

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/8167975>

# Topological determinants of protein unfolding rates

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · FEBRUARY 2004

Impact Factor: 2.63 · DOI: 10.1002/prot.20324 · Source: PubMed

---

CITATIONS

18

---

READS

12

3 AUTHORS, INCLUDING:



Jaewoon Jung

RIKEN

18 PUBLICATIONS 127 CITATIONS

SEE PROFILE



Jooyoung Lee

Korea Institute for Advanced Study

120 PUBLICATIONS 3,890 CITATIONS

SEE PROFILE

# Topological Determinants of Protein Unfolding Rates

Jaewoon Jung,<sup>1</sup> Jooyoung Lee,<sup>2\*</sup> and Hie-Tae Moon<sup>1</sup>

<sup>1</sup>Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon, Korea

<sup>2</sup>School of Computational Sciences, Korea Institute for Advanced Study, Seoul, Korea

**ABSTRACT** For proteins that fold by two-state kinetics, the folding and unfolding processes are believed to be closely related to their native structures. In particular, folding and unfolding rates are influenced by the native structures of proteins. Thus, we focus on finding important topological quantities from a protein structure that determine its unfolding rate. After constructing graphs from protein native structures, we investigate the relationships between unfolding rates and various topological quantities of the graphs. First, we find that the correlation between the unfolding rate and the contact order is not as prominent as in the case of the folding rate and the contact order. Next, we investigate the correlation between the unfolding rate and the clustering coefficient of the graph of a protein native structure, and observe no correlation between them. Finally, we find that a newly introduced quantity, the impact of edge removal per residue, has a good overall correlation with protein unfolding rates. The impact of edge removal is defined as the ratio of the change of the average path length to the edge removal probability. From these facts, we conclude that the protein unfolding process is closely related to the protein native structure. *Proteins* 2005;58:389–395. © 2004 Wiley-Liss, Inc.

**Key words:** protein folding; folding–unfolding rate; protein structure; contact order; clustering coefficient; impact of edge removal

## INTRODUCTION

In the past three decades, there have been growing efforts in studying and understanding the principles that govern the folding mechanism of proteins.<sup>1,2</sup> In particular, the problem of protein folding kinetics is an important issue, and many scientists have tried to seek the quantities that determine the protein folding and unfolding kinetics. However, it has not yet been clearly demonstrated what the most important quantity in folding kinetics is. The early ideas were based on the assumption that there exist specific pathways for folding, and that polypeptide chains can fold into their native structures without searching the whole conformational space.<sup>3</sup> As more data become available, various theories have emerged, such as the framework model, the hydrophobic collapse model, the diffusion collapse model, and so on.<sup>4–6</sup> Some theoretical studies suggested that the length of a protein sequence is an important factor determining the folding rate and mechanism,<sup>7–10</sup> while other studies suggested

that the stability of a protein determines its folding rate.<sup>11–15</sup>

Recently, experimental data of many two-state folding proteins became available, and many investigations have been focused on the wide range of folding–unfolding rates among these proteins.<sup>5,16–33</sup> Based on these data, a simple insight was provided by the work of Plaxco et al.,<sup>34,35</sup> Baker,<sup>36</sup> and Fersht,<sup>37</sup> where a newly introduced quantity, the contact order of protein native structures, was found to have a strong correlation with experimental folding rates.

Along with the protein folding rate, many research efforts have been invested to understand the unfolding rate of a protein. The protein unfolding process plays an important role in controlling the function of proteins, since it provides a cell with a mechanism of removing proteins when their activities are no longer required.<sup>6</sup> In general, one may naively expect that the logarithm of the protein unfolding rate is closely related to the stability of the protein.<sup>4</sup> On the other hand, the tertiary structures of native proteins can play roles as important in unfolding processes as in folding processes, and there are various topological quantities one can construct from the three-dimensional (3D) structures of proteins.

In this article, we investigate various topological quantities from the native structures of two-state-folding proteins and their relationships to the protein unfolding rate. First, we study the relationship between the protein unfolding rate and the contact order of the protein's native structure, and examine if there exists a high correlation between them, as in the case of the folding rate and the contact order. Second, to study various topological properties of the native structure of a protein, a graph corresponding to the native structure is constructed, which is called the protein contact network. Since the two major conventional quantities in the graph/network theory that represent the topological properties of a graph are the clustering coefficient and the average path length, we have calculated them to investigate their relationships with the protein unfolding rate.

Grant sponsor: MOST (J. Jung and H.-T. Moon). Grant sponsor: Basic Research Program of the Korean Science and Engineering Foundation; Grant number: R01-2003-000-11595-0 (J. Lee).

\*Correspondence to: Jooyoung Lee, School of Computational Science, Korea Institute for Advanced Study, 207-43 Cheongryangri-dong, Dongdaemun-gu, Seoul 130-722, Korea. E-mail: jlee@kias.re.kr

Received 26 April 2004; Accepted 2 August 2004

Published online 19 November 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20324

**TABLE I. Thermodynamic and Kinetic Data of 22 Two-State Folding Proteins**

<i>Class</i>	<i>Proteins</i>	<i>PDB</i>	<i>N</i> <sup>a</sup>	$\log k_F$ <sup>b</sup>	$\log k_U$ <sup>b</sup>	$\Delta G_{U-F}$ <sup>c</sup>
$\alpha$ -helical	$\lambda$ repressor	1LMB	80	3.6902	1.4769	3.0
	ACBP	2ABD	86	2.4456	-3.9993	7.1
	Cytochrome c	1HRC	104	3.4472	-1.7692	6.9
	Im9	1IMP	86	3.1614	-1.9062	6.3
$\beta$ proteins	SH3 <sup>d</sup>	1PKS	76 (84)	-0.4559	-3.1734	3.4
		1AEY	62 (56)	0.6127	-1.3468	2.9
		1SHF	58 (67)	1.9731	-3.0038	6.0
		1SRL	56 (64)	0.9085	-0.9998	4.1
	Cold-Shock proteins	1CSP	67	3.0294	1.0792	3.0
		3MEF	69	2.3010	0.6232	3.0
		1C9O	66	2.5682	-4.3468	4.8
		1G6P	66	2.7520	-0.1938	6.3
		2AIT	74	1.8261	-1.7447	8.1
		Tendamistat				
$\beta$ -sandwich domains	TNfn3	1TEN	90	0.4624	-2.5524	4.8
	T1 127	1TIT	89	1.5052	-3.3092	7.5
	TWlg18	1WIT	93	0.1761	-3.5522	4.0
	CI2	2CI2	64	1.6812	-3.7448	7.0
$\alpha/\beta$ proteins	Muscle AcP	1APS	98	-0.6383	-3.9579	5.4
	Hpr	1HDN	85	1.1731	-2.6773	4.6
	ADAh2	1PBA	80	2.9528	-0.1871	4.1
	U1A	1URN	102 (96)	2.4997	-4.2007	9.3
	FKBP12	1FKB	107	0.6335	-3.7689	5.5

<sup>a</sup>*N* in the parenthesis is the size of a protein in the number of amino acids in the Protein Data Bank (PDB) file.

<sup>b</sup> $k_F$  and  $k_U$  are the values of folding and unfolding rates extrapolated to the absence of a denaturant.

<sup>c</sup> $\Delta G_{U-F}$  is the stability of the native structure measured from experiments.

<sup>d</sup>Atomic coordinates of the SH3 domain proteins are all incomplete.

In addition, we have paid a special attention to the change of the average path length and its relationship to the protein unfolding rate. If we assume that the protein unfolding process corresponds to the incremental edge breaking in the contact network, the change of the average path length is more relevant in unfolding than the average path length itself. We name this quantity the impact of edge removal per residue.

From these relationships, we find that the clustering coefficient has no correlation with the unfolding rate, and that the impact of edge removal per residue has a good correlation with the unfolding rate.

## METHODS AND RESULTS

### Experimental Data for Two-State Folding Proteins

We consider proteins that fold by two-state kinetics. The only detectable states for these proteins, on their folding-unfolding pathways, are the unfolded (denatured) and the folded (native) states. In principle, there could exist kinetically undetectable intermediate states that are high in energy.<sup>5</sup>

For two-state folding proteins, the logarithm of unfolding rate is linear to the concentration of a denaturant

$$\log k_U = \log k_U^{H_2O} + m_{k_U}[\text{denaturant}], \quad (1)$$

where  $k_U^{H_2O}$  is the value of the unfolding rate extrapolated to the absence of the denaturant, and  $m_{k_U}$  is a constant of proportionality.<sup>5</sup>

In Table I, we list the two-state folding proteins, along with their folding-unfolding rates and stabilities. The

proteins are  $\alpha$ -helical proteins,  $\beta$  proteins,  $\beta$ -sandwich proteins, and  $\alpha/\beta$  proteins.

### Construction of Protein Contact Networks

A protein contact network is constructed from the geometry of a protein native structure. To study the topology of a protein structure, we construct a graph corresponding to the protein conformation in which nodes represent amino acid residues and edges represent the pairs of amino acid residues in contact. The pairwise contact is defined to exist between amino acid residues  $i$  and  $j$  when the  $C_\alpha$  distance between them is less than the cutoff distance.<sup>38,39</sup> From the graph, we can investigate various quantities that characterize the graph. Figure 1 shows the contact network of the protein 2ABD.

### Contact Order and Its Relation With Folding-Unfolding Rates

Before investigating relationships between the protein unfolding rate and various topological quantities of protein contact networks, we examine the relationship between the unfolding rate and the contact order. The contact order of a protein is one of the quantities that is highly correlated with its folding rate. The contact order is defined as

$$CO = \frac{1}{LN} \sum_{i,j}^N |i - j|, \quad (2)$$

where  $N$  is the total number of contacts,  $|i - j|$  is the sequence separation between residues  $i$  and  $j$  in contact,

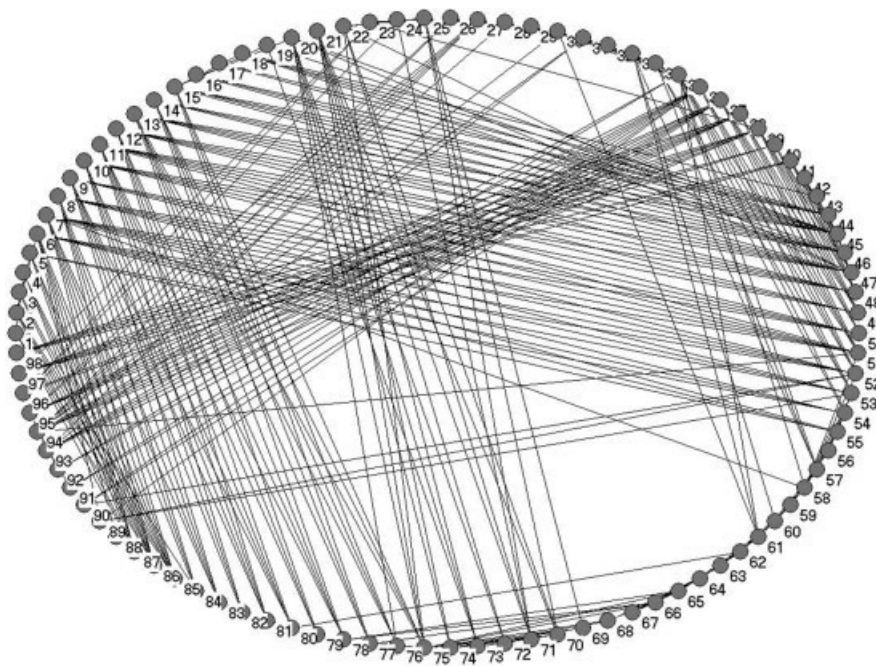


Fig. 1. The protein contact network of 2ABD. Edges are drawn between all pairs of amino acid residues with their  $C_{\alpha}$  distances less than the cutoff distance.

and  $L$  is the size of the protein in the number of amino acid residues. The contact order reflects the extent of local and nonlocal contacts of a native protein structure and is known to be closely related to the folding rates of two-state folding proteins.<sup>34</sup> This high correlation can be understood if one assumes that the topology of the transition state resembles the topology of the native state, and that the folding rate is related to the size and the configurational entropy of a loop.<sup>37</sup> In Figure 2, folding rates and contact order are shown for proteins listed in Table I, and we confirm the overall high correlation of 0.77 with a  $p$  value (probability of achieving the correlation by random chance) less than 0.0001.

On the other hand, the correlation (0.31, with  $p$  value = 0.17) between the unfolding rate and the contact order, as shown in Figure 3, is not as strong as the correlation between folding rate and contact order. From these results, we may expect that there are other quantities strongly related to unfolding rates.

### Clustering Coefficients and the Average Path Length

Here we consider the clustering coefficient of the graph constructed from a native structure. Along with the average path length of a graph, the clustering coefficient is an important quantity specifying the graph. It represents the average fraction of the pairs of neighbors (if there exists an edge between 2 nodes, we say that they are neighbors to each other) that are also neighbors to each other.

The clustering coefficient is defined as

$$C = \frac{1}{N} \sum_k \frac{n_k}{N_k(N_k - 1)/2}, \quad (3)$$

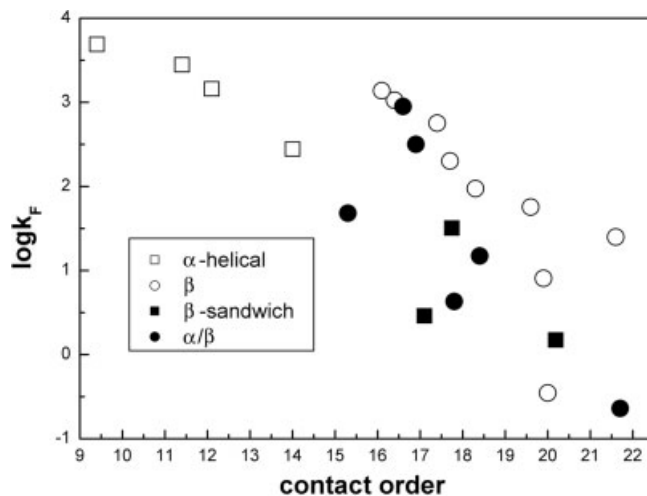


Fig. 2. The relationship between the folding rate  $k_f$  and the contact order constructed from the protein native structure. For all proteins, folding rates are highly correlated with their relative contact orders [correlation = 0.77,  $p$  (the probability of achieving the correlation by random chance) < 0.0001].

where  $n_k$  is the number of edges among the neighbors of node  $k$  and  $N_k$  is the neighbor number of node  $k$ . The maximum number of possible edges between these  $N_k$  neighbors is  $N_k(N_k - 1)/2$  and the ratio  $\frac{n_k}{N_k(N_k - 1)/2}$  denotes the fraction of the existing edges over the maximum number.<sup>40</sup> In a protein structure, the clustering coefficient provides the measure of the extent to which 2 residues, interacting with an identical third residue, also interact each other. The more there exist instances that more than 3 amino acids are in contact to one another, the



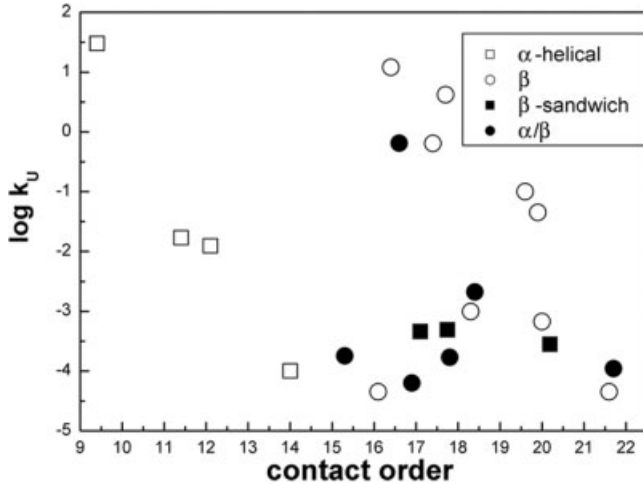


Fig. 3. The relationship between  $\log k_U$  and the contact order. When we consider all proteins, the correlation is poor compared to the corresponding correlation with the folding rate (correlation = 0.31,  $p = 0.17$ ).

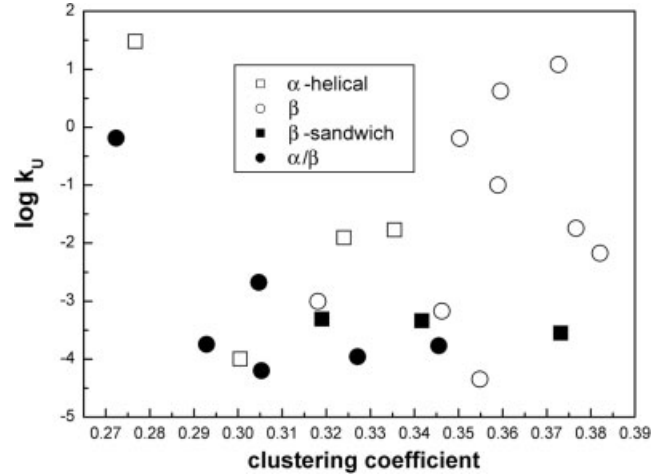


Fig. 4. The relationship between  $\log k_U$  and the clustering coefficient. When all proteins are considered, we do not see much correlation, but when  $\beta$  proteins (denoted by  $\circ$ ) are excluded, the correlation is 0.52,  $p = 0.07$ .

larger the clustering coefficient becomes. We can also calculate the clustering coefficient as

$$C = \frac{3 \times (\text{Number of triangles})}{\text{Number of connected triples}}, \quad (4)$$

where the triangles are the trios of nodes in which each node is connected to both of the other two nodes, and the connected triples are the trios in which at least one is connected to the other two. From this formula, we expect that a protein contact network would have a high value of clustering coefficient when there are a large number of triangles in the network.

The average path length of a graph is defined as

$$L = \frac{1}{N(N-1)} \sum_{k \neq j=1}^N l_{kj} \quad (5)$$

where  $l_{kj}$  is the minimum number of edges one must pass through to reach node  $j$  from node  $k \neq j$ . This quantity measures the extent to which each amino acid residue influences, on average, the rest of the residues of a protein. The more compact the structure of a protein is, the shorter its average path length becomes.

In Figure 4, the unfolding rate  $k_U$  versus the clustering coefficient is shown. When all proteins in Table I are considered, there exists little correlation between the clustering coefficient and  $\log k_U$ . But, when  $\beta$  proteins are excluded, the correlation is 0.52, with  $p$  value = 0.07.

The purpose of this investigation is to find, in the case of *unfolding*, a topological quantity that plays the role of the contact order in folding (which is highly correlated with the *folding* rates of many two-state folding proteins). As stated above, the clustering coefficient is not the universal determinant of the unfolding rates for proteins of various structures. In addition, we have considered the relationship between the average path length and the unfolding rate, and no significant correlation is observed.

### Impact of Edge Removal of a Protein Contact Network

In addition to the contact order and the clustering coefficient, we have considered another topological quantity, which we call as the impact of edge removal per residue of a protein contact network. The concept of the impact of edge removal is similar to that of the error tolerance or the robustness of a network.<sup>41</sup> The error tolerance is determined by the change of the average path length when randomly selected nodes are removed from a network. A network is considered to have a high value of error tolerance if its average path length increases slowly upon the removal of *nodes*. It is shown that scale-free networks have higher values of error tolerance compared to exponential networks. Here, with graphs constructed from protein native structures, we introduce a concept similar to the error tolerance, or the robustness. Since *nodes* (amino acid residues) cannot be removed by unfolding processes (i.e., peptide bonds are not broken), we consider the removal of *edges* (contacts). From a protein contact network, we measure the change of the average path length as a function of the edge removal probability. The edge removal probability is the probability that a particular edge is removed from the graph constructed by a native protein structure.

During the unfolding process of proteins, physical pairwise distances between two nodes increase, which results in removing native contacts. So we interpret the unfolding phenomena in terms of the edge removal process from a protein contact network. By the edge removal, some graphs may lose their characteristics more easily than others. Since the main characteristics of graphs are the average path lengths, we consider the change of the average path length as a function of the edge removal probability. In Figure 5, we show the change of the average path length for  $\alpha$ -helical proteins 1LMB and 2ABD. In Figure 6, the change of the average path length per residue, which is the

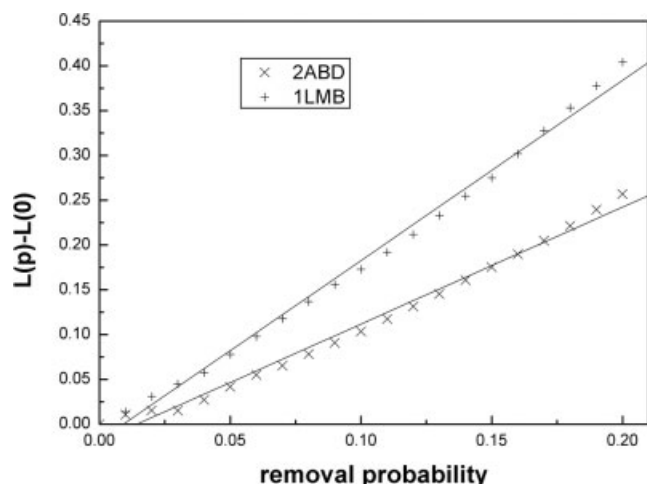


Fig. 5. The relationship between the change of the average path length and the edge removal of probability. The edge removal probability is the probability that a particular edge is removed, and  $L(p) - L(0)$  is the change of the average path length when edges are removed with probability  $p$  (i.e.,  $L(p)$  is the average path length when edges are removed with removal probability  $p$ ). The change of the average path length of 1LMB is greater than that of 2ABD, which indicates that the structure of the protein 2ABD is more conserved than that of 1LMB upon the same value of edge removal probability.

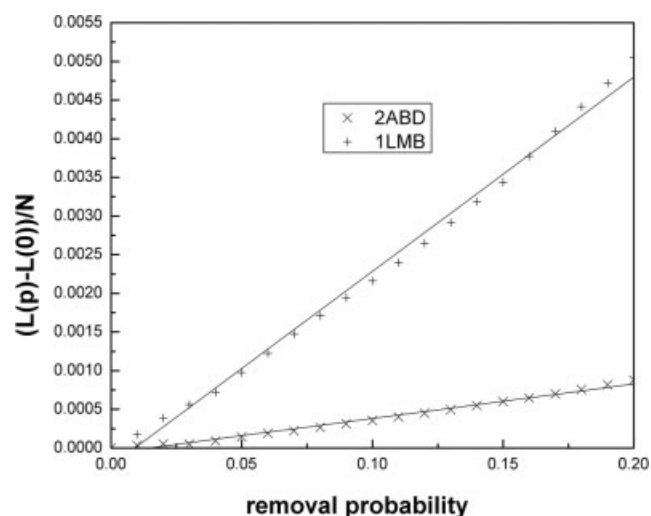


Fig. 6. The same as Figure 5, but the y axis is divided by the size of the protein in the number of amino acids.

change of the average path length divided by the size of the protein in the number of amino acids, is shown. We observe that the change of the average path length and the change of the average path length per residue are approximately linear to the edge removal probability for small values of the removal probability. If the values of the removal probability are too high, nodes are divided into separate clusters with small sizes, and we do not consider these cases. From the linear relationship, we measure the slope of the linear line. In Figures 5 and 6, we observed that the slope of 1LMB is larger than that of 2ABD (i.e., 1LMB has a higher impact of edge removal per residue than 2ABD). This indicates that the graph constructed

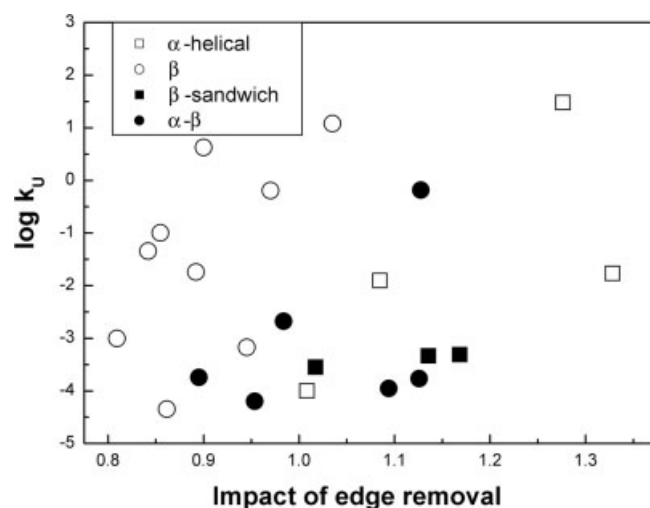


Fig. 7. The relationship between unfolding rate and impact of edge removal. Here, impact of edge removal is defined as the slope of the regression line in Figure 5.

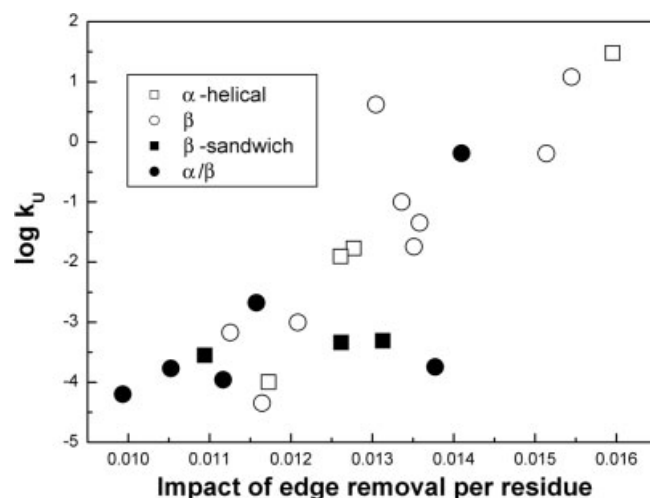


Fig. 8. The relationship between unfolding rate and the impact of edge removal per residue. The impact of edge removal per residue is defined as the slope of the regression line in Figure 6, which has a correlation value of 0.83,  $p < 0.0001$ . Here, the cutoff distance of 8.0 Å is used.

from 1LMB loses its integrity more rapidly than the graph of 2ABD upon the same amount of edge removal. Here we define the impact of edge removal and the impact of edge removal per residue as the slopes of the regression lines in Figures 5 and 6, respectively.

Since the unfolding process can be viewed as the process of losing native contacts, we expect that proteins with higher values of the impact of edge removal have higher values of unfolding rates (unfold easily). Based on this conjecture, for all proteins in Table I, we plot the logarithm of  $k_u$ , as a function of the impact of edge removal and the impact of edge removal per residue in Figures 7 and 8, respectively. We observe significantly better correlations between  $\log k_u$  and the impact of edge removal per residue as compared to Figures 3 and 4. The correlation value is 0.83, with  $p$  value less than 0.0001.

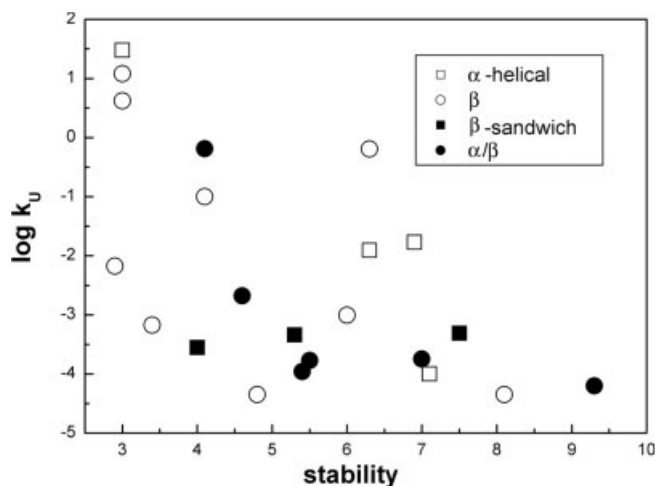


Fig. 9. The relationship between the unfolding rate and the protein stability. Correlation = 0.55,  $p = 0.010$ .

## DISCUSSION

We have investigated the relationships between the logarithms of the unfolding rates  $\log k_U$  and the various topological quantities obtained from the graphs of protein native structures, and observed that the unfolding rate is closely related to the impact of edge removal per residue calculated from the contact network of a protein native structure.

Contrary to the case of the folding rate, the unfolding rate shows much less correlation with the contact order. This may imply that unfolding processes may deviate from folding processes for some proteins. In folding processes, long-range contacts take a longer time to be formed than short-range contacts; consequently, the formation of these long-range contacts can be the rate-limiting step in folding processes. However, in unfolding processes, breaking long-range contacts may not be as important as in folding processes (see Figs. 2 and 3).

As in the case of the contact order, the unfolding rate shows little correlation with the clustering coefficient. And the correlation is better when  $\beta$  proteins are excluded. This means that clustering coefficient cannot serve as a universal topological quantity for unfolding rates.

The impact of edge removal per residue is shown to be highly correlated with unfolding rates for all 22 two-state folding proteins listed in Table I. If we assume that the contact network of the transition state in unfolding corresponds to a network where a certain fraction of native contacts are removed from the native contact network, the impact of edge per residue removal would play an important role to determine the unfolding rate. The average path length of a network increases as edges are removed from the network. Thus, the bigger the impact of edge removal per residue is, the less robust the native contact network upon the random removal of contacts; hence, the protein unfolds more easily.

From the relationships with the unfolding rates, we conjecture that the free-energy barrier of unfolding is closely related to the protein native structure based on the

high correlation between the unfolding rate and the structural characteristic, namely, the the impact of edge removal per residue.

Finally, the relation between  $\log k_U$  and the protein stability is shown in Figure 9. We observe a smaller amount of correlation (0.55, with  $p = 0.010$ ) between them as compared to the case of the impact of edge removal per residue. Another quantity to consider is the compactness of a protein. For this, we studied the correlation between the total number of contacts and the unfolding rate, and the measured correlation is 0.50, with  $p = 0.02$ .

## CONCLUSIONS

We have investigated the relationships between various quantities related to the protein native structures and unfolding rates. By constructing the graph of a native protein structure, we could find quantities that are closely related to unfolding rates. In particular, the impact of edge removal per residue, which is the ratio of the change of the average path length divided by the protein's size to edge removal probability, was shown to have a high correlation with the protein unfolding rate. From these facts, we conclude that the protein unfolding process is closely related to the protein native structure.

## REFERENCES

1. Anfinsen C. Principles that govern the folding of protein chains. *Science* 1973;181:223–230.
2. Branden C, Tooze J. Introduction to protein structure. New York: Garland; 1991.
3. Levinthal C. Are there pathways for protein folding? *J Chem Phys* 1968;65:44–45.
4. Creighton TE. Protein folding. New York: W. H. Freeman; 1992.
5. Fersht AR. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. New York: W. H. Freeman; 1999.
6. Pain RH. Mechanisms of protein folding. New York: Oxford University Press; 2000.
7. Wolynes PG. Folding funnels and energy landscapes of larger proteins within the capillarity approximations. *Proc Natl Acad Sci USA* 1997;94:6170–6175.
8. Finkelstein AV, Badretidinov AY. Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold. *Fold Des* 1997;2:115–121.
9. Kilmov DK, Thirumalai D. Factors governing the foldability of proteins. *Proteins* 1997;26:411–441.
10. Gutin AM, Abkevich VI, Shakhnovich EI. Chain length scaling of protein folding time. *Phys Rev Lett* 1996;77:5433–5436.
11. Finkelstein AV. Rate of beta-structure formation in polypeptides. *Proteins* 1991;9:23–27.
12. Sali A, Shakhnovich EI, Karplus M. How does a protein fold? *Nature* 1994;369:248–251.
13. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 1995;21:167–195.
14. Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND. Toward an outline of the topography of a realistic protein folding funnel. *Proc Natl Acad Sci USA* 1995;92:3626–3630.
15. Pande VS, Grosberg AY, Tanaka T. On the theory of folding kinetics for short proteins. *Fold Des* 1997;2:109–114.
16. Jackson SE. How do small single-domain proteins fold? *Fold Des* 1998;3:R81–R91.
17. Huang GS, Oas TG. Structure and stability of monomeric gamma repressor: NMR evidence for two-state folding. *Biochemistry* 1995;34:3884–3892.
18. Burton RE, Huang GS, Daugherty MA, Fullbright PW, Oas TG. Microsecond protein folding through a compact transition state. *J Mol Biol* 1996;263:311–322.
19. Kragelund BB, Robinson CV, Knudsen J, Dobson CM, Poulsen

- yFM. Folding of a four-helix bundle: studies of acyl-coenzyme A bind protein. *Biochemistry* 1995;34:7217–7214.
20. Kragelund BB, Poulsen FM. Fast and one-step folding for closely and distantly related homologous proteins of a four-bundle family. *J Mol Biol* 1996;256:187–200.
  21. Chan CK, Hofrichter J. Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc Natl Acad Sci USA* 1997;94:1779–1784.
  22. Schindler T, Herrler M, Marahiel MA, Schmid FX. Extremely rapid folding in the absence of intermediates. *Nat Struct Biol* 1995;2:663–673.
  23. Viguera AR, Martinez J, Fillmonov V, Mateo P, Serrano L. Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a 2-state folding transition. *Biochemistry* 1994;33:2142–2150.
  24. Viguera AR, Serrano L, Wilmanns M. Different folding transition-state may result in the same native structure. *Nat Struct Biol* 1996;3:874–880.
  25. Grantcharova VP, Baker D. Folding dynamics of the src SH3 domain. *Biochemistry* 1997;36:15685–15692.
  26. Guijarro JI, Morton CJ, Plaxco KW, Campbell ID, Dobson CM. Folding kinetics of the SH3 domain of PI3 kinase by real-time NMR combined with optical spectroscopy. *J Mol Biol* 1998;276:657–667.
  27. Plaxco KW, Dobson CM. The folding kinetics and thermodynamics of the Fyn-SH3 domain. *Biochemistry* 1998;37:2529–2537.
  28. Ferguson N, Capaldi AP, James R, Kleanthous C, Radford SE. Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J Mol Biol* 1999;286:1597–1608.
  29. Clarke J, Hamill SJ, Johnson CM. Folding and stability of a fibronectin type3 domain of human tenascin. *J Mol Biol* 1997;270:771–778.
  30. Clarke J, Cota E, Fowler SB, Hamill SJ. Folding studies of immunoglobuline-like beta-sandwich proteins suggest that they share a common folding pathway. *Structure* 1999;7:1145–1153.
  31. Villegas V, Azuaga A, Catusus LI, Reverter D, Mateo PL, Aviles FX, Serrano L. Evidence for a two-state transition in the folding process of the activation domain of human procarboxypeptidase A2. *Biochemistry* 1995;34:15105–15110.
  32. Nuland V, Meijberg NAJ, Warner J, Forge V, Scheek RM, Robillard GT, Dobson CM. Slow cooperative folding of a small globular protein HPr. *Biochemistry* 1998;37:622–637.
  33. Friesner RA. Computational methods for protein folding. New York: Wiley; 2001.
  34. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 1998;277:985–994.
  35. Plaxco KW, Simons KT, Ruczinski I, Baker D. Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry* 2000;39:11177–11183.
  36. Baker D. A surprising simplicity to protein folding. *Nature* 2000;405:39–42.
  37. Fersht AR. 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* 2000;97:1525–1529.
  38. Dokholyan NV, Li L, Shakhnovich EI. Topological determinants of protein folding. *Proc Natl Acad Sci USA* 2002;99:8637–8641.
  39. Bendruscolo M, Paci E, Dobson CM, Karplus M. Three key residues from a critical contact network in a protein folding transition state. *Nature* 2001;409:641–645.
  40. Watts DJ, Strogatz SH. Dynamics of “Small-World” networks. *Nature* 1998;393:440–442.
  41. Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. *Nature* 2000;406:378–381.