

Folding rate prediction using complex network analysis for proteins with two- and three-state folding kinetics

Hai-Yan Li¹, Ji-Hua Wang¹

¹Key Lab of Biophysics in Universities of Shandong, Dezhou University, Shandong, China.
Email: tianwaifeixian78@163.com, jhwyh@yahoo.com.cn

Received 5 September 2009; revised 9 October 2009; accepted 10 October 2009.

ABSTRACT

It is a challenging task to investigate the different influence of long-range and short-range interactions on two-state and three-state folding kinetics of protein. The networks of the 30 two-state proteins and 15 three-state proteins were constructed by complex networks analysis at three length scales: Protein Contact Networks, Long-range Interaction Networks and Short-range Interaction Networks. To uncover the relationship between structural properties and folding kinetics of the proteins, the correlations of protein network parameters with protein folding rate and topology parameters contact order were analyzed. The results show that Protein Contact Networks and Short-range Interaction Networks (for both two-state and three-state proteins) exhibit the “small-world” property and Long-range Interaction networks indicate “scale-free” behavior. Our results further indicate that all Protein Contact Networks and Short-range Interaction networks are assortative type. While some of Long-range Interaction Networks are of assortative type, the others are of disassortative type. For two-state proteins, the clustering coefficients of Short-range Interaction Networks show prominent correlation with folding rate and contact order. The assortativity coefficients of Short-range Interaction Networks also show remarkable correlation with folding rate and contact order. Similar correlations exist in Protein Contact Networks of three-state proteins. For two-state proteins, the correlation between contact order and folding rate is determined by the numbers of local contacts. Short-range interactions play a key role in determining the connecting trend among amino acids and they impact the folding rate of two-state proteins directly. For three-state proteins, the folding rate is determined by short-range and long-range interactions among residues together.

Keywords: Protein Contact Networks; Small-World; Scale-Free; Assortative Type; Folding Rate

1. INTRODUCTION

The network concept is increasingly used to describe the topology and dynamics of complex systems. As the essential matter of life, proteins are biological macromolecules made up of a linear chain of amino acids and fold into unique three-dimensional structures (native states). Despite the large degrees of freedom, proteins fold into their native states in a very short time. It is important to understand how proteins consistently fold into their native-state structures and the relationship between structures and function. A protein molecule can be treated as a complex network with each amino acid simplified as a node and the interaction between them as a link. Efforts have been made to model proteins as networks for studying protein topology, small world properties and examining the nucleation in protein folding [1-10]. Bagler and Sinha [11], in their recent protein network analysis, constructed Protein Contact Networks and Long-range Interaction Networks to analyze the assortative mixing of networks and folding kinetics of two-state proteins.

But there is a significant difference in the folding behavior of small proteins with simple two-state kinetics and of larger proteins having a three-state folding kinetics [12]. The two-state proteins have no visible intermediates in the course of folding, which therefore occur as an “all-or-none” process under all experimental conditions. However, the proteins with three-state folding kinetics fold via intermediates, which accumulate during the early stages of folding when it occurs in denaturant-free water [13-16]. Based on the work by Bagler and Sinha, two- and three-state proteins that belong to different structural classes were selected from protein crystal structure data bank to model the native-state protein structures as networks. To investigate various topological properties, the network models were constructed at three different length scales. Protein Contact Networks (PCNs) were built by considering the contacts between atoms in amino acid residues. There is a natural distinction of contacts into two types: long-range and short-

range interactions [7]. We considered the Long-range Interaction Networks (LINs) and Short-range Interaction Networks (SINs) of each protein, which are subsets of the corresponding PCNs. To investigate if the general network parameters can offer any clue to the biophysical properties of the existing three dimensional structure of a protein, these networks were analyzed to focus on their topology including clustering coefficients, shortest path length, average degree, degree distribution and assortative mixing behavior of the amino acid nodes. The determination of folding rate for two- and three-state folding kinetics has a significant difference. To uncover the relationship between the structural properties and the folding kinetics of the proteins, the correlation of protein network parameters with protein folding rate ($\ln k_f$) and topology parameters contact order (CO) was analyzed. The values of $\ln k_f$ and CO are available as given in Reference [12]. Through our coarse-grained complex network model of protein structures, it was found that short-range interactions play a key role in determining the connecting trend among amino acids and impact directly the folding rate of two-state proteins. For three-state proteins, the folding rate is determined by short-range and long-range interactions among residues together.

2. METHODS

2.1. Construction of PCNs, LINs, SINs and Their Random Networks

In this paper, 30 proteins with two-state kinetics and 15 proteins with three-state kinetics were studied and the dataset was taken from the paper [12]. The data of these protein structures were taken from the Protein Data Bank (PDB) to model them as Protein Contact Networks (PCNs) by setting the C_α atoms as the nodes, and established a link between two nodes, if the atoms were within a cut-off distance (0.8nm).

The Long-range Interaction Network (LIN) of a PCN was obtained by considering the interactions which occur between amino acids that were twelve or more amino acids apart in the primary sequence. A LIN was a subset of its PCN with same numbers of nodes (N) but fewer numbers of links due to removal of the short-range contacts. The Short-range Interaction Network (SIN) of a PCN was built with the amino acids separated within twelve. For compare, the random network was constructed with the same numbers of residues (N) and links as those of the PCNs, SINs, LINs.

2.2. Network Parameters

The degree of any node i is represented by $k_i = \sum_{j=1}^N a_{ij}$.

Here a_{ij} is the element of the adjacency matrix, whose value is 1 if an edge connects a node “ P ” to another node

“ J ” and 0 otherwise. N is the number of nodes. Average degree $\langle k \rangle$ of a network is defined as $\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i$.

The shortest path length is related to the link number of a pathway between two nodes and it is the least link number of all the pathways between two nodes. The average shortest path length is defined as $L = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N L_{ij}$, where L_{ij} is the shortest path length between nodes i and j .

The average clustering coefficient C is the average over all vertices of the fraction of the number of connected pairs of neighbours for each vertex. It is calculated as follows: $C = \frac{1}{N} \sum_{i=1}^N C_i$, where C_i is the clustering coefficient for a node i and defined as the fraction of links that exist among its nearest neighbours to the maximum number of possible links among them. It scales the cohesiveness of the neighbours of a certain node from the view of topology.

Many networks show “assortative mixing” on their degrees. The Assortativity Coefficient (r) measures the tendency of degree correlation. It is the Pearson Correlation Coefficient of the degrees at either ends of an edge. Its value was calculated using the function suggested by Newman [17] and was given as

$$r = \frac{M^{-1} \sum_i j_i k_i - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}, -1 \leq r \leq 1$$

The networks having positive r values are assortative in nature and the negative value implies that the network is of disassortative type.

3. RESULTS AND DISCUSSION

3.1. Network Parameters of PCNs, LINs, SINs

3.1.1. Average Degree of the Networks

The average degree $\langle k \rangle$ was calculated for each of the three type networks (PCNs, SINs, LINs) of two- and three-state proteins. **Figure 1** shows the average degree $\langle k \rangle$ as a function of network size N . **Table 1** shows the average degree of three type networks for two- and three-state proteins. The values of $\langle k \rangle$ have no obvious difference between two- and three-state proteins. In other words, the average number of contacts per residue for three-state proteins is similar equal to that of two-state proteins. For two-state proteins, the average number of short-range contacts is smaller than that of

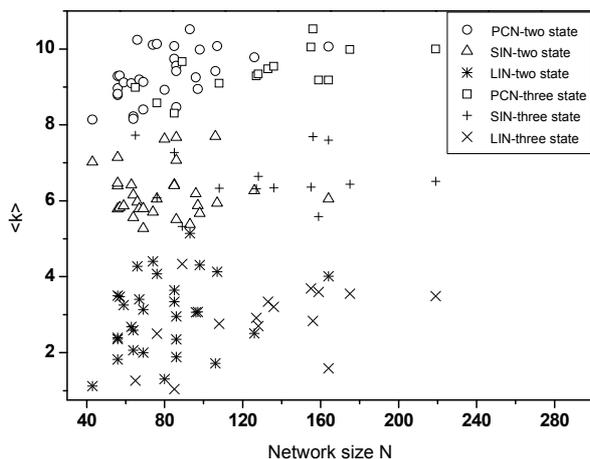


Figure 1. Average degree $\langle k \rangle$ as a function of the network size N for 30 two-state and 15 three-state proteins.

Table 1. Average degree $\langle k \rangle$ of PCNs, SIns, LIns.

Network type	$\langle k \rangle_{\text{two-state}}$	$\langle k \rangle_{\text{three-state}}$
PCNs	9.31±0.664	9.41±0.582
SIns	6.23±0.668	6.56±0.724
LIns	2.99±0.999	2.85±0.938

three-state proteins and the average number of long-range contacts is slightly higher than that of three-state proteins. In general, for coarse-grained complex network model of protein structures, it has been shown, for different folding kinetics, that the short-range interactions and long-range interactions are consistent with each other for a statistical equilibrium. It is observed that the average degree $\langle k \rangle$ for LIns shows lower values than that of SIns and PCNs regardless of their states. It indicates that long-range interactions exhibit a predominant lower average connectivity compared with short-range interactions. Protein structure has the strongest average connectivity by integrating both short-range and long-range interactions. To verify whether the observed trend depends on the network size (i.e., the number of amino acids of the protein), the correlation coefficient between $\langle k \rangle$ and N was calculated. Any significant relationship between $\langle k \rangle$ and N in SIns and LIns for two- and three-state proteins was not found. On the other hand, **Figure 1** indicates that the $\langle k \rangle$ of PCNs (both two- and three-state proteins) show a high positive correlation with N . The correlation coefficients of three-state proteins are higher than that of two-state proteins, and their values are 0.672 ($p=0.006$) and 0.511 ($p=0.004$), respectively.

3.1.2. “Small-World” Property

To examine whether the networks have the “small-world” property, the average clustering coefficient C

and the average shortest path length L of each of the networks and their respective C_r and L_r for the random networks with the same size were calculated. According to Watts and Strongatz [18], a network has the “small-world” property if $C \gg C_r$ and $L \geq L_r$. C_r and L_r can be calculated using the expressions $C_r \approx \langle k \rangle / N$ and $L_r \approx \ln N / \ln \langle k \rangle$. **Table 2** shows the $\langle C \rangle$ and $\langle L \rangle$ of 30 two-state proteins and 15 three-state proteins and the corresponding values of random networks. It is obviously found that PCNs and SIns (both two- and three-state proteins) are characterized by large values of $\langle C \rangle$ and $\langle L \rangle$ compared with the corresponding random networks, which have the typical property of small-world networks. It indicates that any two amino acids are connected with each other via only a few other amino acids in both two- and three-state proteins. Whereas LIns have similar $\langle C \rangle$ with their random networks and their $\langle L \rangle$ are smaller than those of the corresponding random networks. It indicates that LIns do not exhibit the “small-world” property. **Table 2** also shows that two-state proteins have similar values of $\langle C \rangle$ with three-state proteins for three types networks and LIns have remarkable lower $\langle C \rangle$ than those of PCNs and SIns. It suggests that long-range interactions have reduced congregating of amino acids, which may facilitate communication among distant residues in the native structure to some extent, but such a feature can also increase the folding time as it requires distant residues in the chain to come closer during the folding process. **Table 2** also shows that $\langle L \rangle$ of three-state proteins are more higher compared with corresponding two state proteins. It suggests that three-state proteins are packed more loosely than two-state proteins and it has a low global connectivity compared with two-state proteins.

3.1.3. Degree Distribution

The degree distribution is an important feature which characterizes the network topology. **Figure 2** shows the degree distribution of three types’ networks for two- and three-state proteins. The shape of the degree distribution of small-world network is bell-shaped, Poisson-like. It has a pronounced peak at $\langle k \rangle$ and decays exponentially for large k . Thus the topology of the network is relatively homogeneous, all nodes having approximately the same number of edge. The shape of the degree distribution is Poisson distribution, which is another typical property of “small-world” networks. A network lacking a characteristic scale $\langle k \rangle$ and having degree distribution of a power-law form is known as “scale-free” network [19]. From **Figure 2(a)**, the long-range interaction distribution patterns (both two- and three-state), it is noticed that a large number of nodes with a small number of links and a small

Table 2. Values of $\langle C \rangle$ and $\langle L \rangle$ of three types networks of two- and three-state proteins as well as those for the corresponding random networks.

Network type	$\langle C \rangle$	$\langle Cr \rangle$	$\langle L \rangle$	$\langle Lr \rangle$
Two-state				
PCNs	0.589	0.122	2.97	1.947
SINs	0.649	0.085	7.269	2.413
LINs	0.039	0.04	3.796	5.677
Three-state				
PCNs	0.57	0.079	3.694	2.139
SINs	0.65	0.056	10.827	2.584
LINs	0.042	0.023	4.828	13.953

number of links with a large number of nodes in the distribution pattern, indicating the “scale-free” behaviour. But from **Figure 2(b)**, the short-range interaction distribution patterns as well as those of total interactions are of the Poisson type. This indicates again that PCNs and SINs exhibit “small-world” behaviour, while LINs indicate “scale-free” behaviour.

The scale-free degree distribution of LINs indicates that proteins contain hubs, i.e. central residues, which have a large number of long-range interactions with other residues. The kinetic mechanism of transitions from the denatured state to the native state is nucleation [20]. The nucleus is composed of a set of adjacent residues, and is stabilized by long-range interactions that are formed as the rest of the protein collapses around it. The Poisson degree distribution means that protein structures have a much smaller number of hubs than most self-organized networks including most cellular or social networks. The major reason for this deviation from the scale-free degree distribution lies in the limited simultaneous binding capacity of a given amino acid side-chain

(also called as excluded volume effect). The limited amino acid side chain binding capacity contributes to the fact that each amino acid has a characteristic average degree. This depends on the interaction cut-off, which makes hydrophilic amino acids “strong hubs” (observed at high interaction cut-off allowing low overlaps), and hydrophobic amino acids “weak hubs” (at low interaction cut-off allowing high overlaps), respectively. Hubs are integrating various secondary structure elements, and, therefore, it is not surprising that they increase the thermodynamic stability of proteins.

3.1.4. Assortative Mixing Behavior of the Nodes

The assortative mixing concept has been used in social, technological and biological networks [17]. In social networks assortative mixing leads to homophily, i.e., the tendency of individuals to associate with similar partners. This quantity is also important to control epidemics since assortative has a profound impact on the percolation in networks. Contrary to social networks, which tend to be assortative, biological and technological networks tend to be disassortative. Concerning this aspect, the networks are classified as to show assortative mixing, if the degree correlation is positive, a preference for high-degree nodes to attach to other high-degree nodes, or disassortative mixing, otherwise. Assortativity Coefficient (r) for each of the networks was calculated, as shown in **Table 3**. It indicates that all the PCNs and SINs have positive r values regardless of two-state or three-state, while the LINs have both positive and negative r values. The ratio of negative r_L values for two-state is significantly higher than that for three-state. The former is 17/30, while the latter is 3/15.

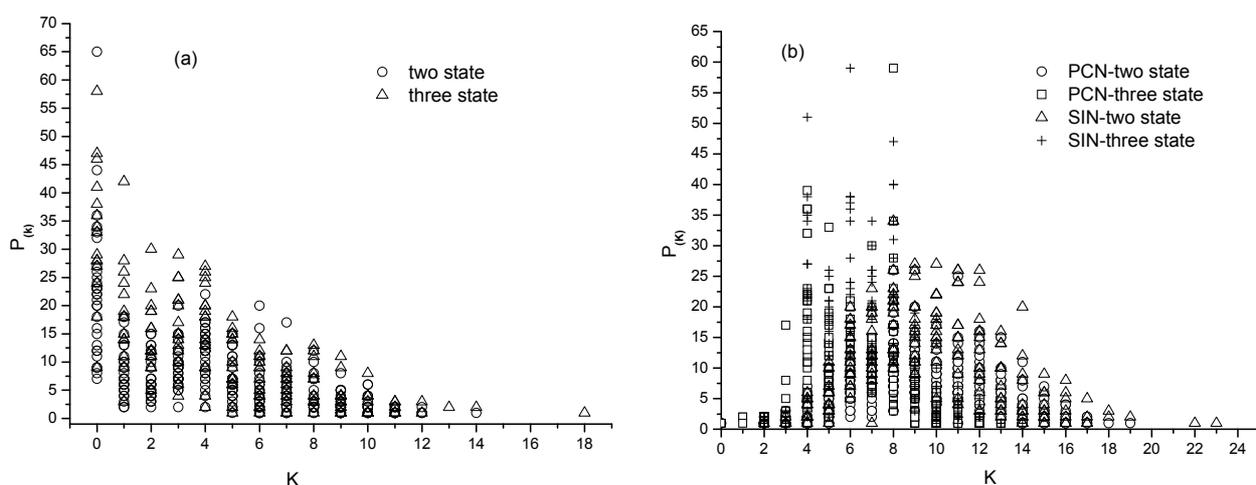


Figure 2. Degree distribution of three type networks for two- and three-state proteins. (a) LINs of two- and three-state proteins; (b) PCNs and SINs of two- and three-state proteins.

Table 3. Assortativity coefficient (r) of three type networks for two- and three-state proteins.

Network type	PCNs	SINs	LINs
r (Two-state)	0.106~0.424	0.106~0.562	-0.566~-0.006 0~0.35
r (Three-state)	0.074~0.562	0.047~0.54	-0.349~-0.117 0.008~0.369

The r values of different networks suggest that all PCNs and SINs are of assortative type, the LINs of three-state proteins (except three) are also of assortative type. While maximum of LINs of two-state have the characteristics of disassortative mixing, few others are of assortative type. Thus it may be said that in all of the PCNs and SINs the residues (nodes) with high degree have tendencies to be attached with the residues having high degree values. The result is consistent with previous study by S. Kundu [21] and Ganesh Bagler [11]. But in some LINs of two-state and three-state proteins having negative r values the mixing pattern of amino acid residues are different. Here the amino acids (nodes) having high degree values have a tendency to be attached with amino acids with smaller degree. This result is not consistent with Ganesh Bagler, who concluded that the assortative mixing in PCNs and LINs is a generic feature of protein structures. Recent research suggests that assortative mixing by degree reduces the stability of networks [22]. In almost all biological networks (e.g. protein interaction network, neural network etc.), nodes of high degree tend to avoid being connected to other highly connected nodes, i.e. these networks show disassortative mixing. This difference of assortative mixing between SINs and LINs may be a possible reason for the stability of native-state proteins and the research of assortative mixing in LINs may give interesting surprises in the future. However, the PCN is a composite network of SIN and LIN. When considering the protein structure networks, the r values had been obtained, which represent a cumulative effect of either all positive r values or a mixture of positive and negative r values. Thus it was found that protein structure networks always have positive r values and they are assortative.

3.2. Correlations of Protein Network Parameters with Folding Rate ($\ln k_f$) and Contact Order (CO)

To uncover the relationship between the structural properties and the folding kinetics of the proteins, the correlation of protein network parameters with protein folding rate ($\ln k_f$) and contact order (CO) was studied. The correlation coefficient between general network parameters (e.g., C , L , $\langle k \rangle$, and r) and the folding rate logarithm ($\ln k_f$) were calculated out. And similar correlation between network parameters and CO was also discussed.

3.2.1. Correlation for Two-State Kinetics

For all the 30 two-state proteins, the clustering coefficients C of PCNs and LINs have not any significant relationship with the $\ln k_f$, and the correlation coefficients are 0.248 ($p=0.186$), -0.118 ($p=0.534$), respectively. However, SINs have high positive correlation between C and $\ln k_f$ (correlation coefficient are 0.602, $p=0.000$). From **Table 2**, the clustering coefficients of LINs are significant lower than those of PCNs and SINs, which show a low correlation with the folding rate of the proteins. It indicates that clustering of amino acids that participate in the long-range interactions, into "cliques" slows down the folding process of two-state proteins. SINs have the highest clustering coefficients among them and C of SINs have significant correlation with the folding rate, indicating that the short-range interactions may be playing a constructive and active role in determining the rate of the two state proteins folding process. The similar correlation occurs between r and $\ln k_f$. The correlation coefficient between r and $\ln k_f$ of SINs is -0.625 ($p=0.000$). For PCNs and LINs, the correlation coefficients are 0.295 ($p=0.181$) and 0.121 ($p=0.753$), respectively. It shows that short-range interactions play a key role in determining the connecting trend among amino acids and influence the folding rate of two-state proteins directly.

Previous studies have found that contact order (CO) has a significant correlation with folding rate of proteins (correlation coefficient of these 30 proteins is -0.72 , $p=0.000$). As an experiential parameter based on 3D structure, though significant correlating with folding rate, the physical meanings of CO is ambiguity. In this study, it is found that C_{SIN} and r_{SIN} have a high correlation with contact order (CO). The correlation coefficients are -0.64 ($p=0.000$) and 0.817 ($p=0.000$), respectively. Since the clustering coefficients depend on the degree of the node, we calculated the correlation coefficients between $C_{\text{SIN}} * \langle k \rangle_{\text{SIN}}$ and $\ln k_f$. It shows high positive correlation (correlation coefficients are 0.733, $p=0.000$) between them for these two-state proteins. A significant high correlation also exists between $C_{\text{SIN}} * \langle k \rangle_{\text{SIN}}$ and CO, the value is -0.796 ($p=0.000$) (see **Figure 3**).

C_{SIN} measures the transitivity in the short-range interaction network and $\langle k \rangle_{\text{SIN}}$ measures the average number of short-range contacts per residue. It indicates that the correlation between CO and $\ln k_f$ is determined by the number of local contacts for two-state proteins. It is consistent with the previous study by Mirny and Shakhnovich [23]. It is interesting to note that despite dissimilar quantities that CO and C_{SIN} measure, the similar correlation coefficients essentially indicate the important role of short-range contact formation in the rate of folding for two-state proteins.

3.2.2. Correlation for Three-State Kinetics

For three-state proteins, the clustering coefficients C of

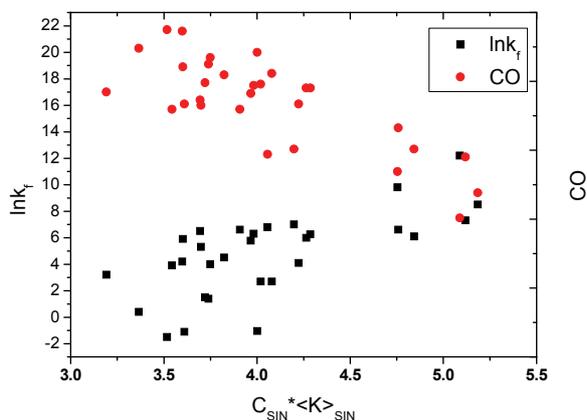


Figure 3. Correlation between $C_{SIN} * \langle k \rangle_{SIN}$ with $\ln k_f$ and CO. constructed to uncover the different influence of long-range and short-range interactions on two- and three-state folding kinetics. It was found that PCNs and SINS (both two- and three-state proteins) have the typical property of small-world networks, whereas LINS exhibit the “scale-free” property.

the PCNs show a high positive correlation with the folding rate (correlation coefficient is 0.652, $p=0.001$). However, C of LINS and SINS have not significant relationship with the $\ln k_f$, and the correlation coefficients between C and $\ln k_f$ are -0.278 ($p=0.315$) and 0.405 ($p=0.081$), respectively. The similar correlation occurs between r and $\ln k_f$. The correlation coefficient between r and $\ln k_f$ of PCNs is -0.603 ($p=0.017$). For LINS and SINS, the correlation coefficients are -0.394 ($p=0.146$) and -0.474 ($p=0.075$), respectively. It shows that, for three-state proteins, the folding rate is determined by short-range and long-range interactions among residues together.

4. CONCLUSIONS

The network concept is increasingly used to describe the topology and dynamics of complex systems. In this paper, the three type networks (PCNs, LINS, SINS) were. All of PCNs, SINS and nearly all LINS of three-state proteins are of assortative type. While maximum of LINS of two-state are of disassortative type. This different assortative mixing behaviour of LINS may be a possible reason for the stability of native-state proteins and the research of assortative mixing in LINS may give interesting surprises in the future.

For two-state proteins, C_{SIN} and r_{SIN} show high correlation with $\ln k_f$ and CO, which indicates the correlation between CO and $\ln k_f$ is determined by the numbers of local contacts. Short-range interactions play a key role in determining the connecting trend among amino acids and influence directly the folding rate of two-state proteins. For three-state proteins, C_{PCN} and r_{PCN} also show high correlation with $\ln k_f$ and CO, which shows that the

folding rate is determined by short-range and long-range interactions among residues together.

5. ACKNOWLEDGMENTS

This work was supported by a grant from Chinese National Key Fundamental Research Project (No. 30970561) and Shandong Fundamental Research Project (NO. Y2005D12).

REFERENCES

- [1] M. Vendruscolo, N. V. Dokholyan, E. Paci and M. Karplus. (2002) Small-world view of the amino acids that play a key role in protein folding, *Physical Review E*, **65**, 061910.
- [2] N. V. Dokholyan, L. Li and F. Ding. (2002) Topological determinants of protein folding, *Proc. Natl. Acad. Sci. (USA)*, **99**, 8637–8641.
- [3] A. R. Atilgan, P. Akan and C. Baysal. (2004) Small-world communication of residues and significance for protein dynamics, *Biophys. J.*, **86**, 85–91.
- [4] G. Amitai, A. Shemesh and E. Sitbon. (2004) Network analysis of protein structures identifies functional residues, *J. Mol. Biol.*, **344**, 1135–1146.
- [5] D. J. Jacobs, A. J. Rader and L. A. Kuhn. (2001) Protein flexibility predictions using graph theory, *Proteins: Structure, Function, and Genetics*, **44**, 150–165.
- [6] X. Jiao, S. Chang, C. H. Li, W. Z. Chen and C. X. Wan. (2007) Construction and application of the weighted amino acid network based on energy, *Phys. Rev. E.*, **75**, 051903.
- [7] L. H. Greene and V. A. Higman. (2003) Uncovering network systems within protein structures, *J. Mol. Biol.*, **334**, 781–791.
- [8] M. Aftabuddin and S. Kundu. (2006) Weighted and unweighted network of amino acids within protein, *Physica A.*, **396**, 895–904.
- [9] S. Kundu. (2005) Amino acids network within protein, *Physica A.*, **346**, 104–109.
- [10] K. V. Brinda and S. Vishveshwara. (2005) A network representation of protein structures: implications for protein stability, *Biophys J.*, **89**, 4159–4170.
- [11] G. Bagler and S. Sinha. (2007) Assortative mixing in protein contact networks and protein folding kinetics, *Bioinformatics*, **23(14)**, 1760–1767.
- [12] O. V. Galzitskaya, S. O. Garbuzynskiy, D. N. Ivankov and A. V. Finketics. (2003) Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics, *Proteins: Structure, Function, and Genetics*, **51**, 162–166.
- [13] S. E. Jackson. (1998) How do small single-domain proteins fold? *Fold Des.*, **3**, 81–91.
- [14] A. R. Fersht. (1999) Kinetics of protein folding, In: Hadler GL. Editor: *Structure and mechanism in protein science*, W.H. Freeman & Co., New York, 40–572.
- [15] A. R. Fersht. (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism, *Proc Natl Acad Sci USA*, **97**, 1525–1529.
- [16] W. A. Eaton, V. Munoz, S. J. Hagen, G. S. Jas, L. J. Lapidus, E. R. Henry and J. Hofrichter. (2000) Fast ki-

- netics and mechanisms in protein folding, *Annu Rev Biophys Biomol Struct*, **29**, 327–359.
- [17] M. E. J. Newman. (2002) Assortative mixing in networks, *Phys. Rev. Lett.*, **89**, 208701–208704.
- [18] D. J. Watts and S. H. Strogatz. (1998) Collective dynamics of “small-world” networks, *Nature*, **393**, 440–442.
- [19] M. E. J. Newman. (2003) The structure and function of complex networks, *SIAM Rev.*, **45(2)**, 167–256.
- [20] A. R. Fersht. (1995) Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications, *Proc Natl Acad Sci USA*, **92(24)**, 10869–10873.
- [21] M. Aftabuddin and S. Kundu. (2007) Hydrophobic, hydrophilic and charged amino acids’ networks within protein, *Biophysical Journal*, **93(1)**, 225–231.
- [22] M. Brede and S. Sinha. (2005) Assortative mixing by degree makes a network more unstable, eprint arXiv: cond-mat/0507710.
- [23] L. Mirny and E. Shalhovich. (2001) Protein folding theory: From lattice to all-atom models, *Annu Rev Biophys Biomol Struct*, **30**, 361–396.