chem. Phys.* **98** 7420
[32] Wang J and Wang W 1999 *Nature Struct. Biol.* **6** 1033
[33] Betancourt M R and Thirumalai D 1999 *J. Mol. Biol.* **287** 627
[34] Li W, Wang P Y, Dou S X and Tong P Q 2003 *Chin. Phys.* **12** 226