Folding Behaviour for Proteins BBL and E3BD with Gō-like Models *

ZUO Guang-Hong(左光宏), ZHANG Jian(张建), WANG Jun(王骏), WANG Wei(王炜)**

National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, Nanjing 210093

(Received 7 March 2005)

Folding behaviour of protein BBL and its homologue domain E3BD are studied by using an off-lattice Gō-like model. It is found that the folding behaviours of these two proteins are different. Protein BBL folds in a downhill manner, which is consistent with experiments. In contrast, protein E3BD folds cooperatively and has a bimodal distribution of the Q values (the similarity to the native state). By analysing the native structures of the two proteins, it is found that the difference in folding behaviours can be attributed to the different structural features described by the number of nonlocal contacts per residue.

PACS: 87.14.Ee, 87.15.Cc, 87.64.Aa

It is generally accepted that the folding of a protein is determined by its sequence,[1−5] and moreover, the native topology of the protein is related to many features of the free-energy funnel.[6] Therefore, it is plausible to consider that homologous proteins which have similar sequences and native conformations may fold in a similar manner. However, this is not always true. Experiments have shown that some proteins in a same superfamily may fold with different mechanisms.[7] Studies on these proteins suggested that this is due to the role of local stability in the folding mechanism.[7] Here we discuss the importance of the native topology in the folding behaviour.

Recently, four homologous proteins of protein BBL have been studied experimentally.[8,9] These proteins, e.g., protein BBL (PDB: 1BBL) and E3BD (PDB: 1EBD), are similar in sequences and in native structures. Sequence alignment shows that their sequence identity is 34%, and the *RMSD* of 33 residues (exclude the terminal residues) is even smaller than 1.76 Å (by CE software). Superimposed three-dimensional structures of both the proteins are shown in Fig. 1. However, different folding behaviours were declared by researchers.[8,9] It was suggested that protein BBL folds in a downhill manner with a low cooperativity, while protein E3BD folds cooperatively. These two kinds of folding behaviour are all considered as the typical behaviour from the theory of funnel-like landscapes.[5,10−12]

To understand such differences for proteins with similar structures, we report a study on proteins BBL and E3BD using molecular dynamical simulations based on a simplified Gō-like model. It is found that protein BBL folds in a downhill manner indeed without any rate-limiting barriers in the free-energy landscape. The distribution of the formed fraction of native contacts in equilibrium is not bimodal, even at its folding transition temperature. Differently, protein E3BD folds cooperatively and a weak bimodal distribution of the formed fraction of native contacts is observed. That is to say, protein E3BD folds in a weak two-state manner, instead of in a downhill one. A statistics on the native structures of various proteins have been carried out. The results imply in some degree that there is a transition following the change of native structures of proteins, and the difference between proteins BBL and E3BD indicates the existence of this transition. Such an observation suggests a method to identify the downhill folders.

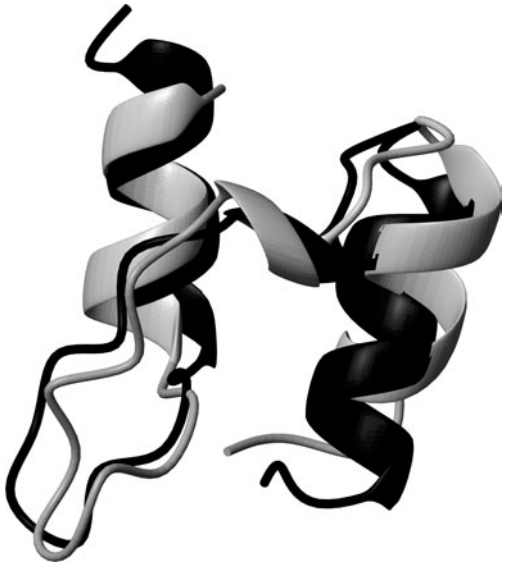

Fig. 1. Superimposition of the native structures of proteins BBL and E3BD in dark grey and light grey cartoons, respectively. The overall fold is a three-helix bundle with two main helices at the N and C termini. These helices are joined by a flexible loop that contains a 3_{10} helix.

The model used in this work is an off-lattice

* Supported by the National Natural Science Foundation of China under Grant Nos 90103031, 90403120, 10021001, 10474041, and 10204013, the Nonlinear Project of the National Key Basic Research Special Foundation of China, and the FANEDD of the Ministry of Education of China.

** To whom correspondence should be addressed. Email: wangwei@nju.edu.cn

Gō-like model which describes the protein as connected beads with Gō-like potentials.[13] Such a model has been widely used in the study on the protein folding.[14−19] A native contact between two residues is defined if the distance between any pair of non-hydrogen atoms of these two residues is less than 5 Å in the native state. The folding dynamics is simulated by the Langevin dynamics. The random coil conformation is used as the initial conformation and the leap-frog algorithm is used to integrate the equation of the Langevin dynamics. The thermodynamic quantities are calculated by the WHAM algorithm.[14,20,21] Detailed descriptions, such as the various potentials and parameters, can be found in Refs. [14,15].

Based on the Gō-like model, the folding of the two proteins shown in Fig. 1 are simulated. The folding behaviour can be characterized by their thermodynamics. The barrier-less feature of the downhill folding indicates that there is no barrier between the folded state and the unfolded states, i.e., there is no clear separation between the native state and the denatured states. Therefore, along the reaction coordinate, the population of native state and that of denatured states mix with each other. That is, there would be a broad distribution with a single peak. However, for the two-state folders, the free-energy barrier would result in a bimodal distribution with two peaks corresponding to ordered native state and disordered denatured states, respectively. In our simulations, the probabilities of the states along reaction coordinate Q as a function of temperature are shown in Fig. 2. The insets show the snapshots at the folding transition temperature T_f. According to Fig. 2(a), the probability distribution of protein BBL displays a broad ensemble of states only with a maximum that progressively and continuously slides from a high Q value to a low value as temperature increases. Interestingly, it is found that every conformation can be significant populated at some temperature. This is in contrast to the folding of protein E3BD, where only two sets of conformations can be populated, i.e. $Q \approx 1.0$ conformations at low temperature and $Q \approx 0.2$ conformations at high temperature (see Fig. 2(b)). At the same time, the intermediate region between these two states, e.g. at about $Q = 0.6$, always has little population at any temperature. This implies that the system switches between these two conformational sets of distinct states, which is a typical character of the cooperative folding. Particularly, the difference between two proteins is more remarkable in the distribution of Q values at the transition temperature T_f (insets in Fig. 2), where the distribution for protein BBL is not bimodal, while that for protein E3BD is bimodal. Note that the free-energy barrier separating the unfolding and native states for protein E3BD is not as large as the typical two-state folder, such as Chymotrypsin Inhibitor 2 (CI2). This indicates that the folding of protein E3BD is of weak two-state, which may be attributed to its native topology. As a result, different shapes of the distributions illustrate the difference in the folding manners for these two proteins, namely the downhill manner for protein BBL and the weakly two-state one for protein E3BD. Thus, our simulations support the idea that these homologous proteins fold with different behaviours.

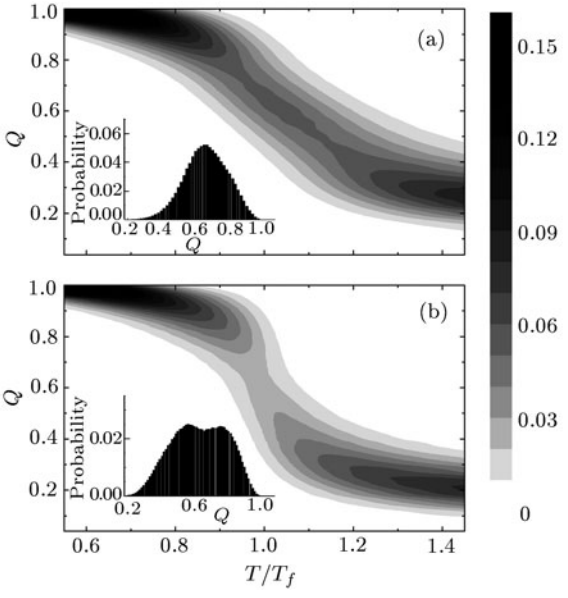

Fig. 2. Distribution of equilibrium probabilities of the Q values as a function of temperature for protein BBL (a) and protein E3BD (b), respectively. The grey level indicates the values of the probability. The darker the colour, the higher the probability is. The insets show the distribution of equilibrium probability of Q for proteins BBL (a) and E3BD (b) at the folding temperature T_f. Here T_f for proteins E3BD and BBL are 0.95 and 0.86, respectively.

More detailedly, the temperature-dependent specific heat and the ensemble average of the fraction of formed native contact, Q, are shown in Fig. 3. The peak height of the specific heat for protein BBL is only half the protein E3BD, while its peak is much wide (Fig. 3(a)). This indicates that the folding cooperativity of protein BBL is weaker in comparison with that of its homologous protein. Similarly, Figure 3(b) shows that the average of Q for protein BBL exhibits a boarder transition region than that of protein E3BD. These results are consistent with the observations in experiments,[8,9] and further illustrate the differences between the folding of these proteins.

The difference in the folding behaviour implies a large change in the shape of free-energy landscape. What produces the difference between the homologous proteins BBL and E3BD? To identify the physical origin of the difference may help us to understand the factors governing the free-energy landscapes. Here, similar to the starting point of the Gō-like models, we

analyse the native structures of proteins. The native structure of a protein can be described by a contact map which is a widely used method for characterizing protein structures. Let us consider the sequentially long-range (nonlocal) contacts which make important contributions to the cooperativity of proteins.[22,23] Practically, the nonlocal contacts are defined as those with sequential distance larger than $L_c^c = 6$ residues. The nonlocal contacts of these two proteins are aligned as shown in Fig. 4. The sequences are divided into four pieces according to their secondary structures, such as the N-helix, loop-helix, loop, and C-helix. It is found that there are more nonlocal contacts in protein E3BD, and these additional nonlocal contacts are mainly distributed in the regions between N-helix and loop, loop-helix and C-helix, and the loop itself. This suggests qualitatively that the difference in the folding behaviour might come from the difference in the number of nonlocal contacts, resulting from the structural features of both proteins.

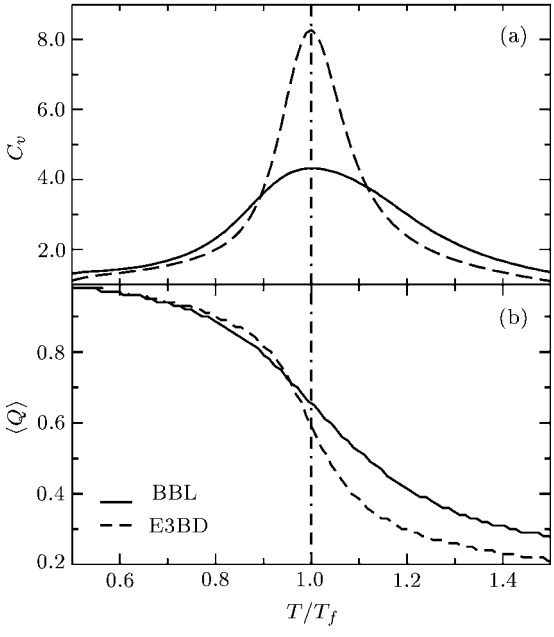

Fig. 3. Thermodynamical properties for protein BBL (solid lines) and E3BD (dashed lines). (a) Specific heat as a function of temperature. (b) Temperature dependence of average formed fraction of native contact Q. The vertical dotted line indicates the folding temperature T_f.

Based on the native structures of these two proteins, we have made a calculation of the number of nonlocal contacts per residue N_N which is taken as a structural indicator of the proteins. The factor N_N is defined as $N_{\rm nonlocal}/L$, where $N_{\rm nonlocal}$ and L are the total number of nonlocal contacts and the length of protein chain, respectively. Note that L is used as the denominator in the definition to eliminate the effects of protein chain length. As shown in Fig. 4, there are 40 nonlocal contacts for E3BD and 27 for the BBL, and the lengths of these two protein are 41 and 37, respectively. Therefore, the protein E3BD has its $N_N = 0.976$ and the protein BBL has its $N_N = 0.730$. In addition, a statistics on other two-state folding proteins shows a good correlation between the parameter N_N and the height of the free-energy barriers.[25] The interpolation of N_N with null height of the barrier is about $N_N = N_N^c \simeq 0.9$. This implies that there is a transition of the free-energy landscapes with barrier to without barrier. As a special case, the homologous proteins BBL and E3BD lie at different sides of the transition. Since there has been just one downhill folder found up to now, the determination for the transition point is not precise. Yet, there should be a sudden change (or at least a crossover) of cooperativity at a certain value of N_N, which separates the importance of interaction of the local and nonlocal contacts. This is consistent with the idea of the importance of the native topology on protein folding.[6] In addition, our statistics provides a method to identify the downhill folder from the native topology of proteins only, which is much easier than regular experimental methods. However, such a prediction deserves further testing.

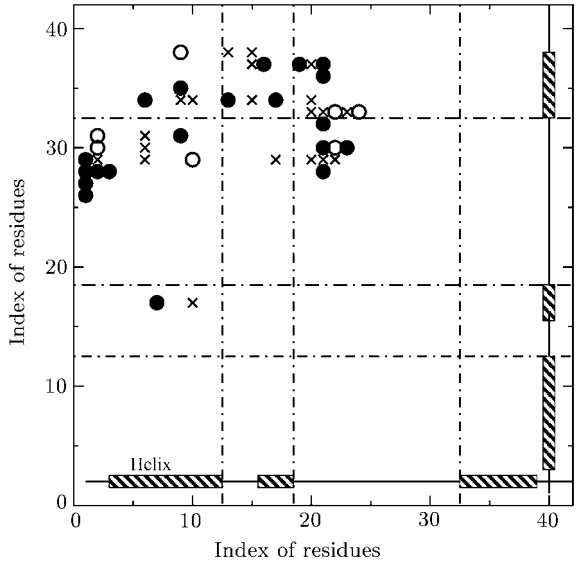

Fig. 4. Nonlocal contacts alignment of protein BBL and E3BD. Here •, ○ and × denote the nonlocal contacts appeared only in protein E3BD, only in protein BBL and in both proteins, respectively. The dotted lines divide the sequence into four pieces. It is obvious that the number of nonlocal contacts of E3BD is more than that of BBL at each region, especially in the region between N-helix and loop, the region between loop-helix and C-helix, and the loop region. There are 4, 3 and 3 extra nonlocal contacts for E3BD than BBL in these regions.

In conclusion, the folding behaviour of protein BBL, and of its homologous protein E3BD, are studied using the Gō-like model. Although proteins BBL and E3BD are homologous proteins and very similar

in both sequence and structure, the folding behaviour of these two proteins are different. Protein BBL shows downhill folding behaviour, while protein E3BD shows weak two-state folding behaviour and folds more cooperatively. By analysing the native structures of these two proteins, the origin of the difference in the folding behaviour is attributed to the different number of the nonlocal native contacts. There should be a critical value around $N_N^c = 0.9$, which separates the downhill folders and two-state folders. The present results are clearly related to the experimental findings.

References

[1] Anfinsen C B 1973 *Science* **181** 223
[2] Fersht A R 1995 *Curr. Opin. Struct. Biol.* **5** 79
[3] Brooks C L 1998 *Curr. Opin. Struct. Biol.* **8** 222
[4] Mirny L and Shakhnovich E 2001 *Ann. Rev. Biophys. Biomol. Struct.* **30** 361
[5] Onuchic J N, Luthey-Schulten Z and Wolynes P G 1997 *Ann. Rev. Phys. Chem.* **48** 545
[6] Baker D 2000 *Nature* **405** 39
[7] Gianni S et al 2003 *Proc. Natl. Acad. Sci.* USA **100** 13286
[8] Garcia-Mira M M et al 2002 *Science* **298** 2191
[9] Ferguson N et al 2004 *J. Mol. Biol.* **344** 295
[10] Socci N D, Onuchic J N and Wolynes P G 1996 *J. Chem. Phys.* **104** 5860
[11] Bryngelson J D et al 1995 *Proteins* **21** 167
[12] Eaton W A et al 2000 *Ann. Rev. Biophys. Biomol. Struct.* **29** 327
[13] Ueda Y, Taketomi H and Gō N 1975 *Int. J. Pept. Res.* **7** 445
[14] Clementi C, Nymeyer H and Onuchic J N 1991 *J. Mol. Biol.* **298** 937
[15] Kaya H and Chan H S 2003 *J. Mol. Biol.* **326** 911
[16] Koga N and Takada S 2001 *J. Mol. Biol.* **313** 171
[17] Hoang T X and Cleplak M 2000 *J. Chem. Phys.* **112** 6851
[18] Xu W X, Wang J and Wang W 2005 *Chin. Phys. Lett.* **22** 258
[19] Qin M, Wang J and Wang W 2003 *Chin. Phys. Lett.* **20** 1883
[20] Ferrenberg A M and Swendsen R H 1989 *Phys. Rev. Lett.* **63** 1195
[21] Shea J E and BrooksIII C L 2001 *Ann. Rev. Phys. Chem.* **52** 499
[22] Abkevich V I, Gutin A M and Shakhnovich E I 1995 *J. Mol. Biol.* **252** 460
[23] Chan H S 1998 *Nature* **392** 761
[24] Jackson S E 1998 *Fold. Des.* **3** R81
[25] Zuo G H, Wang J and Wang W (in preparation)