

## Effect of interaction energy fluctuation on the folding of a proteinlike model

Zhongwen Xing, Jun Wang, and Wei Wang

*National Laboratory of Solid State Microstructure and Department of Physics, Institute for Solid State Physics,  
Nanjing University, Nanjing 210093, China*

*and Center of Advanced Science and Technology in Solid State Physics, Nanjing 210093, China*

(Received 25 November 1997; revised manuscript received 6 April 1998)

The folding of a proteinlike model with two kinds of distributions of interaction energy fluctuation—a frozen fluctuation and a fluctuation varying with time—describing disorder effects of frustration in the energy spectrum is studied. It is found that the former is unfavorable for the folding, while the latter is favorable for the folding. The folding time versus the fluctuation is simulated via the Monte Carlo method. An analytical argument for the folding time is also made, which is in good agreement with the simulations. In addition, a scaling behavior of the folding time with chain length and a range of temperature for fast folding are also obtained. [S1063-651X(98)11908-6]

PACS number(s): 87.10.+e, 64.60.Cn, 82.20.Db

The folding of a protein to a native structure in molecular biophysics has been subjected to intense investigation for a long time [1,2]. Some new understandings have recently been achieved for a physical picture of folding, namely, the folding funnel and energy landscape, resulting from advances in experimental methods and from theoretical studies based on some simple proteinlike models [3–17]. This new view replaces the pathway concept of sequential events with the funnel concept of parallel events during the folding of a protein, and the new perspective sees folding as a diffusion-like process in a complex “rugged” energy landscape [3,17] (also see the references therein). Folding is easy if the landscape resembles a many-dimensional funnel leading through a myriad of pathways to the native structure. This folding process is argued to depend on the driving effects which include the slope of the folding funnel resulting from a decrease in entropy or conformations, and, of course, also on the interactions between the amino acids in proteins [3,17]. The latter determine the native structure of the protein [18].

Recently, a study on the folding of a proteinlike model was made in Ref. [19]. In this model a polypeptide chain was assumed to have  $M$  monomers, and only two-body connectivity to be accessible. Due to the saturation of the interactions there are at most  $N = M/2$  bonds in the chain. To ensure that the nativelike state is not only possible but also stable, these connections are classified into two classes—namely, native and non-native—according to their binding energies  $-E_n$  and  $-E_{nn}$ , with  $E_n > E_{nn}$ , respectively. The native contacts are defined as a set of  $N$  distinct pairs of residues which are closest in space in the native conformation.

The ratio  $\Delta = E_{nn}/E_n$  describes the effect of frustration. The folding of the chain is a diffusion process between the levels in an energy spectrum constructed by different combinations of  $i$  native and  $j$  non-native bonds. An effective folding transition to an order structure, and a folding time scaled as  $M^\lambda$  with  $\lambda \approx 3$  have been found.

The energy spectrum shown in Ref. [19] may relate to a smooth folding funnel [3,17] or a uniform arrangement of the levels, since the binding energies are only taken with two values. It is well known that the interaction strengths or inter-residue contact energies between any two types of amino

acids in protein can be tabulated by a matrix [18]. However, in some studies the heterogeneity of interactions of proteins is further simplified as a Gaussian distribution [8,13]. Actually, for real proteins, since the bonds may have different distances and angles of interaction, etc., the combining energy between two monomers, regardless of the native or non-native bond, is not an equal constant. Furthermore, the binding energies may vary with time or fluctuate around their values of the native state [20]. These, to certain degree, may be associated with the disorder effects of frustration. Thus the energy spectrum would be nonuniform, and would fluctuate with time resulting from various values of the energies due to different contacts during the folding. The folding features, such as the folding time, are different. We expect that such a situation is close to the real cases of the protein in nature.

Motivated by the aforementioned considerations, in this work we report a study on the folding by taking into account two nonuniform distributions of the energy levels. *Case A* is a frozen fluctuation with respect to the uniform levels, and *case B* a fluctuation varying with time. Such a model including fluctuation of the interaction energy is expected to describe the heterogeneity of the heteropolymers or proteins. Following the model studied in Ref. [19], by using Monte Carlo simulation and a simple analytical treatment, we find that the folding time is slower for case A and faster for case B as the fluctuation increases than that of the zero fluctuation in Ref. [19]. In addition, a range of temperature of the fast folding and a scaling of the folding time with chain length  $M$  are also obtained.

The energy level  $[i, j]$ , including all possible distinct conformations of the model, is given by

$$E(i, j) = -iE_n - jE_{nn}, \quad (1)$$

where  $i$  and  $j$  are the number of native bonds and non-native bonds, respectively. If the combining energy of a bond between two monomers either  $E_n$  or  $E_{nn}$  is a constant, i.e.,  $E_n = E_n^0$  and  $E_{nn} = E_{nn}^0$  being the energies without fluctuation, the energy spectrum is uniform like the one in Ref. [19]. Figure 1(a) shows such a spectrum and a folding pathway

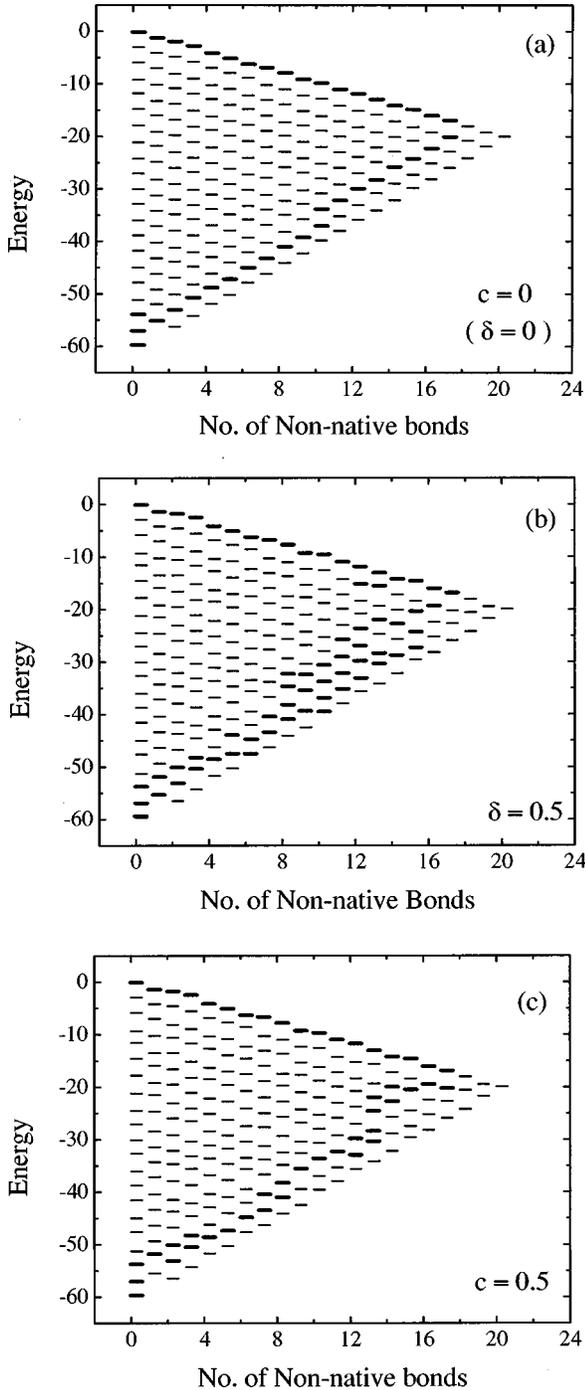


FIG. 1. The energy spectra and folding pathways for a chain  $M=20$ ,  $E_{\text{nn}}^0/E_n^0 = \frac{1}{3}$ , and  $E_{\text{nn}}^0 = 1$ . (a) Zero fluctuation or uniform spectrum [Eq. (1)]. (b) A frozen fluctuation with  $\delta=0.5$  in Eq. (2). (c) A fluctuation with time with  $c=0.5$  in Eqs. (5) and (6). Here the levels are shown at a time after the system reaches its native state. The thick bars define the pathway of the maximum number of visits over all runnings.

(see following discussion). In this work, we consider two different cases of the nonuniform spectrum for Eq. (1).

*Case A—frozen fluctuation:* We introduce an energy fluctuation  $2\delta$  in Eq. (1),

$$E(i,j) = -iE_n^0 - jE_{\text{nn}}^0 \pm \delta, \quad (2)$$

where  $\pm$  is assumed to be randomly distributed on each level  $[i,j]$ , and  $\delta$  describes the degree of fluctuation. Once the chain length  $M$  is set and the degree of the fluctuation  $\delta$  is chosen, the fluctuation is initially distributed on each level according to its sign in Eq. (2). Physically, Eq. (2) represents a frozen fluctuation or a fluctuation independent of time for each level, which describes the disorder effects of frustration. The spectrum and a pathway of the folding with  $\delta=0.5$  are shown in Fig. 1(b). The Monte Carlo simulation is programmed as follows. First, a nativelylike state with  $N=M/2$  native bonds or pairs between specific monomers is assumed, or an *ad hoc* set of native couples of monomers is given. If two monomers do not belong to one such pair, they form only non-native bonds. It is stipulated that there is no binding ( $i=j=0$ ) in the chain at the initial time. Then, starting from this initial state, we randomly select two monomers in the chain. We have several choices for the next steps. (1) If they are neither bound to each other nor bound to other monomers, we let them bind to each other to form a native bond (or a non-native bond) for a native couple (or for a non-native couple). (2) If both selected monomers or one of these two monomers are bound with other monomers, they are abandoned and the program goes to random selection again. (3) If the two monomers are bound to each other with a native or non-native bond, they are broken with a penalty probability  $P = \exp[-(E_{\text{new}} - E_{\text{old}})/k_B T]$ , where  $E_{\text{new}}$  and  $E_{\text{old}}$  are the energies [see Eq. (2)] after and before breaking, and  $T$  is the temperature, respectively. A Monte Carlo step is counted whatever happens for the random sampling. When all monomers are bound to each other by the native bond ( $i=N$ ), the total energy of the system is at a minimum, and the system reaches its ground state or folded state. The Monte Carlo steps (MC's) from the initial state to the folded state are defined as the folding time. The statistical average is taken over 100 to 200 runnings depending on the system size. The energies are measured in units of  $E_{\text{nn}}^0$ , i.e.,  $E_{\text{nn}}^0 = 1$ , and  $E_n^0$  is set as  $E_n^0 = 3$ , and the temperature is taken as  $T=0.3$ . Notice that we have chosen the Boltzmann constant  $k_B=1$ .

From Fig. 1(b), we can see that the folding processes are different for different values of the fluctuation  $\delta$  [the pathway for  $\delta=0$  relates to the folding in Ref. [19]; see Fig. 1(a)]. The folding process could be divided into two stages. In the first stage, most of the bonds become non-native within a short time. The pathway, defined as the maximum number of visits over all runnings, goes along the non-native direction. In the second stage, however, these non-native bonds are continuously broken one by one, and at the same time native bonds are formed. The pathway turns in the native direction, and most of the folding time is elapsed in this stage. Finally, the system reaches the nativelylike state or folded state quickly.

In Fig. 2, we show the folding time versus the fluctuation  $\delta$  for a chain  $M=40$ . We see that there is an increase in the folding time as the fluctuation  $\delta$  increases. It is found that the folding time can be scaled as  $t_\delta \sim t^0 \exp(\lambda_1 \delta)$ , with  $\lambda_1 \approx 3.89$  and  $t^0$  the folding time without fluctuation  $\delta=0$ . This means that as the fluctuation  $\delta$  increases, the folding slows down. This result can be understood by the following argument. Since most of time is elapsed in the second stage of the folding, the time is mainly determined by that of continuous

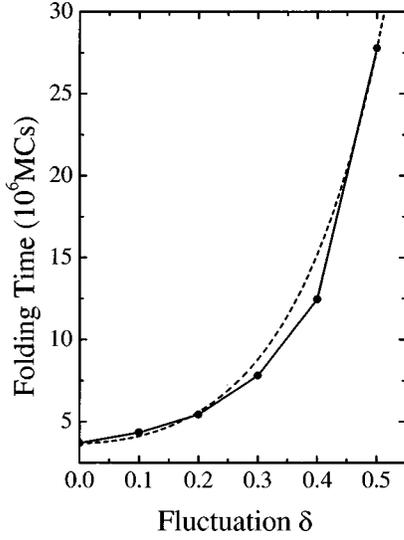


FIG. 2. The folding time (MC's) vs the energy fluctuation  $\delta$ , for  $M=40$ ,  $T=0.3$ ,  $E_n^0=3.0$ , and  $E_{nn}^0=1.0$ . The spectrum is frozen [see Fig. 1(b)]. Solid circles show the results from the Monte Carlo simulations. The dashed line shows the results by quantitative analysis [see Eq. (4)].

breaking of the non-native bonds and the forming of the native bonds (some non-native bonds are also formed). Physically, these relate to the probabilities of breaking and forming of both kinds of bonds. In a uniform spectrum model, on each level, the time of breaking a non-native bond is given by

$$t^0 \sim 1/P_{nn}^0 = \exp(E_{nn}^0/T), \quad (3)$$

where  $P_{nn}^0$  is the probability of breaking a non-native bond in the case of zero fluctuation. In the presence of the fluctuation, there are three possibilities for a non-native bond. If the energy interval increases ( $\Delta E = E_{nn}^0 + 2\delta$ ), the breaking time goes up by  $t_+ = \exp(\Delta E) = t^0 \exp(2\delta/T)$ .

The probability of this possibility is equal to 0.25. In the opposite case,  $\Delta E = E_{nn}^0 - 2\delta$ , the breaking time decreases by  $t_- = t^0 \exp(-2\delta/T)$ . This probability is also equal to 0.25. The remainder of the probability, 0.5, belongs to a case where the two neighbor energy levels increase or decrease simultaneously, i.e., the energy interval  $\Delta E = 0$ . These give the folding time

$$t_\delta \sim \frac{t^0}{4} [\exp(2\delta/T) + \exp(-2\delta/T) + 2] = t^0 [ch(\delta/T)]^2. \quad (4)$$

From Eq. (4), one can see that  $t_\delta$  is always greater than  $t^0$ . As a result, the present model is unfavorable for the folding. Obviously,  $t_\delta$  returns to the uniform spectrum model in the limit of  $\delta \rightarrow 0$ .

*Case B—fluctuation with time:* Now we consider a case where the distribution of the spectrum can fluctuate with time. In this case, we choose  $E_n$  and  $E_{nn}$  in Eq. (1) by

$$E_n = E_n^0 [1 + 2c(r - 0.5)], \quad (5)$$

$$E_{nn} = E_{nn}^0 [1 + 2c(r - 0.5)]. \quad (6)$$

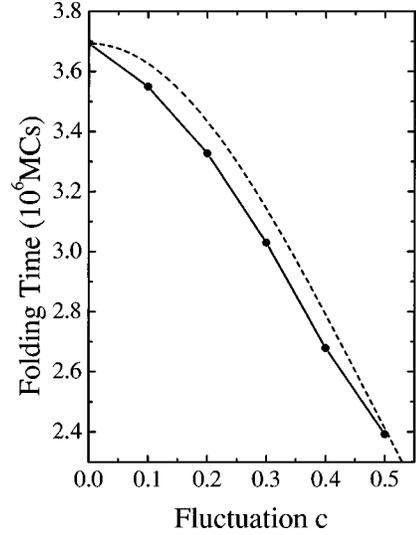


FIG. 3. The folding time (MC's) vs the fluctuation  $c$ , for  $M=40$ ,  $T=0.3$ ,  $E_n^0=3.0$ , and  $E_{nn}^0=1.0$ . The spectrum fluctuates with time [see Eqs. (5) and (6) or Fig. 1(c)]. Solid circles show the results from the Monte Carlo simulation. The dashed line shows the relation by quantitative analysis [see Eq. (9)].

Here  $r$  is a random number in a region (0,1), and changes its value after each Monte Carlo step, and  $c$  is a parameter describing the fluctuation. Equations (5) and (6) include small energy fluctuations around  $E_n^0$  and  $E_{nn}^0$ . Here again, the energies without fluctuation are set as  $E_n^0=3$  and  $E_{nn}^0=1$ . The binding energy for each bond is determined by a random number  $r$ . In this model, the folding pathways for different amplitudes  $c$  of the fluctuation are similar to that of case A [see Fig. 1(c)]. However, the folding time  $t^c$  is found to decrease with the increasing in  $c$ , as shown in Fig. 3. This implies that the folding is fast as the fluctuation becomes large. This behavior can be well scaled by  $t^c = t^0 - \lambda_2 c$  with  $\lambda_2 \approx 2.7 \times 10^6$  MC's. Such a result can be understood as follows. For each level  $[i, j]$  with  $E_{nn} = E_{nn}^0(1+x)$ , the probability of breaking a non-native bond is

$$P_{nn}^c(x) = \exp[-E_{nn}^0(1+x)/T] = P_{nn}^0 \exp(-x/T), \quad (7)$$

where  $E_{nn}^0=1$  has been used. Since the binding energy changes continuously from  $E_{nn}^0(1-c)$  to  $E_{nn}^0(1+c)$ , the average probability of breaking a non-native bond is

$$P_{nn}^c = \frac{P_{nn}^0}{2c} \int_{-c}^c \exp(-x/T) dx = \frac{P_{nn}^0 T}{c} \sinh(c/T). \quad (8)$$

Thus the folding time is approximately given by

$$t^c \sim t_{nn}^c \propto t^0 \frac{c}{T} [\sinh(c/T)]^{-1}, \quad (9)$$

where  $t_{nn}^c$  is the time for breaking a non-native bond with  $t_{nn}^c \sim 1/P_{nn}^c$ . From Eq. (9), it follows that  $t^c$  is always smaller than  $t^0$ . As a result, this model is favorable for the folding, and the increasing in the fluctuation speeds up the folding, i.e., the folding time decreases. From Fig. 3, we can see that the analytical argument is well in agreement with

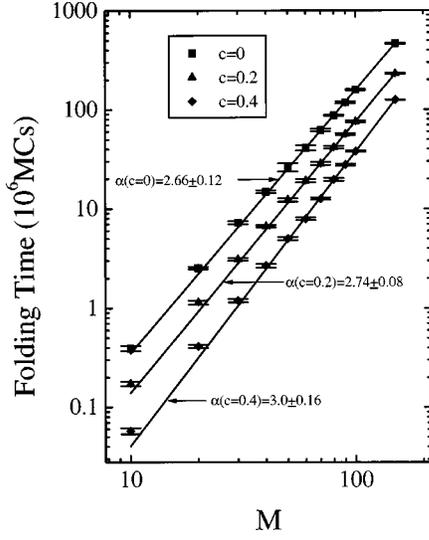


FIG. 4. The logarithmic-logarithmic plot for the folding time (MC's) vs the chain length  $M$  with  $T=0.3$ ,  $E_n^0=3.0$ , and  $E_{nn}^0=1.0$ . The spectrum fluctuates with time [see Eqs. (5) and (6) or Fig. 1(c)]. The symbols denote  $c=0.0$  (squares),  $0.2$  (triangles), and  $0.4$  (diamonds), respectively. The curves for  $c=0$  and  $0.2$  are shifted by constants  $\log_{10}4$  and  $\log_{10}2$  in the vertical direction, respectively.

the simulation. It is noted that the breaking of native bonds slows down the folding as the fluctuation increases. Similar to the argument for obtaining Eq. (9), the breaking probability of a native bond is

$$P_n^c = \frac{P_n^0}{2c} \int_{-c}^c \exp(-E_n^0 x/T) dx = \frac{P_n^0 T}{E_n^0 c} \sinh(E_n^0 c/T). \quad (10)$$

From Eq. (10), we can see that the probability  $P_{nn}^c$  is always larger than  $P_n^0$  (the breaking probability for the native bond at zero fluctuation  $c=0$ ). This means that the breaking of the native bonds always slows down the folding, since it takes a time  $t_n^c \sim 1/(1-P_n^c)$ . However, this effect of slowing down does not dominate the folding since the number of the native bonds to be broken is small.

In order to check the effect of the system sizes, we further study the folding processes of different chain lengths  $M$  with different amplitudes  $c$ , as shown in Fig. 4. From Fig. 4, first we can see that as the fluctuation increases the folding of the system becomes fast. For three different amplitudes of the fluctuation the folding time can be scaled as  $t^c \sim M^{\alpha(c)}$ , with  $\alpha(c=0)=2.66 \pm 0.12$ ,  $\alpha(c=0.2)=2.74 \pm 0.08$ , and  $\alpha(c=0.4)=3.00 \pm 0.16$ . The exponents  $\alpha(c)$  increase as the amplitude  $c$  of the fluctuation increases, and they seem to satisfy a simple relation of  $\alpha(c) = \alpha(0) + 2.25c^2$ . The increase in these exponents results from the decrease in  $t_{nn}^c$  [see Eq. (9)] as the value of  $c$  increases. Second, as the chain length  $M$  increases, we can see that the folding time increases. When a system has more monomers  $M$ , i.e., a longer chain, the number of the non-native bond is larger. This implies that there are more non-native bonds to be broken and more native bonds to be formed (the number of breaking native bonds is also large). These factors make the system reach its nativelike state at a longer time, i.e., the folding becomes

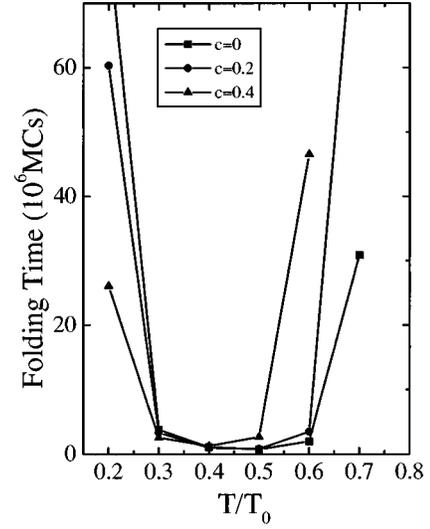


FIG. 5. The folding time (MC's) vs the temperature  $T/T_0$  ( $T_0 = E_{nn}^0/k_B = 1$ , with  $k_B = 1$ ) for  $M=40$ ,  $E_n^0=3.0$ , and  $E_{nn}^0=1.0$ . The symbols denote  $c=0.0$  (squares),  $0.2$  (circles), and  $0.4$  (triangles), respectively.

slow. Interestingly, when  $M > 120$  the existence of the fluctuation is no longer favorable for the folding, and the folding time  $t^c$  is longer even than that of the zero fluctuation for  $M > 200$ . This is because the effect of the nonuniformity of the energy spectrum is diminished when the chain length  $M$  reaches a certain value. Physically, the nonuniformity in the spectrum or the fluctuation in the interaction energy describe some features of the frustration, while the large chain length  $M$  relates to large number of conformations or large entropy. Therefore, for a short chain (or small  $M$ ), the folding is dominated by the effect of frustration, whereas for a long chain (or large  $M$ ) the folding is dominated by the effect of entropy. This further confirms the main conclusion of Ref. [19].

It is worth noting that the scaled exponent  $\alpha(c=0) = 2.66 \pm 0.12$  appears smaller than the one obtained in Ref. [19]. This results from the definition of the model. On the one hand, since the sampling probability for a native couple of two specific monomers is given by  $P_n^s = 1/(M-2q-1)$ , with  $q=i+j$  the total number of bonds [for a non-native couple, the sampling probability is  $P_{nn}^s = (M-2q-2)/(M-2q-1)$ ]; the folding speeds up as  $q$  increases. On the other hand, after two free monomers are randomly chosen, they form absolutely either a native bond or a non-native bond since the energy of the system decreases (this is the same case as treated by Socci and Onuchic for a probability  $p = \exp[-(E_{\text{new}} - E_{\text{old}})] > 1$  in Ref. [11]). The chain performs its folding to a native structure quickly and the folding time is short with a small exponent  $\alpha$ . Physically, our results imply that if an *ad hoc* set of native couples of monomers is specified, the folding time is short, or the protein behaves like a fast folding. Such a specification mimics the compact nativelike state, and cannot be implied as an initial state. As a result, the scaled exponent  $\alpha(c=0)$  is close to  $\alpha = 2.70$ , as obtained in Ref. [21] for the model of Ref. [22] in which an idea sequence is designed. In some sense, our specification may be associated with the design of the native interaction in the model of Ref. [22]. Differently, in Ref. [19], the formation of a new bond between any two monomers relates to a

factor  $k(q) \sim (M-2q)(M-2q-1)/2(q+1)$  due to the requirement of a detailed condition, which turns out to be the same for both native and non-native bonds. Because most of the folding time is spent with  $q \sim M/2$  (most of the monomers are connected by non-native bonds within a short time), the factor  $k(q)$  is small, and  $k(q) \sim 1/M$  decreases as  $M$  increases. Thus the folding becomes slower, and the exponent  $\alpha$  is larger due to the random pairing of monomers with native or non-native bonds [23].

Finally, we study the relation between folding time and temperature for case B. It is well known that the protein can reach its native state only in a small range of temperature. As shown in Fig. 5, we also find such a range. This range becomes narrow and shrinks to the low temperature region as the fluctuation increases. At the same time, the folding speeds up at lower temperatures for larger energy fluctuation. Nevertheless, the folding slows down at high temperatures. This is because the breaking probabilities of both native and non-native bonds are directly proportional to the temperature, and are inversely proportional to the amplitude of the fluctuation  $c$  ( $P_n^c$  and  $P_{nn}^c \sim T/c$ ). That is, the folding time is  $t \sim t_n^c + t_{nn}^c$ , with  $t_n^c \sim 1/(1-P_n^c)$  and  $t_{nn}^c \sim 1/P_{nn}^c$ . Since the folding time is dominated by the effect of breaking of the non-native bonds at large fluctuation and low temperature,

the folding speeds up. Conversely, the effect of breaking the native bonds is apparently large at large fluctuation and high temperature, which makes the folding slow down, or the folding time increase.

In conclusion, in this work, the effect of a fluctuation of the interaction energy on the folding dynamics for a protein-like model was studied. We found that an interaction energy spectrum with a frozen fluctuation, or a frozen energy landscape that is less kinetically dependent, is unfavorable for the folding, whereas the spectrum constructed with a kinetic landscape is favorable for the folding. The folding is really a self-organized kinetic process [24], and the “folding funnel” should have spatiotemporal cooperative features in order to make sure that the “water flows along different routes down mountainsides” [17] or to behave like a fast folding of protein in nature.

One of the authors (W.W.) acknowledges support by the National Laboratory of Solid State Microstructure, the Young Research Foundation of the Committee of National Education, the National Natural Science Foundation (Grant No. 19604006), and the Outstanding Young Research Foundation of the National Natural Science Foundation (Grant No. 19625409).

- 
- [1] C. Levinthal, *J. Chim. Phys. Phys.-Chim. Biol.* **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] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, *Proteins: Struct., Funct., Genet.* **21**, 167 (1995).  
 [5] J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten, and N. D. Socci, *Proc. Natl. Acad. Sci. USA* **92**, 3626 (1995).  
 [6] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, *J. Chem. Phys.* **104**, 5860 (1996).  
 [7] J. D. Bryngelson and P. G. Wolynes, *J. Phys. Chem.* **93**, 6902 (1989).  
 [8] E. Shakhnovich, G. Farztdinov, A. M. Gutin, and M. Karplus, *Phys. Rev. Lett.* **67**, 1665 (1991).  
 [9] C. J. Camacho and D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **90**, 6369 (1993).  
 [10] H. S. Chan and K. A. Dill, *J. Chem. Phys.* **100**, 9238 (1994).  
 [11] N. D. Socci and N. Onuchic, *J. Chem. Phys.* **101**, 1519 (1994).  
 [12] V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich, *J. Chem. Phys.* **101**, 6052 (1994).  
 [13] A. Sali, E. Shakhnovich, and M. Karplus, *Nature (London)* **369**, 248 (1994).  
 [14] K. A. Dill *et al.*, *Protein Sci.* **4**, 561 (1995).  
 [15] M. H. Hao and H. A. Scheraga, *J. Phys. Chem.* **98**, 9882 (1994).  
 [16] M. H. Hao and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **93**, 4984 (1996).  
 [17] K. A. Dill and H. S. Chan, *Nat. Struct. Bio.* **4**, 10 (1997).  
 [18] S. Miyazawa and R. L. Jernigan, *Macromolecules* **18**, 534 (1985); *J. Mol. Biol.* **256**, 623 (1996).  
 [19] C. J. Camacho, *Phys. Rev. Lett.* **77**, 2324 (1996); **80**, 207 (1998).  
 [20] As pointed out by Baldwin [*Nature (London)* **369**, 183 (1994)] the pairwise interaction between two residues may be different between the close-packed native structures and the folding intermediates.  
 [21] A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich, *Phys. Rev. Lett.* **77**, 5433 (1996).  
 [22] H. Taketomi, Y. Ueda, and N. Go, *Int. J. Pept. Protein Res.* **7**, 445 (1975).  
 [23] If the monomers are not specified in the simulation, the factor  $k(q)$  is included automatically in the kinetic rules, and the exponent  $\alpha(c=0)$  is close to that obtained in Ref. [19].  
 [24] It is noted that such a self-organized process does not relate to the so-called self-organized criticality [see P. Bak, C. Tang, and K. Wiesenfeld, *Phys. Rev. Lett.* **59**, 381 (1987)], but to an *ab initio* organization with a unique final state.