Model study of prionlike folding behavior in aggregated proteins

Yong-Yun Ji,[1] You-Quan Li,[1] Jun-Wen Mao,[1,2] and Xiao-Wei Tang[1]
[1]*Department of Physics, Zhejiang University, Hangzhou 310027, People's Republic of China*
[2]*Department of Physics, Huzhou Teachers College, Huzhou 313000, People's Republic of China*
(Received 9 June 2005; published 12 October 2005)

We investigate the folding behavior of protein sequences by numerically studying all sequences with a maximally compact lattice model through exhaustive enumeration. We get the prionlike behavior of protein folding. Individual proteins remaining stable in the isolated native state may change their conformations when they aggregate. We observe the folding properties as the interfacial interaction strength changes and find that the strength must be strong enough before the propagation of the most stable structures happens.

DOI: 10.1103/PhysRevE.72.041912 PACS number(s): 87.10.+e, 87.14.Ee, 87.15.−v

I. INTRODUCTION

The biological function of the protein is tightly related to its conformation. The loss of biological activity of the proteins, such as in the case of insoluble protein plaques consisting of amyloid fibrils in organs [1,2], may arise from the aggregation of misfolded protein which frequently causes various diseases, such as prion diseases, Alzheimer's disease, and Parkinson's disease [3]. For the prion diseases, the prion protein is regarded as the origin of some brain-attacking diseases which are known as spongiform encephalopathies. The structures of the normal form of prion protein (PrP^C) have been obtained [4–6]. Both the PrP^C and the corresponding misfolded form (PrP^{Sc}) have identical sequences. The only difference is in their conformations which are considered to be responsible for the aggregation and disease [7]. Current experiment methodologies encounter difficulties in obtaining atomic details of seed formation and conversion from PrP^C to PrP^{Sc} [8]. Large challenges are still in existence for people to theoretically investigate the mechanism of seed formation and propagation at atomic level.

The importance of protein folding has been recognized for a long time [9–11]. It is widely believed that for most single domain proteins, the native structure is the global free-energy minimum state, and the amino-acid sequence alone encodes sufficient information to determine its 3-D structure [11]. A great deal of research has been performed to get the general properties of protein folding and interpret the basis of misfolding diseases. Li *et al.* [12] presented a meaningful interpretation on the structure selection of nature proteins by proposing a concept of designability on the basis of the *HP* lattice model proposed by Dill [13], where *H* and *P* stand for hydrophobic and polar amino acid. Protein refolding to an alternative form has been observed by Harrison *et al.* [14] using the lattice model in a propagatable manner. Under conditions where the normal native state is marginally stable or unstable, two chains refold from monomeric minimum (the normal native state) to an alternative multimeric minimum in energy, comprising a single refolded conformation that can propagate itself to other protein chains. Harrison *et al.* treated both 2-D and 3-D *HP* models to investigate the dimer formation and found that the structures in the homodimeric native state rearrange so that they are very different in conformation from those at the monomeric native state [15]. Giugliarelli *et al.* showed how the average inter-amino-acid interaction affects the properties of both single and interacting proteins in a highly ordered aggregation, but only in 2-D lattice model. They found the propensity to structural changes of aggregated protein, namely, the prionlike behavior of protein [16]. A Monte Carlo simulation has been used by Bratko and Blanch to examine the competition between intramolecular interactions which is responsible for the native protein structure, and intermolecular association which results in the aggregation of misfolded chains [17]. These works enhanced our understanding on the diseases caused by protein misfolding and aggregating. However, the folding behavior of all sequences is mostly studied within two-dimensional models.

As we have known that the lattice model used to be a valuable model for the designability, it is therefore interesting to investigate the behavior of protein misfolding and aggregating together with the changes of designability and stability on the basis of lattice model. In this paper we apply the 3-D *HP* model to elucidate the mechanism of protein aggregation diseases such as prion diseases. In Sec. II, we introduce our model that consists of a stack of 27 toy bricks, each of them standing for a $3\times3\times3$ cubic lattice. This imitates 27 aggregated model proteins interacting with each other through interface. In Sec. III, we discuss the effect of interaction between model proteins on the structural stability. Our results show the conversion of most stable structures between isolated state and aggregated state. The change of most stable structures is strongly correlated with the strength of the interfacial interaction.

II. "TOY BRICK" MODEL

Both two- and three-dimensional lattice models which describe the protein folding into the native structure as free energy minimization are NP hard [18]. It is significant to introduce simplified models to achieve some essential properties of protein folding. The *HP* lattice model introduced first by Dill [13] has helped in understanding essential properties of protein folding and evolution. That model and its extended ones are widely used up to date [19–21], and we have investigated the medium effects on the selection of se-

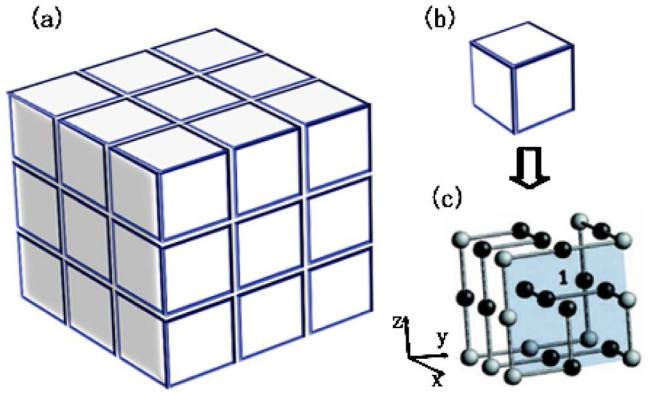

FIG. 1. (Color online) The multimer stacked with 27 polymers (a) and an example structure of single polymer (b) and (c) with hydrophobic residues (gray) and polar residues (black).

quences folding into stable proteins with the simple *HP* model and got some meaningful results [22].

In order to investigate the folding behavior of aggregated protein (multimer), we reconstruct the original *HP* model by putting a number of individual model proteins (polymers) together, namely, building the cubic toy bricks layer by layer in order. Each polymer has identical sequence and structure but is in different orientation. In this paper, we study the multimer which is stacked by $3\times3\times3$ ordered polymers [Fig. 1(a)]. As is shown in Fig. 1(c), each polymer is figured as a cube formed by a chain of 27 beads occupying the discrete sites of a lattice in a self-avoiding way, with two types of beads of polar (P) and hydrophobic (H) amino acids respectively. These polymers interact with each other through the contact amino acids at the surfaces of adjacent polymers. In this model, the total energy of the multimer includes the pair contact energy of the amino acids inside each polymer and the interfacial potential (the additional energy) caused by the contact of amino acids from two adjacent polymers. In our case, 54 interfaces between $3\times3\times3$ ordered polymers should be considered. Thus, the energy of the multimer is given by

$$H = 27 \times \sum_{i<j} E_{\sigma_i\sigma_j}\delta_{|r_i-r_j|,1}(1-\delta_{|i-j|,1}) + \alpha \sum_{\{f,f'\}} \varepsilon_{ff'}, \quad (1)$$

where the former part represents the total energy of 27 isolated polymers, i,j denote the successive labels of residues in a sequence, r_i is for the position (of the ith residue) on the lattice sites, and σ_i is for $H(P)$ corresponding to hydrophobic (polar) residue, respectively. Here the delta notation is adopted, i.e., $\delta_{a,b}=1$ if $a=b$ and $\delta_{a,b}=0$ if $a\neq b$. As the hydrophobic force drives the protein to fold into a compact shape with as many hydrophobic residues inside as possible [23], the *H-H* contacts are more favored in this model, which can be characterized by choosing $E_{PP}=0$, $E_{HP}=-1$, and $E_{HH}=-2.3$ as adopted in Ref. [12]. The second term in Eq. (1) is introduced to describe the interfacial energy caused by the contact of polymers. The sum will be done over all the 54 pairs of surfaces. The α denotes for the strength of the interfacial interaction and it ranges from 0.1 to 0.9. The $\varepsilon_{ff'}$ is the pair interfacial energy of two contact faces f and f' which belong to two adjacent polymers, respectively.

For calculation convenience, we label the six faces of the polymer with $1,2,3,\bar{1},\bar{2},\bar{3}$ for a given structure, of which the normal directions correspond to the six directions along $\mathbf{x},\mathbf{y},\mathbf{z},-\mathbf{x},-\mathbf{y}$, and $-\mathbf{z}$, respectively [Fig. 1(c)]. For each face there are still four states related by rotation of $\pi/2$ that are specified by second label $0,\pi/2,\pi$, and $3\pi/2$. Then we can easily denote each face of the polymer with brackets: $|\mu,\theta\rangle$, where μ is the face label and θ is the rotation label. When the structure and sequence of the polymer are fixed, we can write down 24 pairs of kets and bras, $|\mu,\theta\rangle,\langle\mu,\theta|$. The interfacial energy between polymers can be represented as $\varepsilon_{ff'}=\langle f|f'\rangle=\langle\mu,\theta|\mu',\theta'\rangle$; it denotes the contact energy between the θ-th rotated state of the μ-th face of a polymer and the θ'-th rotated state of the μ'-th face of the adjacent polymer, and the f stands for the θ-th rotated state of μ-th face. The second term of energy in Eq. (1) is then given by

$$\alpha \sum_{\{f,f'\}} \varepsilon_{ff'} = \alpha \sum_{\{f,f'\}} \langle f|f'\rangle = \alpha \sum_{\{\mu,\theta,\mu',\theta'\}} \langle\mu,\theta|\mu',\theta'\rangle. \quad (2)$$

Both $\langle\mu,\theta|$ and $|\mu',\theta'\rangle$ can be specified by matrices with entities being either H or P. In terms of these matrixes, we can easily get the contact energy of two faces by accounting for the pairs from the corresponding entities, namely,

$$\langle\mu,\theta|\mu',\theta'\rangle = \sum_{k=1}^{3}\sum_{l=1}^{3} E_{\sigma_{M_{kl}},\sigma_{N_{kl}}}. \quad (3)$$

Here M and N denote the the aforementioned matrices corresponding to $\langle\mu,\theta|$ and $|\mu',\theta'\rangle$. To avoid ambiguity in defining the matrix entities related to $\langle\mu,\theta|$ and $|\mu,\theta\rangle$, we make the following point. The matrix is determined by the face of the cubic by taking the right hand and head along positive directions of two axes, respectively. An additional point is that we look toward the inside of the cubic to define a ket $|\,\rangle$ and toward the outside to define a bra $\langle\,|$. For example, in the $3\times3\times3$ case as shown in Fig. 1(c), the 1-th face contact with $\bar{1}$-th face of another polymer, we write down the matrices related to $|1,0\rangle$ and $\langle\bar{1},0|$ as

$$N_{|1,0\rangle} = \begin{bmatrix} H & P & H \\ H & P & H \\ H & P & H \end{bmatrix}, \quad M_{\langle\bar{1},0|} = \begin{bmatrix} H & P & H \\ P & P & P \\ H & H & H \end{bmatrix}. \quad (4)$$

Then the contact energy of these two faces is evaluated as

$$\begin{aligned}\langle\bar{1},0|1,0\rangle &= \sum_{k=1}^{3}\sum_{l=1}^{3} E_{\sigma_{M_{kl}},\sigma_{N_{kl}}} \\ &= E_{HH}+E_{PP}+E_{HH}+E_{PH}+E_{PP}+E_{PH}+E_{HH} \\ &\quad + E_{HP}+E_{HH}. \end{aligned} \quad (5)$$

Clearly, $\langle\mu,\theta|\overline{\mu',\theta}\rangle=\langle\overline{\mu',\theta}|\mu,\theta\rangle$ in our definition. To minimize the total energy of the multimer, the optimal configuration is adjusted by rotating the bra $\langle\overline{\mu',\theta}|$ to an appropriate $\langle\overline{\mu',\theta'}|$ such that the inner product $\langle\overline{\mu',\theta'}|\mu,\theta\rangle$ takes the smallest magnitude. Clearly, the system for $1\times1\times1$ poly-

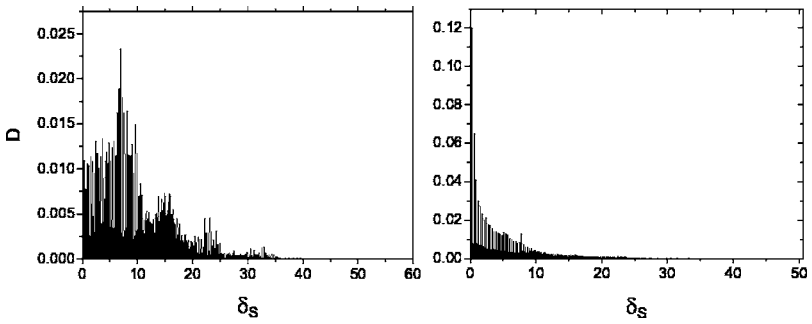

FIG. 2. The distribution of sequences (D) versus the energy gap δ_s for normal sequences (left) and prionlike sequences (right).

mers corresponds to an Ising model, while that for $2\times2\times2$ or larger ones corresponds to the 24-state potts model.

III. RESULTS AND DISCUSSION

It has been noticed by Li *et al.* [12] that some structures can be designed by a large number of sequences, while some can be designed by only a few sequences. To elucidate this difference, they introduced the designability of a structure which is measured by the number of sequences that take this structure as their unique lowest energy state. In conclusion, structures differ drastically according to their designability, i.e., highly designable structures emerge with a number of associated sequences much larger than the average ones. Additionally, the energy gap represents the minimum energy a particular sequence needs to change from its ground-state structure into an alternative compact structure. And the average energy gap for a given structure is evaluated by averaging the gaps over all the sequences which design that structure. In the single-polymer *HP* model, the structures with large designability have much larger average gaps than those with small designability, and there is an apparent jump around $N_s=1400$ in the average energy gap. This feature was first noticed by Li *et al.* [12], thus these highly designable structures are thermodynamically more stable and possess proteinlike secondary structures into which the sequences fold faster than the other structures.

In our model, the designability and the energy gap for the multimer are similarly defined, with the energy of the single polymer replaced by the total energy of the multimer. Considering that the different orientation of the polymers contributes differently to the total energy of the multimer in spite of the identical sequence and structure of each polymer, the designability (N_s) in this case denotes the number of sequences that take the individual structure as the unique lowest energy state of the multimer. Similarly, the energy gap δ_s is the minimum energy for the multimer to covert from its ground-state conformation (including 27 particularly oriented polymers) into an alternative form.

In our simulation, we search the maximal compact cubic structure of the polymer with various designability first. Based on this, we calculate the total energy of the multimer by stacking 27 polymers one by one to form the multimer. At each step the state with minimal energy will be preserved. After all polymers are on their positions, we regulate the orientation of each polymer again to search the minimal energy of multimer. After the exhaustive enumeration of all possible sequences, we find that many structures with high average gap in the isolated case are no more highly designable in aggregated case when $\alpha=1$. There are 2 197 634 prionlike sequences taking one structure as the native state (unique energy minimum) in the isolated case, but taking another structure when aggregated, while 1 905 960 sequences take the same structures as their native states in both isolated and aggregated cases.

By investigating the sequences of these two cases, we find that the normal sequences possess much more proportion of sequences with larger energy gap, though there are more prionlike sequences. Figure 2 shows the percentage of total sequences versus the energy gap. The distribution of normal sequences (left) has a peak at $\delta_s=6.9$; on the right of the figure, plotted for prion-like sequences, the distribution decreases monotonously with the increase of δ_s. In this simplified model, the energy gap is considered to be a good parameter to indicate the stability of a sequence at its native state. The larger the energy gap is, the more stability the structures

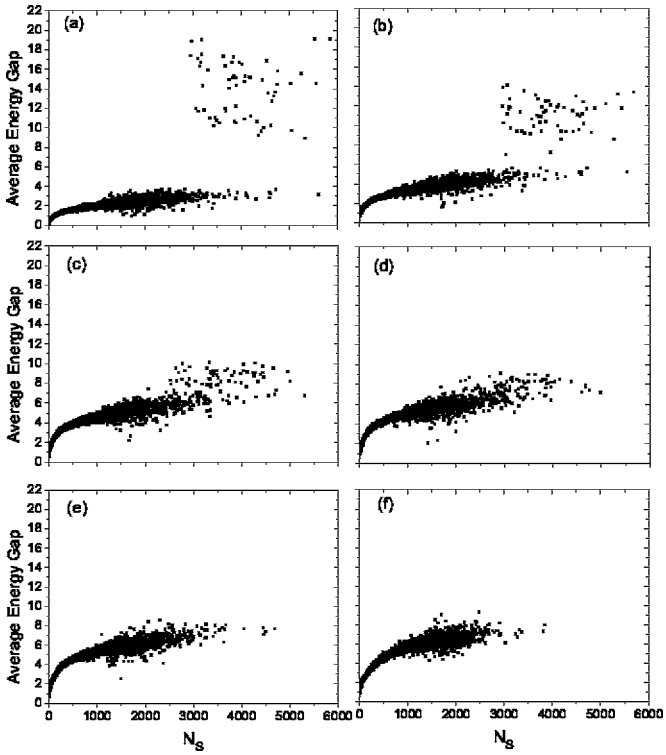

FIG. 3. The average energy gap versus the N_s for different interfacial interaction strength: (a) $\alpha=0.1$, (b) $\alpha=0.3$, (c) $\alpha=0.5$, (d) $\alpha=0.6$, (e) $\alpha=0.7$, and (f) $\alpha=0.9$.

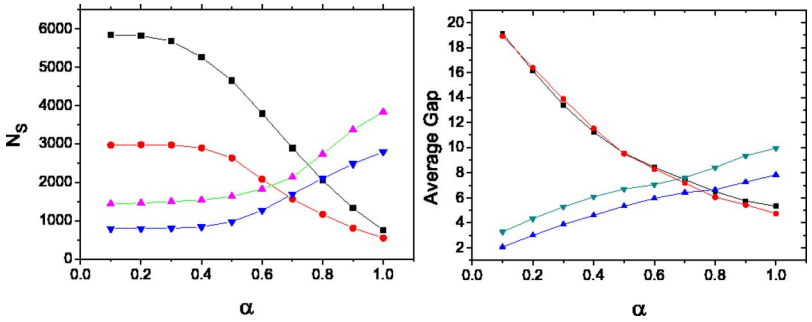

FIG. 4. (Color online) The N_s (left) and average gap (right) versus α for four particular structures: the highest designable structure in isolated (square) and in aggregated (up triangle) case, the largest average gap one in isolated (circle) and in aggregated (down triangle) case.

have. Consequently, minority sequences are prionlike. Actually, in nature there should be few prionlike sequences and most protein sequences are stable in their native states. We analyze some of the prionlike sequences by assuming that the decrease of energy can drive the polymers in a multimer to refold to new lower energy states. When the isolated native polymers aggregate to form a multimer, some replacement of isolated native structure by new Prp^{Sc}-like structure will reduce the energy of the multimer. In this case, all polymers in the multimer will prefer to change their structures for staying much lower total energy level.

In our simulation, we obtain that the average gaps of the structures with high designability diminish with increasing α from 0.1 to 0.9, while the average gaps of most structures with low designability increase. Additionally, a number of new structures, which no sequences take as native state in an isolated case, emerge. In isolated case, there is an abrupt jump in the average energy gap which separates the structures into two groups, the highly designable and lowly designable. These two groups are mixed around $\alpha=0.6$ in aggregated case (Fig. 3). This implies that the stable structures in the isolated case may become unstable and the new stable structures emerge only when the strength of interaction is large enough. This may be consistent with the result in Ref. [24] that most misfolding diseases have a broad incubation period before they cause symptoms and some patients will not be injured. After the seed formation of the misfolding protein, it maybe still take some time for misfolded proteins to get strong enough as condition changes. If the strength stays below the transition point, the injury will not be induced. We investigate some particular structures. The designability and average gaps of highest designable and largest average gap structures in the isolated case diminish continuously with increasing α, but some other structures (lowly designable in isolated case) enhance their designability and average gaps when aggregate (Fig. 4).

We compare the change of average gap versus the designability in isolated and aggregated cases, respectively. In the latter case with $\alpha=1$, our simulation shows that the average gap increases almost continuously with the increasing of N_s, and there is no abrupt jump in average gap (Fig. 5). Thus the structures cannot be distinguished by the designability and the average gap obviously. The largest average gap reaches 9.945, which is much larger than that in isolated case. Considering an individual sequence, there are two sequences whose gaps reach the value of 55.4 in the aggregated case (2.6 in the isolated case). There are no sequences with energy gap larger than 70.2, the interfacial energy counterpart of these gaps. As is shown in Ref. [12], there are 60 highly designable structures which are distinguished by large average gap from other ones in the isolated case. When the proteins aggregate, the average gap increases with designability N_s continuously. But there are large differences in the structures. We find the structures with largest average gap or the highest designability are no longer the ones in the isolated case, which implies that the most stable structures change when the proteins aggregate. This is similar to the situation that prionlike proteins differ in their conformations, PrP^{C} and PrP^{Sc}, in isolated and aggregated cases, respectively [14], which is thought to deduce the diseases.

When $\alpha=1$, 36.7 percent of sequences take one structure as their unique ground states, which is much larger than that in the isolated case (4.75%). In the aggregated case, the highest designability N_s is of 3831. Hence there must be much more sequences which average over the other structures and make some lowly designable structures possessing more sequences. As is shown in Fig. 6, in the aggregated (left) case, there are many more structures with large N_s, i.e., there are 1497 structures whose N_s are larger than 1400. In the situation of $\alpha=1$, the distribution of structures with N_s will reach a maximum when $N_s=113$ (Fig. 6), and it is noticeable that the distributions are the same when we change the α from

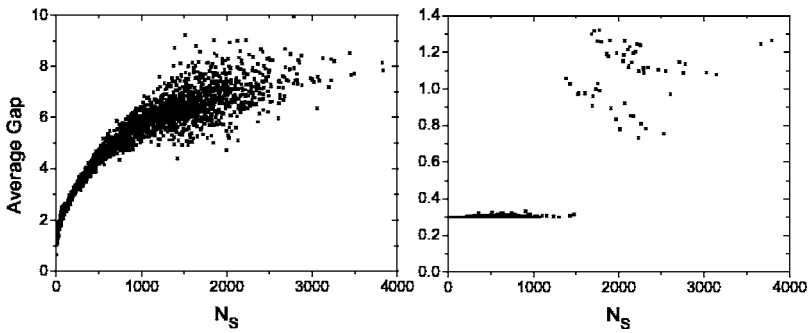

FIG. 5. The average energy gap versus the N_s: 27 polymers, aggregated case ($\alpha=1$) (left) and isolated polymer case (right).

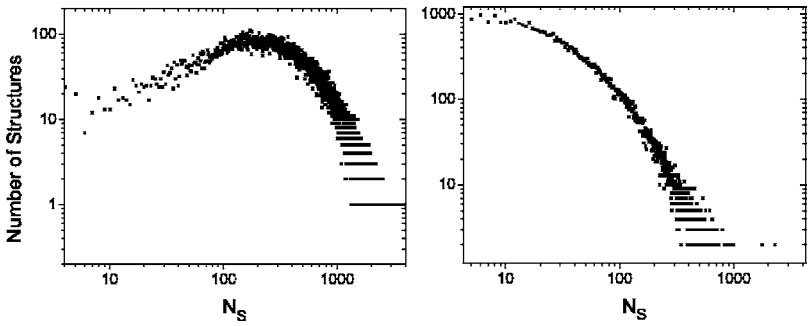

FIG. 6. The number of structures versus the N_s in the situation: 27 polymers, aggregated($\alpha=1$)(left) and single polymer (right).

0.000 001 to 1. In comparison with the isolated case, there are many fewer structures that take for individual small designability: in the aggregated case the largest number of structures gets 111 at $N_s=113$, which is 1109 with N_s being 2 in the isolated case. The sequences are much more averaged in all structures in the aggregated case, but there are still some structures which are occupied by large number of sequences, and particularly with large average gap. The average gap changes from 0.65 ($N_s=1$) to 9.95 ($N_s=2785$), and the largest designability N_s is 3831 with average gap being 7.85.

In summary, by 3-D *HP* lattice model we observed that there are some sequences which do have prionlike behavior in aggregated proteins. It has been known that the thermodynamic stability of proteins in solution is affected by a variety of factors including temperature, hydrostatic pressure, and presence of additives, such as salts and cosolvent species, which implies that the interfacial interactions between isolated proteins are ubiquitous. Many different environments likely influence the prion's ability to misfold and aggregate. Based on the experiments on human prion protein, Swietnick *et al.* reported that the synthetic PrP^C can be converted to PrP^{Sc} in vitro as the *p*H changes [25]. At low *p*H, PrP^C gains β-like extended structure and tends to aggregate. In support of this result, similar experiments on mouse prion protein [26] also demonstrate a conformational transition at low *p*H. Though there is strong evidence supporting for the protein-only hypothesis of prion propagation [27], the underlying mechanism of conversion of PrP^C to PrP^{Sc} remains elusive. On the basis of our simulation, we hypothesize that the environments affect the interaction between prion proteins (e.g., parameter α in our calculation), and hence facilitate or induce the conversion. If the concentration of protein is sufficiently high and some other conditions are changed, the isolated proteins have a tendency to aggregate reaching a lower energy state. Our calculation showed that a few proteins will propagate the aggregated normal form to abnormal conformation to get more stable multimer (with lower energy), namely, the prionlike behavior. Particularly, we found that the most stable structures are no longer the ones in the isolated case when the proteins aggregate, such as the structures with largest average gap or the highest designability. Furthermore, we obtained that the average gaps of the structures with high designability diminish with increasing α, while the average gaps of most structures with low designability increase. Thus there is no obvious jump in average gap between lowly and highly designable structure for the aggregated proteins. We expect this result can give some hints on the study of misfolding diseases. Since this is a simplified model, further understanding about protein aggregation and misfolding diseases is expected. A more realistic proteinlike model with the various interactions being close to the real proteins is in progress. In our simulation, for the feasibility of exhaustive enumeration, we only search the maximally compact structures. However, Schiemann *et al.* [28] reported that the proteins did not necessarily prefer an overall compact shape, and the hydrophobic core is maximally compact in many ground-state structures, but not the whole conformation for the simple peptide model. It is likely that the study on the structures beyond maximally compact will give more information in the future.

ACKNOWLEDGMENTS

This work is supported by NSFC No. 10225419 and the Key Project of Chinese Ministry of Education No. 02046.

[1] G. Glenner, N. Engl. J. Med. **302**, 1283 (1980).
[2] C. C. F. Blake and L. C. Serpell, Structure (London) **4**, 989 (1996).
[3] E. McKintosh, S. J. Tabrizi, and J. Collinge, J. Neurovirol **9**, 183 (2003).
[4] R. Riek, S. Hornemann, G. Wider, M. Billeter, R. Glockshuber, and K. Wuthrich, Nature (London) **382**, 180 (1996).
[5] T. L. James, H. Liu, N. B. Ulyanov, and S. Farr-Jones *et al.*, Proc. Natl. Acad. Sci. U.S.A. **94**, 10086 (1997).
[6] S. B. Prusiner, Proc. Natl. Acad. Sci. U.S.A. **95**, 13363 (1998).
[7] C. A. Ross and M. A. Poirier, Nat. Med. **10**, S10 (2004).
[8] M. Buyong and N. Ruth, Protein Sci. **11**, 2335 (2002).
[9] L. Pauling and R. B. Corey, Proc. Natl. Acad. Sci. U.S.A. **37**, 235 (1951).
[10] L. Pauling and R. B. Corey, Proc. Natl. Acad. Sci. U.S.A. **37**, 251 (1951).
[11] C. Anfinsen, Science **181**, 223 (1973).
[12] H. Li, R. Helling, C. Tang, and N. S. Wingreen, Science **273**, 666 (1996).
[13] K. A. Dill, Biochemistry **24**, 1501 (1985).
[14] P. M. Harrison, H. S. Chan, S. B. Prusiner, and F. E. Cohen, Protein Sci. **10**, 819 (2001).

[15] P. M. Harrison, H. S. Chan, S. B. Prusiner, and F. E. Cohen, J. Mol. Biol. **286**, 593 (1999).
[16] G. Giugliarelli, C. Micheletti, J. M. Banavar, and A. Maritan, J. Chem. Phys. **113**, 5072 (2000).
[17] D. Bratko and H. W. Blanch, J. Chem. Phys. **114**, 561 (2001).
[18] A. S. Fraenkel, Bull. Math. Biol. **55**, 1199 (1993).
[19] J. R. Banavar, M. Cieplak, and A. Maritan, Phys. Rev. Lett. **93**, 238101 (2004).
[20] G. Tiana, B. E. Shakhnovich, N. V. Dokholyan, and E. I. Shakhnovich, Proc. Natl. Acad. Sci. U.S.A. **101**, 2846 (2004).
[21] G. Salvi and P. De Los Rios, Phys. Rev. Lett. **91**, 258102 (2003).
[22] Y. Q. Li, Y. Y. Ji, J. W. Mao, and X. W. Tang, Phys. Rev. E **72**, 021904 (2005).
[23] W. Kauzmann, Adv. Protein Chem. **14**, 1 (1959).
[24] D. N. Irani, M. Otto, and T. Weber, Curr. Treat. Options in Infect. Dis. **5**, 477 (2003).
[25] W. Swietnick, R. Peterson, P. Gambetti, and W. K. Surewicz, J. Biol. Chem. **272**, 27517 (1997).
[26] S. Hornemann, R. Glockshuber, Proc. Natl. Acad. Sci. U.S.A. **95**, 6010 (1998).
[27] C. Soto and J. Castilla, Nat. Med. **10**, S63 (2004).
[28] R. Schiemann, M. Bachmann, and W. Janke, J. Chem. Phys. **122**, 114705 (2005).