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 researchers have revealed 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-domain. 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 3α-helix bundle and the other is a 4β-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 2FS1 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 (GB domain related) (Figure 1b and c). In Figure 1c, the positions of α-helices and β-stands 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 noticed that the location of the central α-helix is similar in 3α and in 4β + α. The α-helix in 4β + α is somewhat longer than that in 2FS1. This may reflect the α-helix in 4β + α is more stable than in 3α. 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 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.

**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: +81-77-561-6908; Fax: +81-77-561-2699; E-mail: tkikuchi@sk.ritsumei.ac.jp

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


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 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) (a) and C-terminal 4 helices (residues 173-270) in Circadian Clock Protein KaiA (PDB code: 1R8J) (b). The α-helices in 1R8J corresponding to those in 1MBN are labeled as (E), (F), (G) or (H). The DALI score 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.

Figure 4: An example of such partial structures is E, F, G, H helices (residues 51-148) in myoglobin (PDB code 1MBN) (a) and C-terminal 4 helices (residues 173-270) in Circadian Clock Protein KaiA (PDB code: 1R8J) (b). The α-helices in 1R8J corresponding to those in 1MBN are labeled as (E), (F), (G) or (H). The DALI score 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 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.

Figure 5: 3D structures of hen egg white lysozyme (PDB code: 2VB1) (a) and Tapes lysozyme (PDB code: 2DQA) (b). The configurations of the helices αA, αB, αC, and the β sheet are similar in both proteins. (Blue segments correspond to them).

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
