

Recent Topics in Protein Folding

Takeshi Kikuchi*

Department of Bioinformatics, College of Life Sciences, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan

Short Communication

The protein folding is a long-standing problem and there have been a lot of experimental and theoretical/computational studies so far [1-3]. These efforts by many researches have reveals mysterious folding mechanisms of proteins. However, there are still several unsolved problems on protein folding. An interesting one is how the information of the folding mechanism of a protein is encoded in its amino acid sequence. In the present short article, I discuss the remaining interesting problems on protein folding.

One interesting and yet unsolved problem is the folding of GA- and GB- domains. In general, the 3D structures or the topologies of two proteins are similar if these proteins share more than 30% sequence identity. However, protein sequences which do not follow this empirical rule have been designed. It is known that the GA- and GB-domains bind to human serum albumin (HSA) [4,5] and the constant (Fc) region of IgG [6,7] respectively. Alexander et al. [8] and He et al. [9] designed two proteins from sequences of GA-domain and GB-domain.

These proteins share about 60% sequence identity but different structures, that is, one is a 3 α -helix bundle and the other is a 4 β -sheet + α -helix structures using the phage display technique to introduce mutations. The 3D structures of GA- and GB-domains are shown in Figure 1a (PDB codes of GA- and GB-domains are 1FS1 and 1PGA respectively).

He et al. [10] further designed another sequences with 98% identity but exhibiting different 3D structures. 98% identity means that two related sequences differ by only one amino acid in 56 residues. The PDB codes of these proteins are 2LHG (GA domain related) and 2LHE

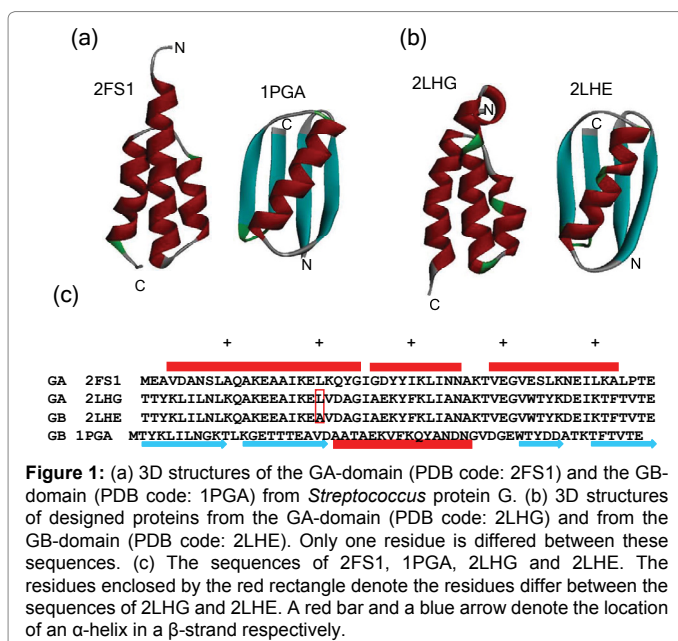


Figure 1: (a) 3D structures of the GA-domain (PDB code: 2FS1) and the GB-domain (PDB code: 1PGA) from *Streptococcus* protein G. (b) 3D structures of designed proteins from the GA-domain (PDB code: 2LHG) and from the GB-domain (PDB code: 2LHE). Only one residue is differed between these sequences. (c) The sequences of 2FS1, 1PGA, 2LHG and 2LHE. The residues enclosed by the red rectangle denote the residues differ between the sequences of 2LHG and 2LHE. A red bar and a blue arrow denote the location of an α -helix in a β -strand respectively.

(GB domain related) (Figure 1b and c). In Figure 1c, the positions of α -helices and β -strands in 2FS1 and/or 1PGA are presented. The positions of the central α -helices in 2LHG and 2LHE are almost same. From this figure, it is notified that the location of the central α -helix is similar in 3 α and in 4 β + α . The α -helix in 4 β + α is somewhat longer than that in 3 α . Furthermore, I present the hydrophobic packing around 20-L in 2LHG and around 20-A in 2LHE, that is, the different residues between these proteins in Figure 2a and b.

In 2LHG, the sufficient hydrophobic packing is observed around 20-L, while only one residue, 25-I, makes hydrophobic contact with 20-A in 2LHE. (Two residues are supposed to make a contact when a heavy atom in a residue is close to a heavy atom in another residue within 5Å.) This difference in the hydrophobic packing may produce the difference in the 3D structures.

These facts urge us to reconsider the relationship between the sequence identity and the structural similarity of two proteins. Many researchers including computational works on GA- and GB-domain

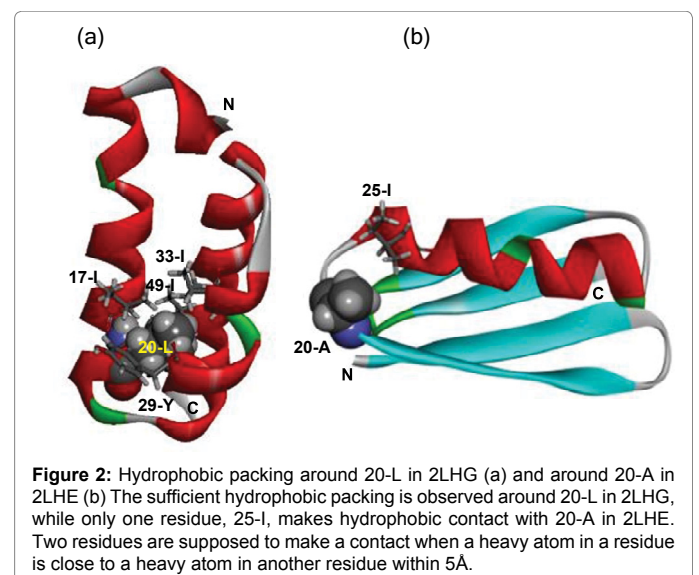


Figure 2: Hydrophobic packing around 20-L in 2LHG (a) and around 20-A in 2LHE (b) The sufficient hydrophobic packing is observed around 20-L in 2LHG, while only one residue, 25-I, makes hydrophobic contact with 20-A in 2LHE. Two residues are supposed to make a contact when a heavy atom in a residue is close to a heavy atom in another residue within 5Å.

*Corresponding authors: Takeshi Kikuchi, Department of Bioinformatics, College of Life Sciences, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan, Tel: +81-77-561-5909; Fax: +81-77-561-2659; E-mail: tkikuchi@sk.ritsumei.ac.jp

Received February 15, 2018; Accepted March 16, 2018; Published March 23, 2018

Citation: Kikuchi T (2018) Recent Topics in Protein Folding. J Proteomics Bioinform 11: 075-078. doi: [10.4172/jpb.1000469](https://doi.org/10.4172/jpb.1000469)

Copyright: © 2018 Kikuchi T. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

formations of these homologous structures pointed out the difficulty of this problem. That is, this is a very challenging problem. Our recent studies [11,12] indicate that in the respective structures different residues form contacts during the early stage of folding. This may suggest that the respective structures form depending whether packing of hydrophobic residues is dominant or secondary structure formation is dominant in the early stage of folding.

Next I consider the following problem. That is, similar partial structures appear in the protein structure space.

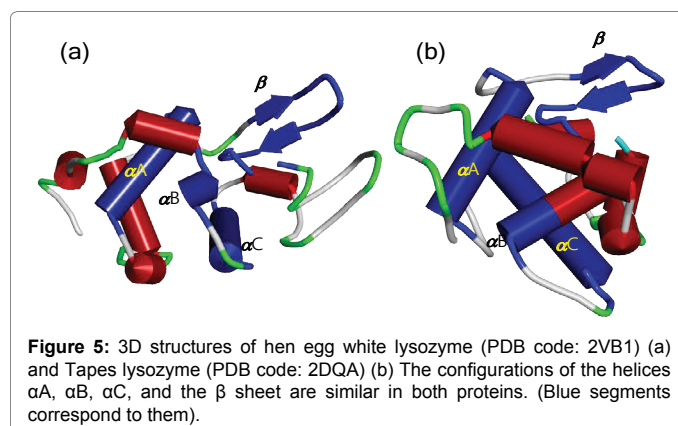
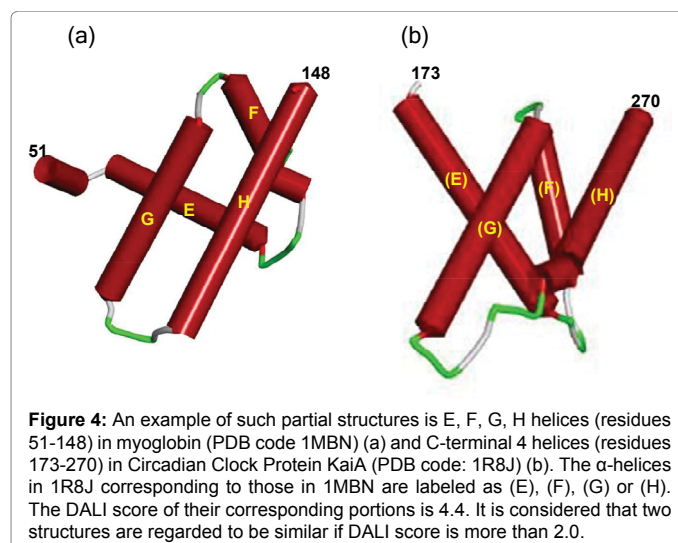
The existence of “foldon” has been recognized by many authors. For example, Rooman et al. [13,14] pointed out that some short segments in a protein correspond well to the segments which fold in the early stage of folding. Panchenko et al. [15,16] defined a foldon as a segment exhibiting maximum “foldability” based on the ratio of stability of the native structure of a given protein and root mean square fluctuation of the energy of the non-native structures of this protein.

Foldons in proteins from different folds sometimes show similar structures [15]. Figure 3a and b shows the example of a pair of foldons defined by Panchenko et al. [15] of gamma-II crystalline 1-25 and myoglobin 65-91 and rmsd of this part is 3.6Å. It is interesting whether the folding mechanisms of such similar foldon structures are also similar.

Besides foldons, Kister et al. [17,18] indicated the common regular structures in β-sandwich proteins. They call partial segments forming such regular structures key strands. The partial structure formed by alpha-helices E, F, G and H appeared in the Globin-like fold proteins is also observed in proteins beyond the Globin-like fold [19]. The common hydrophobic packing is recognized in such partial structures. An example of such partial structures is E, F, G, H helices (residues 51-148) in myoglobin (PDB code 1MBN) and C-terminal 4 helices (residues 173-270) in Circadian Clock Protein KaiA (PDB code: 1R8J). These structures are presented in Figure 4a and b.

The DALI score [20] of their corresponding portions is 4.4. It is considered that two structures are regarded to be similar if DALI score is more than 2.0. The lysozyme-like fold proteins exhibit various 3D structures but partial structures are very similar and such similar structures exhibit the common function of lysozyme proteins. I present the similarity of the partial structures in hen egg white lysozyme (PDB code: 2VB1) and Tapes lysozyme (PDB code: 2DQA) in Figure 5a and b.

Hen lysozyme and Tapes lysozyme are classified into c-type and i-type superfamilies respectively. The comparison of these two structures reveals that the whole structures differ considerably, but the configurations of the helices α-A, β-B, α-C, and the β sheet are similar in both proteins as shown in Figure 5. (Blue segments correspond to them.) We demonstrated [21] the common hydrophobic interactions



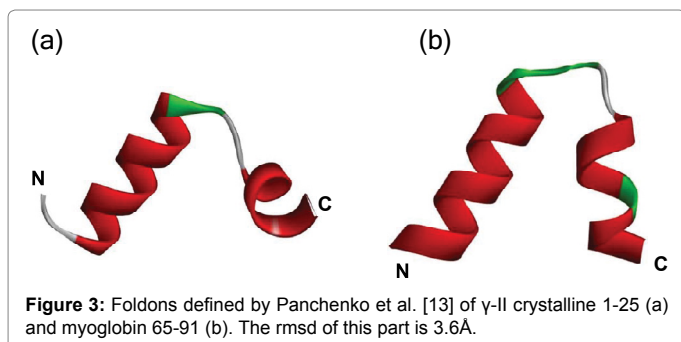
connecting the evolutionary conserved folding units for lysozyme-like fold proteins. These interactions seem to be essential to form similar partial structures. If some common property of amino acid sequences of common partial structures is identified, modeling of the structure of a partial sequence will be possible. Such partial structures will be used to construct a whole protein structure as the fragment assembly technique by Baker and coworkers [22].

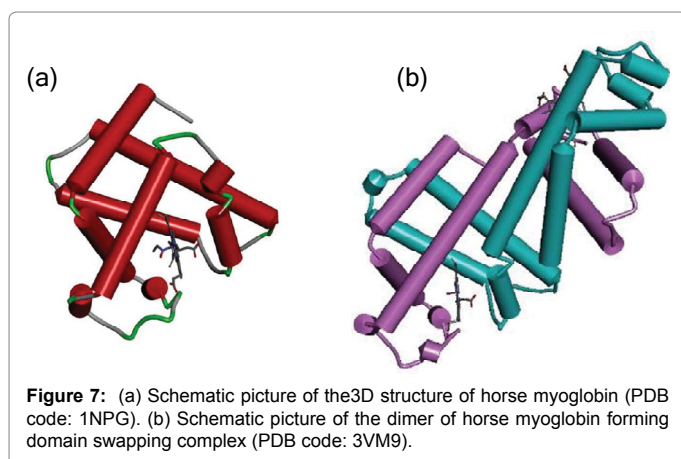
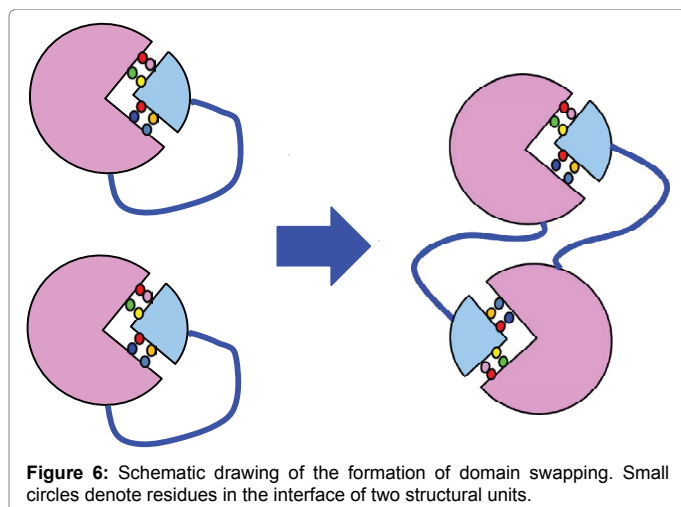
The other problem is domain swapping. Phenomena of domain swapping have been observed by many authors [23-25]. Domain swapping is an oligomer formation by two or more identical proteins exchanging structural units as schematically shown in Figure 6.

Colored small circles in Figure 6 denote residues at the interface of two structural units. The interactions between residues are expected to be maintained after the formation of a dimer. The earliest example of domain swapping was the case of RNase A [26]. The first 3D structure determination of domain swapped dimer was dimeric diphtheria toxin [27,28]. Figure 7a and b presents an example of domain swapping formed by two horse myoglobins.

Furthermore, domain swapping mechanisms probably relate to the mechanisms of regulation of protein functions, molecular diseases and aggregation. The increase of its function by forming domain swapped structure of RNase A was reported [29].

It is expected that the domain swapping mechanism of a protein relates to the mechanism of the protein folding. However, many of





the details of mechanisms of domain swapping are still unclear. We are clarifying so far that folding of a protein occurs at the evolutionary conserved hydrophobic residues forming high contact frequency. The frequency of contacts in the early stage of folding can be predicted by an inter-residue potential derived from inter-residue average distance statistics [11,12,19,21,30]. It would be possible to design a new domain swapped protein dimer if the knowledge on protein folding is utilized. Furthermore, it may be also possible to design a domain swapped protein dimer with a new function.

As shown in this short article, still many problems to be solved are remaining on protein folding and new knowledge will be further accumulated on the protein folding problem. The utilization of such knowledge about folding reveals new possibility of new applications.

References

1. Compiani M, Capriotti E (2013) Computational and Theoretical Methods for Protein Folding. *Biochemistry* 52: 8601-8624.
2. Englander SW, Mayne L (2014) The nature of protein folding pathways. *Proc Natl Acad Sci USA* 111: 15873-15880.
3. Dill K, Ozkan SB, Shell MS, Weikl TR (2008) The Protein Folding Problem. *Annu Rev Biophys* 37: 289-316.
4. Falkenberg C, Bjork L, Åkerstrom B (1992) Localization of the binding site for streptococcal ProteinG on human serum albumin. Identification of a 55-kilodalton protein G binding albumin fragment. *Biochemistry* 31: 1451-1457.
5. Frick IM, Wikstrom M, Forsen S, Drakenberg T, Gomi H, et al. (1992)

Convergent evolution among immunoglobulin G-binding bacterial proteins. *Proc Natl Acad Sci USA* 89: 8532-8536.

6. Myhre EB, Kronvall G (1997) Heterogeneity of nonimmune immunoglobulin Fc reactivity among gram-positive cocci: Description of three major types of receptors for human immunoglobulin G. *Infect Immun* 17: 475-482.
7. Reis KJ, Ayoub EM, Boyle MDP (1984) Streptococcal Fc receptors. II. Comparison of the reactivity of a receptor from a group C streptococcus with staphylococcal protein A. *J Immunol* 132: 3098-3102.
8. Alexander PA, Rozak DA, Orban J, Bryan PN (2005) Directed evolution of highly homologous proteins with different folds by phage display: Implications for the protein folding code. *Biochemistry* 44: 14045-14054.
9. He Y, Yeh DC, Alexander PA, Bryan PN, Orban J, et al. (2005) Solution NMR Structures of IgG binding domains with artificially evolved high levels of sequence identity but different folds. *Biochemistry* 44: 14055-14061.
10. He Y, Chen Y, Alexander PA, Bryan PN, Orban J, et al. (2012) Mutational tipping points for switching protein folds and functions. *Structure* 20: 283-291.
11. Kikuchi T (2008) Analysis of 3D structural differences in the IgG-binding domains based on the interresidue average-distance statistics. *Amino Acids* 35: 541-549.
12. Matsuoka M, Suguta M, Kikuchi T (2014) Amino acid sequence analyses of proteins with high sequence identity but different 3D structures. *BMC Res Notes* 7: 654.
13. Rooman MJ, Kocher JPA, Wodak SJ (1991). Prediction of protein backbone conformation based on seven structure assignments: influence of local interactions. *J Mol Biol* 221: 961-979.
14. Rooman MJ, Kocher J-PA, Wodak SJ (1992). Extracting information on folding from the amino acid sequence: accurate predictions for protein regions with preferred conformation in the absence of tertiary interactions. *Biochemistry* 31: 10226-10238.
15. Panchenko AR, Luthey-Schulten Z, Cole R, Wolynes PG (1997) The Foldon Universe: A Survey of Structural Similarity and Self-recognition of Independently Folding Units. *J Mol Biol* 272: 95-105.
16. Panchenko AR, Luthey-Schulten Z, Wolynes PG (1996) Foldons, protein structural modules, and exons. *Proc Natl Acad Sci USA* 93: 2008-2013.
17. Kister AE, Finkelstein AV, Gelfand IM (2002) Common features in structures and sequences of sandwich-like proteins. *Proc Natl Acad Sci USA* 99: 14137-14141.
18. Kister AE (2015) Amino Acid Distribution Rules Predict Protein Fold: Protein Grammar for Beta-Strand Sandwich-Like Structures. *Biomolecules* 5: 41-59.
19. Matsuoka M, Fujita A, Kawai Y, Kikuchi T (2014) Similar Structures to the E-to-H Helix Unit in the Globin-Like Fold are Found in Other Helical Folds. *Biomolecules* 4: 268-288.
20. Holm L, Park J (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* 16: 566-567.
21. Nakashima T, Kabata M, Kikuchi T (2017) Properties of Amino Acid Sequences of Lysozyme-Like Superfamily Proteins Relating to Their Folding Mechanisms. *J Proteom Bioinf* 10: 94-107.
22. Handl J, Knowles J, Vernon R, Baker D, Lovell SC, et al. (2012) The dual role of fragments in fragment-assembly methods for de novo protein structure prediction. *80: 490-504.*
23. Liu Y, Eisenberg D (2002) 3D domain swapping: as domains continue to swap. *Protein Sci* 11: 1285-1299.
24. Rousseau F, Schymkowitz J, Itzhaki LS. (2012) Implications of 3D domain swapping for protein folding, misfolding and function. *Adv Exp Med Biol* 747: 137-152.
25. Mascarenhas NM, Gosavi S (2017) Understanding protein domain-swapping using structure-based models of protein folding. *Prog Biophys Mol Biol* 128: 113-120.
26. Crestfield AM, Stein WH, Moore S (1962) On the aggregation of bovin pancreatic ribonuclease. *Arch Biochem Biophys* 1: 217-222.
27. Bennett MJ, Choe S, Eisenberg D (1994) Domain swapping: entangling alliances between proteins. *Proc Natl Acad Sci USA* 91: 3127-3131.
28. Bennett MJ, Choe S, Eisenberg D (1994) Refined structure of dimeric diphtheria toxin at 2.0 Å resolution. *Protein Sci* 3: 1444-1463.

29. Gotte G, Bertoldi M, Libonati M. (1999) Structural versatility of bovine ribonuclease A – Distinct conformers of trimeric and tetrameric aggregates of the enzyme. Eur J Biochem 265: 680-687.
30. Kirioka T, Aumpuchin P, Kikuchi T (2017) Detection of Folding Sites of β -Trefoil Fold Proteins Based on Amino Acid Sequence Analyses and Structure-Based Sequence Alignment. J Proteom Bioinf 10: 222-235.