Evolving model of amino acid networks

Shan Chang, Xiong Jiao, Xin-qi Gong, Chun-hua Li, Wei-zu Chen, and Cun-xin Wang*
College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100022, China
(Received 28 November 2007; revised manuscript received 17 May 2008; published 24 June 2008)

The three-dimensional structure of a protein can be treated as a complex network composed of amino acids, and the network properties can help us to understand the relationship between structure and function. Since the amino acid network of a protein is formed in the process of protein folding, it is difficult for general network models to explain its evolving mechanism. Based on the perspective of protein folding, we propose an evolving model for amino acid networks. In our model, the evolution starts from the amino acid sequence of a native protein and it is guided by two generic assumptions: i.e., the neighbor preferential rule and the energy preferential rule. We find that the neighbor preferential rule predominates the general network properties and the energy preferential rule predominates the specific biological structure characteristics. Applied to native proteins, our model mimics the features of amino acid networks well.

DOI: 10.1103/PhysRevE.77.061920 PACS number(s): 87.14.E−, 87.15.B−, 87.10.−e

I. INTRODUCTION

From the perspective of topology, the three-dimensional (3D) structure of a protein molecule can be represented as a complex network. In the network, amino acids are vertices and the interactions between them act as edges. The interactions between amino acids include hydrogen bonds, ion pairs, disulfide bonds, and van der Waals interactions, which stabilize the shape of a protein and keep it from falling apart. The information on interactions can help us to develop the contact potentials and predict the structural class, secondary structures of globular proteins, solvent accessibility, fold recognition, and folding rates [1]. Therefore, many previous studies implemented amino acid networks to understand the relationship between structure and function [1,2], and it is usually found that they are small-world networks. Vendruscolo *et al.* [3] analyzed the network parameter betweenness to identify the "key residues." Dokholyan *et al.* [4] found that the network's topological properties are crucial for the protein's kinetic ability to fold. Amaitai *et al.* [5] identified the functional site residues by the network parameter closeness. Atilgan *et al.* [6] found the correlation between the average shortest path lengths and residue fluctuations. Jiao *et al.* [7] studied the network parameter change for the protein of Chymotrypsin Inhibitor 2 (CI2) on its high-temperature unfolding pathway. The above studies have focused on applications of amino acid networks to specific proteins to study the correlation between network properties and protein functions. Ideally, an amino acid network should be constructed from an available protein structure. Yet, for the limitation of experiment, the protein structure information is unknown in most cases. This restricts the tool of amino acid networks to be used for studying the protein functions.

An evolving mechanism is another important aspect of complex network studies, which is usually used to explain the construction process from single individual nodes to the real complex network. Likewise, these mechanisms or models may help us to understand the process from residue sequence to amino acid network. Thus, without the protein structure information the amino acid network can be constructed by the evolving model and used to study the protein functions. Furthermore, the evolving mechanism may give some useful information on protein folding. However, it is difficult for existing network models to explain the evolving mechanism of amino acid networks. First, growing models [8,9] assume that the number of vertices, N, increases throughout the lifetime of the network, which is common to some real networks. However, in the evolution of an amino acid network, the initial state is the primary structure of a protein—i.e., an amino acid sequence—and no vertex is increased or decreased throughout the process of folding. Therefore, general growing models are not suitable for amino acid networks.

Second, small-world evolving models [10,11] assume that the vertices are uniform and the probability that two vertices are connected is random or uniform. For amino acid networks, every residue has its own specific characteristic, which will lead to different orders of interaction between different amino acids [12,13]. Therefore, amino acid networks should exhibit some preferential rules. For example, the hydrophobic collapse model points out that the first event of protein folding consists of a collapse occurring via long-range interactions between hydrophobic residues [14]. The "framework" model emphasizes the role of short-range interactions in directing protein folding [15]. It assumes that the secondary structures are formed in an early step of the folding process before the tertiary structure occurs. These examples indicate that the connection probability of two vertices in an amino acid network is not uniform. Meanwhile, the preferential rule is different from the "rich get richer" dynamics of the scale-free model [8]. Some especial features of proteins, such as the residue type and the sequence distance, will play an important role in the preferential connectivity of the amino acid network. Therefore, it is necessary to define the preferential connectivity rules depending on these special features of proteins.

In the native condition, amino acid networks of proteins are formed in the process of protein folding. The sequence information of the primary structure can be used to determine the protein's native 3D structure [16]. Therefore, the evolving model for amino acid networks should meet two

*Corresponding author.
FAX: +86-10-67392837. cxwang@bjut.edu.cn

requirements: mimicking the characteristics of a small-world network and agreeing with the features of the real process of protein folding to some extent. Our goal here is to give a suitable model for these problems.

II. MATERIAL AND RESEARCH SYSTEM

Two data sets were used to deduce and verify the applicability of our model. One is a data set of 101 low-homology (<10%) proteins, which are selected from the Protein Data Bank (PDB). The proteins vary from 200 to 600 residues, with resolutions better than 1.8 Å and R values lower than 0.2. The other one includes eight proteins whose folding nuclei have been identified by the original experimental groups [17].

III. THEORY AND MODEL

Based on the perspectives of protein folding, we propose two generic assumptions of the evolution model for amino acid networks. According to the energy landscape perspective [18,19], a nascent polypeptide chain is navigated by the funnel-like energy landscape and folded into the native state. Therefore, the evolution of amino acid networks should be first considered to be energy-driven. Thus, it is needed to define the energy preferential attachment—i.e., the energy-based connection probability of two vertices—in the evolution of amino acid networks. In this work, we define that this probability depends on the residue pairing energy. However, if we only take account of the energy preferential attachment, it is difficult to explain the preferential short-range interactions in directing protein folding. Baker [20] also pointed out that interactions between residues close together in sequence are more likely to form early in folding than those between widely separated residues. Therefore, besides the energy preferential attachment, it is deemed that a residue leans to interact with residues of sequence neighbor—i.e., the neighbor preferential attachment. Then, the connectivity probability Π_{ij} of residues i and j depends on the residue pairing energy E_{ij} and the neighbor preferential weight η_{ij}. This reweighting idea is similar with the topomer search model [21], which can quantitatively account for the folding rates of two-state proteins. In this study, the basic definition of connectivity probability Π_{ij} is given by

$$\Pi_{ij} = \frac{\eta_{ij} E_{ij}}{\sum_i \sum_j \eta_{ij} E_{ij}}, \tag{1}$$

where E_{ij} is the link weight according to the magnitude of the contact energy between residues i and j suggested by Miyazawa and Jernigan [22], which is related to the types of the two interacting amino acids. However, the definition of the neighbor preferential weight η_{ij} is not intuitive, so more details about it are now given.

First, we analyze the sequential distance effects for residue interactions. It is needed to define the interaction or edge between residues in the native proteins. The interaction between two residues is usually defined based on the distance between them. In other researches, the cutoff distance is between 4.5 and 8.5 Å [1]. In this study, we choose 6.5 Å as the cutoff distance and two residues are considered to interact if their $C\alpha$ atoms are within a distance of 6.5 Å. Figure 1 shows the number of interactions per residue as a function of the sequential distance between the interacting residues up to the 20th neighbor for the 101 low-homology proteins. As we know, the interaction between i and $i \pm 1$ residues is the covalent bond, so the number is 1. There is a significant borderline between the sequence distances of 4 and 5. The interaction numbers of i with $i \pm 2$, $i \pm 3$, and $i \pm 4$ residues are larger than those with from $i \pm 5$ to $i \pm 20$ residues in the amino acid sequence. This result agrees with previous studies [23]. Thus, we can define residues from $i \pm 2$ to $i \pm 4$ as the sequence neighbor residues of i, which are the preferentially chosen residues for the residue i. Then, the two residues whose sequence distances are more than 4 can be defined as long-range interaction residues. To simplify the expression, we define the weights η of long-range residues as 1. Meanwhile, it is noted that the interaction number of sequence distance 2 is larger than those of sequence distances 3 and 4. The numbers of sequence distances 3 and 4 are similar, but the number of sequence distance 2 is about double of those of sequence distances 3 and 4. It shows that the probability of interaction for sequence distance 2 is about double of those for 3 and 4. Therefore, we define the neighbor preferential weight η for sequence distance 2 as 2 times those for 3 and 4.

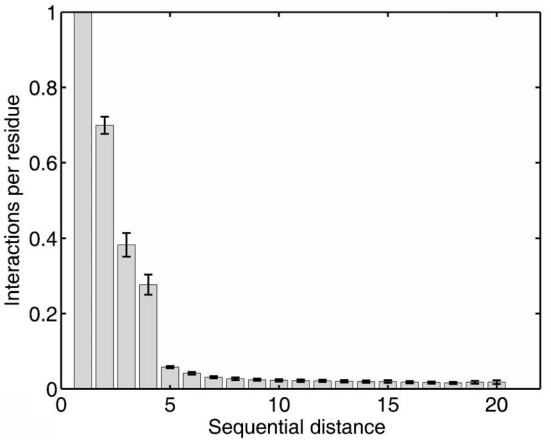

FIG. 1. The average number of interactions per residue as a function of sequential distance. Error bars represent 95% confidence intervals.

However, it is not enough to define the weight η only based on this figure. In the process of protein folding, if a long-range interaction exists between i and j, their sequence neighbor residues, such as $i \pm 1$ and $j \pm 1$, often have some interactions. In this study, we call them space neighbor residues. It is obvious that the weights η of space neighbor residues are larger than those of long-range residues. Thus, this situation should be considered by the neighbor preferential rule and the weights of the space neighbor residues should be estimated. Assuming that i and j have a long-range interaction, it is needed to adjust the neighbor preferential weights η between their neighbor residues $i \pm 1$ and $j \pm 1$. It can be expressed as follows:

$$\begin{array}{c} \\ i-1 \\ i \\ i+1 \end{array} \begin{array}{c} \begin{array}{ccc} j-1 & j & j+1 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array} \Rightarrow \begin{array}{c} \begin{array}{ccc} j-1 & j & j+1 \end{array} \\ \begin{pmatrix} \eta & \eta & \eta \\ \eta & 0 & \eta \\ \eta & \eta & \eta \end{pmatrix} \end{array} \begin{array}{c} \\ i-1 \\ i \\ i+1 \end{array}. \quad (2)$$

Although the sequence distance between them is only 1, the real interaction zone will expand automatically with the evolving process.

To estimate η, we assume that i' and j' are space neighbor residues. All the long-range residues for i' can be considered as a whole. Residue i' will preferentially choose the space neighbor residue to connect. Therefore, the connectivity probability of $\Pi_{i'j'}$ should be larger than the probability sum of all long-range residues. Assuming k' as one of the long-range residues for residue i', we obtain

$$\Pi_{i'j'} > \sum_{k'} \Pi_{i'k'}. \quad (3)$$

After replacing Π with Eq. (1), we get

$$\frac{\eta_{i'j'}E_{i'j'}}{\sum \eta E} > \sum_{k'} \frac{\eta_{i'k'}E_{i'k'}}{\sum \eta E}, \quad (4)$$

and hence

$$\eta_{i'j'} > \frac{\sum_{k'} \eta_{i'k'}E_{i'k'}}{E_{i'j'}}. \quad (5)$$

Since k' is one of the long-range residues for residue i', $\eta_{i'k'}$ is 1 and expression (5) can be reduced. Then we get

$$\eta_{i'j'} > \sum_{k'} \frac{E_{i'k'}}{E_{i'j'}}. \quad (6)$$

Since $\frac{E_{i'k'}}{E_{i'j'}} \sim 1$ and the number of space neighbor and sequence neighbor residues is much less than that of long-range residues, we have

$$\eta_{i'j'} \sim N, \quad (7)$$

where N is the number of the amino acids. From Eq. (7), we can estimate that the neighbor preferential weights η of space neighbor residues are about N times of that of long-range residues. The weight will be adjusted automatically according to the size of the proteins. As we know, the volumes and sequences of different proteins are not uniform. If we define the neighbor preferential weight η of space neighbor residues as the same constant for all proteins, it will be too large or too small for different proteins because of their different sizes. Therefore, this automatically adjusted preferential weight η is appropriate for evolving all the amino acid networks of proteins with different sizes.

We have defined the sequence neighbor residues, the space neighbor residues, and the long-range residues. Meanwhile, the neighbor preferential weights for the space neighbor residues and the long-range residues have also been set. However, the neighbor preferential weights for the sequence neighbor residue are not defined. Here we assume that the neighbor preferential weights η for sequence distances 3 and 4 are w times that of the space neighbor residue. Then, the neighbor preferential weight η can be expressed as

$$\eta_{ij} = \begin{cases} 2wN, & |i-j| = 2, \\ wN, & 2 < |i-j| \le 4, \\ N, & \text{residues } i \text{ and } j \text{ are space neighbor residues}, \\ 1, & \text{residues } i \text{ and } j \text{ are long-range residues}. \end{cases} \quad (8)$$

There is only one parameter w in the evolving model, and we can discuss the effect of w on the evolution of the network.

In addition, due to the spatial constrains imposed by neighboring residues, the vertices in amino acid networks have a limited number of interactions between them [24]. To avoid the interactions number increasing unboundedly, we take an upper limit 15 for it. Therefore, in all, the proposed model is defined as the following scheme.

Step 1. The evolution starts from the primary structure, an amino acid chain with only the covalent bond connected. Then, assign the energy preferential weights for all residues and the neighbor preferential weights for the sequence neighbor residues and the long-range residues.

Step 2. According to the preferential attachment rule, choose residues i and j with degree values less than 15 to connect. The connectivity probability Π_{ij} is given by Eq. (1).

Step 3. If the sequential distance between i and j is larger than 4, the connection between them is the long-range interaction. Then, we adjust the neighbor preferential weights of space neighbor residues around them as Eq. (7).

Repeat steps 2 and 3 until the edges of the network are same as those of the native proteins. For each protein, 1000 independent evolution simulations are performed and the values of the network features are averaged over all simulations.

IV. RESULTS AND DISCUSSION

To verify that both the neighbor and energy preferential rules are necessary, we investigated three forms of the

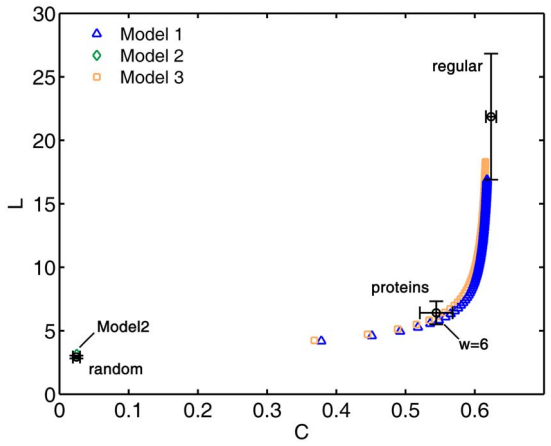

FIG. 2. (Color online) Distribution of the values of the path length and clustering index for the 101 low-homology proteins. Error bars represent the standard deviations of the distributions. For comparison, we also plot data points for random graphs and regular graphs. For evolving models 1 and 3, w increases from 1 to 100.

model. Model 1 takes account of the two preferential rules. Model 2 has only the energy preferential rule. In contrast with model 2, model 3 keeps only the neighbor preferential rule; i.e., it does not distinguish the types of residues. First, the general network characteristics are used for comparison, such as the clustering coefficients C and the characteristic path lengths L. Figure 2 shows the distribution of the network characteristics for the amino acid network of the 101 low-homology proteins and those for the relevant random graphs and regular graphs. These results are consistent with previous studies [3]. Then, we compared the evolving results of our three models. Since model 2 has only the energy preferential rule, its evolving process is directed only by the residue contact energy and does not need to change any parameter. Therefore, in Fig. 2, the evolving result of model 2 has only one point. As shown in Fig. 2, model 2 does not exhibit small-world phenomena, but shows the features of random graphs with small clustering coefficients and small characteristic path lengths. The random graphs have many long-range connections, which lead to an immediate drop of the clustering coefficient C. Since model 2 has only the energy preferential rule, the interactions of residues are determined by the residue pairing energy, but not influenced by the sequence distance between residues. Therefore, many long-range interactions are formed in the evolution process and the evolution network of model 2 is similar to the random graphs. On the contrary, considering the neighbor preferential rule, many short-range interactions are formed to keep the large C and a few long-range interactions are formed to keep the small L. Thus, models 1 and 3 can both exhibit small-world phenomena similar to the native protein. Therefore, the neighbor preferential rule is the major reason for the small-world character of the amino acid network. From Fig. 2, it can be seen that the changes of w have an important effects on models 1 and 3. This parameter is very similar to the parameter p in the small-world model [10,11]. When w is increased from 1 to 100, the network characteristics of models 1 and 3 are changed from the random graph to the regular graph. When w increases, the neighbor preferential weights of sequence neighbor residues increase. Then, the short-range interactions are increasing and the characters of networks tend to those of the regular graphs. However, when w decreases, the long-range interactions are increasing and the characters of networks tend to those of the random graphs. When w is 5–7, the characters of networks are much more similar to those of the native protein structures. We can set $w=6$ in the neighbor preferential rule. It can be found that the neighbor preferential weights of sequence neighbor residues are larger than those of space neighbor residues. In fact, space neighbor residues are of the long-range behavior in sequence distance. Therefore, the result is consistent with Fig. 1, which shows that sequence neighbor residues are easier to connect than long-range residues.

Second, we analyze some biological structure characteristics of native proteins and compare the evolved results of the three models. Nucleation is an important phenomenon in protein folding. Thus, the structure of folding nucleus is the biological structure characteristics of amino acid networks and can differentiate it from other networks. Then, we analyze the connections of folding nucleus residues to compare the three evolving models. To measure the folding nucleus connections, we define $T(F,n)$ as the average contacts of the folding nucleus residues and $P(F,n)$ as the rate of the connections of folding nucleus residues in the noncovalent connections, where F represents the folding nucleus residues and n is the number of noncovalent connections formed in the process of evolution. In our models, the number of noncovalent connections n can approximately express the time of evolution. They can be calculated from the equations

$$T(F,n) = \frac{S_F}{N_F} \tag{9}$$

and

$$P(F,n) = \frac{\frac{S_F}{2n} \times 100\%}{N_F}, \tag{10}$$

where N_F is the number of folding nucleus residues. Along with forming the noncovalent connections, the contacts of the folding nucleus residues will increase; S_F is the sum of these contacts. In Eq. (9), the high value of $T(F,n)$ shows that folding nucleus residues have many contacts. In contrast, the low value of $T(F,n)$ shows that folding nucleus residues have few contacts. If we replace the folding nucleus residues with all residues in a protein, the sum of increased contact numbers for all residues in n noncovalent connections is $2n$. The average value of the contacts for all residues $T(A,n)$ is $\frac{2n}{N}$, where A represents all residues and N is the number of residues. In Eq. (10), the high value of $P(F,n)$ shows that the connections among folding nucleus residues constitute the major part of the connections. In contrast, the low value of $P(F,n)$ shows that the connections of nonfolding nucleus residues are the major part of the connections. Similarly, the average value of the connections for all residues $P(A,n)$ is $\frac{1\times 100\%}{N}$. The value will not change in the

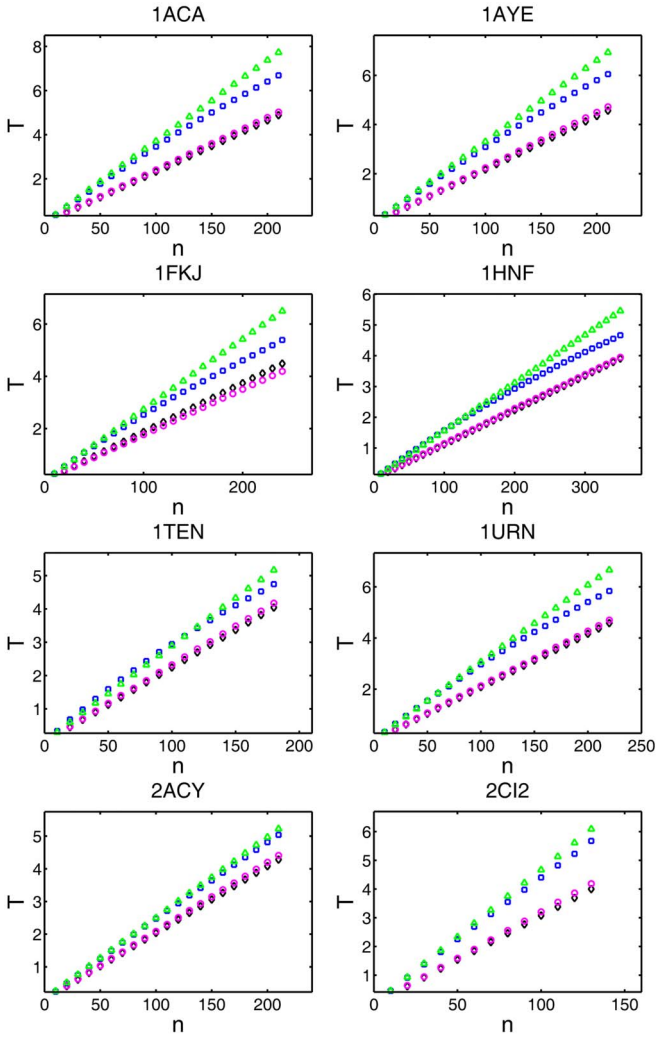

FIG. 3. (Color online) T as a function of noncovalent connection number n is shown for the eight proteins whose folding nuclei have been identified. The blue square is $T(F,n)$ for model 1. The green triangle is $T(F,n)$ for model 2. The pink circle is $T(F,n)$ for model 3. The black diamond shows the average value of the contacts for all residues $T(A,n)$ as the control.

process of evolving, so we can take it as a control for comparison.

Figures 3 and 4 show T and P as functions of noncovalent connection number n for the eight proteins whose folding nuclei have been identified. The values of $P(F,n)$ for the native proteins are higher than the average value of the connections for all residues $P(A,n)$. The folding nucleus residues are usually hydrophobic and collapsed in the core of the proteins, and so they have more connections than other residues. With the energy preferential rule, models 1 and 2 can mimic the higher contact number for the folding nucleus residues $T(F,n)$ than the average value of all residues $T(A,n)$ and show a high value of $P(F,n)$ after the process of adding edges. However, if we do not distinguish the types of residue as model 3, the contact number for the folding nucleus residues $T(F,n)$ is similar to $T(A,n)$, and $P(F,n)$ is not different from the average value $P(A,n)$ in the process of adding edges. It shows that without the energy preferential rule, the folding nucleus residues will not be chosen preferentially in the process of adding edges. Therefore, the energy preferential rule is the major reason for the biological structure character of the amino acid network. Furthermore, $P(F,n)$ of model 1 decreases with the increasing of the noncovalent connections. As the statement of the folding nucleation-growth (or nucleation-condensation) model [25,26], a nucleation step is followed by a rapid propagation. In the beginning of folding, the noncovalent connections are quite few. These connections will form the folding nucleus first. Thus, the average contacts of the folding nucleus residues $T(F,n)$ increase quickly and the rate of the connections of the folding nucleus residues $P(F,n)$ will be high. After the accomplishment of the folding nucleus, the average contacts of the folding nucleus residues will increase slowly and the rate of the connections of the folding nucleus will decrease with the propagation. Therefore, $T(F,n)$ increases slowly and $P(F,n)$ will be low. By using model 1, we find that the simulation result is in agreement with the nucleation-growth model. However, without the neighbor preferential rule as model 2, $T(F,n)$ keeps increasing quickly and $P(F,n)$ holds high values in the process of adding edges.

The folding rate is also an important parameter in protein folding. It is found that there are experimental data [27] of the folding rate k_f for four proteins 1AYE, 1TEN, 1URN, and 2ACY in Fig. 3, whose $\ln k_f$ are 6.8 s^{-1}, 1.1 s^{-1}, 5.8 s^{-1}, and 0.92 s^{-1}, respectively. Some studies show that long-range contact orders have strong negative correlation with the protein folding rate [27]. Our model can imply some information on long-range contacts, so it may reflect some meaningful points for the protein folding rate. As we know, proteins with high long-range contact orders should have much more long-range contacts in their structures. If proteins have low values of the folding rate, such as 1TEN and 2ACY, they will have high values of the long-range contact order. Then, a proper evolution model used to simulate this kind of proteins with low folding rates should give out more long-range contacts. In the networks evolved with the energy preferential rule as models 1 and 2, many long-range interactions are formed. As shown in Figs. 3 and 4, we find that the evolution process of model 1 is similar to that of model 2 for these proteins. In contrast, for proteins with high values of the folding rate, such as 1AYE and 1URN, a proper evolution model should give out less long-range contacts. Then, the evolution processes of model 1 are far from those of model 2.

The degree distribution is an important statistical characteristic for complex networks and often determines important global characteristics [28]. The correlation coefficients of the degree distribution are used to compare the similarity between the two networks of simulation and native protein. In this study, besides the general correlation of the degree distribution, we define another correlation for the single residue degree. Since every residue has its special sequence number and degree value, each sequence number will correspond to a degree value. The correlation of the residue degree between simulation and native protein can be also used to show the similarity. We call it the degree correlation of the single residue. The number of residues is much more than the number of degrees, so the degree correlation for the single residue is

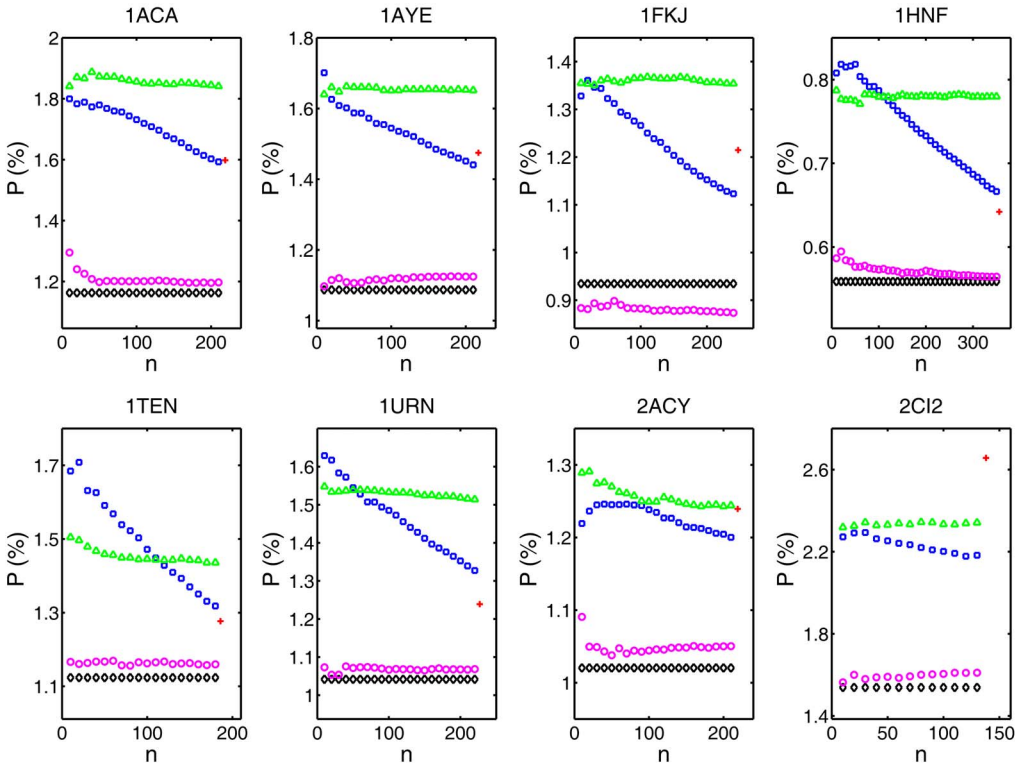

FIG. 4. (Color online) P as a function of noncovalent connections number n is shown for the eight proteins whose folding nuclei have been identified. The blue square is $P(F,n)$ for model 1. The green triangle is $P(F,n)$ for model 2. The pink circle is $P(F,n)$ for model 3. The red cross is $P(F,n)$ for the native protein. The black diamond shows the average value of the connections for all residues $P(A,n)$ as the control.

more sensitive to compare the similarity than the general correlation of degree distribution. Figure 5 shows the correlation coefficient of the degree for the single residue and the correlation coefficient of the degree distribution for 101 low-homology proteins. The degree values of the single residue in model 1 are in strong correlation with those of the native proteins. This result shows that the residues with many interactions in the native proteins also have high degree values by evolving with the two preferential rules. In contrast, residues with few interactions in the native proteins have small degree values. However, without the energy preferential rule as model 3, the correlation is not strong. By using model 3, the evolution of networks does not consider the difference of the residue types. Therefore, it is difficult for the energy preferential residues, such as the hydrophobic residues, to exhibit the high degree. Meanwhile, without the neighbor preferential rule as model 2, the correlation is not strong yet. The energy preferential residues will keep much higher probability to link than those of the native proteins. Therefore, the connectivity is also not similar to the native proteins. As a result, compared with the degree distribution, the correlation of model 1 is stronger than those of models 2 and 3. From the comparisons, it can be seen that by taking account of the two rules, model 1 has the better performance than models 2 and 3 in the reflection of the structure of native proteins.

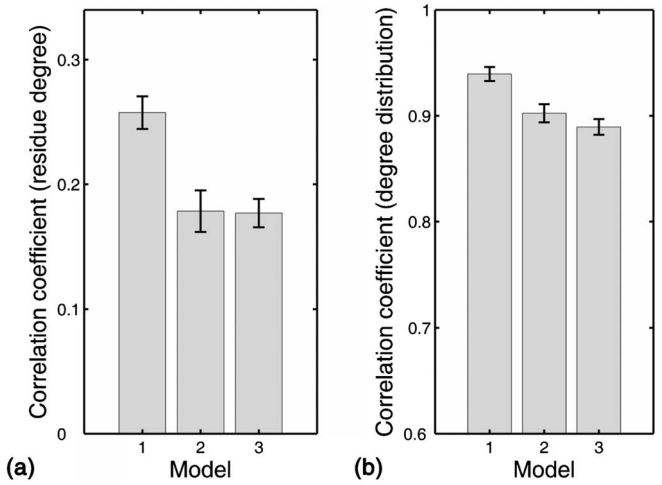

FIG. 5. The correlation coefficient of the degree for the single residue (a) and the correlation coefficient of the degree distribution (b) for the 101 low-homology proteins. Error bars represent 95% confidence intervals.

Our model is a simple evolving model, and it can give some global and/or statistical results for the amino acid network. Similarly, many evolving models [29] have been used to explain the global and/or statistic characteristics of real-world networks. For example, the scale-free model [8] can explain a series of networks with a special degree distribution. Since studies of the evolving model emphasize particularly the description of global and/or statistical properties, it is difficult for them to give elaborate structures of real networks. Our model also has the same problem. It is difficult to describe and show the protein structures only with informa-

tion on contacts. Therefore, it is unsuitable for our model to describe the detailed coordinates of atoms and compare with atom-level protein-folding models. The main advantage of our model is to easily give out the contact information between residues through the perspectives of network evolution. However, these contacts are not completely consistent with the native contacts of proteins. We use this simple model just to explore the evolving mechanism of amino acid networks and get some global and/or statistical results.

V. CONCLUSION

Based on the perspectives of protein folding, we propose an evolution model of amino acid networks. In our model, the evolution starts from the amino acid sequence of a native protein and it is guided by the two generic assumptions of the neighbor preferential and the energy preferential rules. Applied to two data sets of native proteins, it is found that our model not only can mimic the characters of general networks, but also can agree with the real process of protein folding to some extent. Furthermore, we find that the neighbor preferential rule predominates the general network characters and the energy preferential rule predominates the specific biological structure characters. The model with one type of preferential rules can only obtain part of the characters of the amino acid network. In addition, the model shows that both topology and energy play important roles in protein folding. These mechanisms might provide some insights for future studies of protein folding.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grants No. 10574009, No. 20773006, and No. 30670497).

[1] M. M. Gromiha and S. Selvaraj, Prog. Biophys. Mol. Biol. **86**, 235 (2004).
[2] C. Bode, I. A. Kovacs, M. S. Szalay *et al.*, FEBS Lett. **581**, 2776 (2007).
[3] M. Vendruscolo, N. V. Dokholyan, E. Paci, and M. Karplus, Phys. Rev. E **65**, 061910 (2002).
[4] N. V. Dokholyan, L. Li, F. Ding *et al.*, Proc. Natl. Acad. Sci. U.S.A. **99**, 8637 (2002).
[5] G. Amitai *et al.*, J. Mol. Biol. **344**, 1135 (2004).
[6] A. R. Atilgan, P. Akan, and C. Baysal, Biophys. J. **86**, 85 (2004).
[7] X. Jiao, S. Chang, C. H. Li, W. Z. Chen, and C. X. Wang, Phys. Rev. E **75**, 051903 (2007).
[8] A. L. Barabasi and R. Albert, Science **286**, 509 (1999).
[9] A. L. Barabasi, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[10] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[11] M. E. J. Newman and D. J. Watts, Phys. Lett. A **263**, 341 (1999).
[12] K. W. Plaxco, K. T. Simons, and D. Baker, J. Mol. Biol. **277**, 985 (1998).
[13] E. Shakhnovich, Chem. Rev. **106**, 1559 (2006).
[14] K. A. Dill, Biochemistry **24**, 1501 (1985).
[15] O. B. Ptitsyn and A. A. Rashin, Dokl. Akad. Nauk SSSR **213**, 473 (1973).
[16] C. B. Anfinsen, Science **181**, 223 (1973).
[17] L. Mirny and E. Shakhnovich, Annu. Rev. Biophys. Biomol. Struct. **30**, 361 (2001).
[18] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[19] J. N. Onuchic, H. Nymeyer, A. E. Garcia *et al.*, Adv. Protein Chem. **53**, 87 (2000).
[20] D. Baker, Nature (London) **405**, 39 (2000).
[21] D. E. Makarov and K. W. Plaxco, Protein Sci. **12**, 17 (2003).
[22] S. Miyazawa and R. L. Jernigan, J. Mol. Biol. **256**, 623 (1996).
[23] Z. Gugolya, Z. Dosztanyi, and I. Simon, Proteins: Struct., Funct., Bioinf. **27**, 360 (1997).
[24] M. Aftabuddin and S. Kundu, Physica A **369**, 895 (2006).
[25] D. B. Wetlaufer, Proc. Natl. Acad. Sci. U.S.A. **70**, 697 (1973).
[26] V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich, Biochemistry **33**, 10026 (1994).
[27] A. Y. Istomin, D. J. Jacobs, and D. R. Livesay, Protein Sci. **16**, 2564 (2007).
[28] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[29] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).