Ring and zipper formation is the key to understanding the structural variety in all-β proteins

Ryotaro Koike[a,b], Kengo Kinoshita[b,c], Akinori Kidera[b,*]

[a] *Department of Chemistry, Graduate School of Science, Kyoto University, Kitashirakawa-Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan*
[b] *Department of Science of Biological Supramolecular Systems, Graduate School of Integrated Science, Yokohama City University, 1-7-29 Suehirocho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan*
[c] *Structure and Function of Biomolecules, PRESTO, Japan Science and Technology Corporation, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan*

Received 11 September 2002; revised 13 November 2002; accepted 15 November 2002

First published online 2 December 2002

Edited by Robert B. Russell

Abstract **A novel structural classification of β proteins is presented from the viewpoint of the ring-shaped structure and the zipper-like contact pattern, based on the fact that 92% and 60% of β proteins have the ring topology and the zippered contact pattern, respectively. We discuss the implication of the unexpectedly high preference for the ring and zippered structures in connection with the folding process of β proteins.**

Key words: Protein three-dimensional structure; Classification of β proteins; Ring structure; Zippered structure; Relative contact order

1. Introduction

The structural classification of protein domains has mostly been performed by focusing on the similarity in the spatial arrangement and connectivity of the secondary structure elements (SSEs), α-helices and β-strands [1,2]. This view of structural similarity, represented by 'fold' or 'topology', has established good correspondence with the evolutional or functional relationship, and has illuminated distant homologues beyond the sequence similarity [3,4].

In parallel with the fold classification, a number of studies have been devoted to providing a clear perspective on the structural diversity of protein folds from the taxonomic and physicochemical viewpoints [5–7], and various structural rules have been proposed explaining the limited variety of the protein folds [8–10]. Here, we focus our attention on β proteins. Efimov [11] proposed a core structure of β proteins that consists of four SSEs and described the variation in β proteins by stepwise addition of other SSEs to the core structure. Zhang and Kim [12] combined the β-sandwich and β-barrel architectures by regarding these two as variants of the Greek key motif. Ruczinski et al. [13] surveyed the distribution of proteins with an open β-sheet structure and examined how well proteins obey the rules known before.

In this study, we re-examined the folds classified as 'all-β proteins' in the SCOP database [2] from a viewpoint totally different from the SSEs' spatial arrangement and found that two structural features repeatedly appear, namely the ring-shaped structure of the main chain trace and the zipper-like contact pattern in the ring structure. First, we give a definition of the ring structure and the zippered contact pattern. The limited variety of the protein folds is then explained by means of these two features, and a possible connection to the folding process is discussed.

2. Materials and methods

2.1. Dataset

In the SCOP database (ver. 1.55, July 2001) [2], there are 93 folds classified as all-β proteins. One representative for each fold was selected, which was determined by X-ray crystallography with the best resolution among proteins in the same fold. For three folds with duplication (PNGase F-like, Trypsin-like serine proteases, and Acid proteases), we used one of the duplicates. Two folds (gp9 and Ev matrix protein) are divided into two domains according to the annotations of SCOP. Thus, 95 protein domains were used in this study.

2.2. Definition of ring structure

A ring structure (Fig. 1a) was identified in a protein domain as follows. Consider pairs of Cα atoms with a spatial distance of less than 7.0 Å. We call the chain part, bounded by the Cα pair with the largest sequence separation, the closed region of the domain. When the ratio, R_{ring}, of the number of residues in the closed region to the total residue number exceeds 0.6, the domain is considered to have a ring structure. The distribution of R_{ring} for the 95 β domains is shown in Fig. 2a.

2.3. Definition of zippered contact pattern

The zipper-like contact pattern (Fig. 1b) was identified in the ring structure by the following method. The contacting residue pairs are defined as those satisfying $r_{ij} \leq r_{\mathrm{cut}}$, where r_{ij} is the distance (Å) between the ith and jth Cα atoms, and $r_{\mathrm{cut}} = 12$ Å. The zipper-like contact pattern can be recognized in a path following the contacting pairs starting from the left bottom and ending at the right top on the distance matrix (Fig. 1c). Such a pattern can be detected in a quantitative manner by the dynamic programming (DP) algorithm [14]. We used a score matrix for DP having the form, $(a-r_{ij})^2$ for $r_{ij} \leq r_{\mathrm{cut}}$ and $-b$ for $r_{ij} \geq r_{\mathrm{cut}}$ (Fig. 1g), where $a = 20$ Å, $b = 10$ Å^2, and the periodic boundary condition (Fig. 1f) was introduced to detect the zippered pattern even in the structure depicted in Fig. 1e, which we call here a circular permutated structure. No gap penalty was used. Instead, horizontal and vertical paths are considered to represent either a one-to-many contact for the contacting pairs (e.g. residue pairs 3–8 and 3–11 in Fig. 1c) or an insertion for non-contacting pairs (e.g. 3–9 and 3–10 in Fig. 1c). A zippered contact pattern is then identified when the ratio, R_{zipper}, of the number of residues involved in the contact along the optimal path to the number of residues in the closed region exceeds 0.8. The distribution of R_{zipper} is shown in Fig. 2b.

*Corresponding author. Fax: (81)-45-508 7367.
E-mail addresses: koike@tsurumi.yokohama-cu.ac.jp (R. Koike), kinoshita@tsurumi.yokohama-cu.ac.jp (K. Kinoshita), kidera@tsurumi.yokohama-cu.ac.jp (A. Kidera).

0014-5793/02/$22.00
PII: S0014-5793(02)03729-8

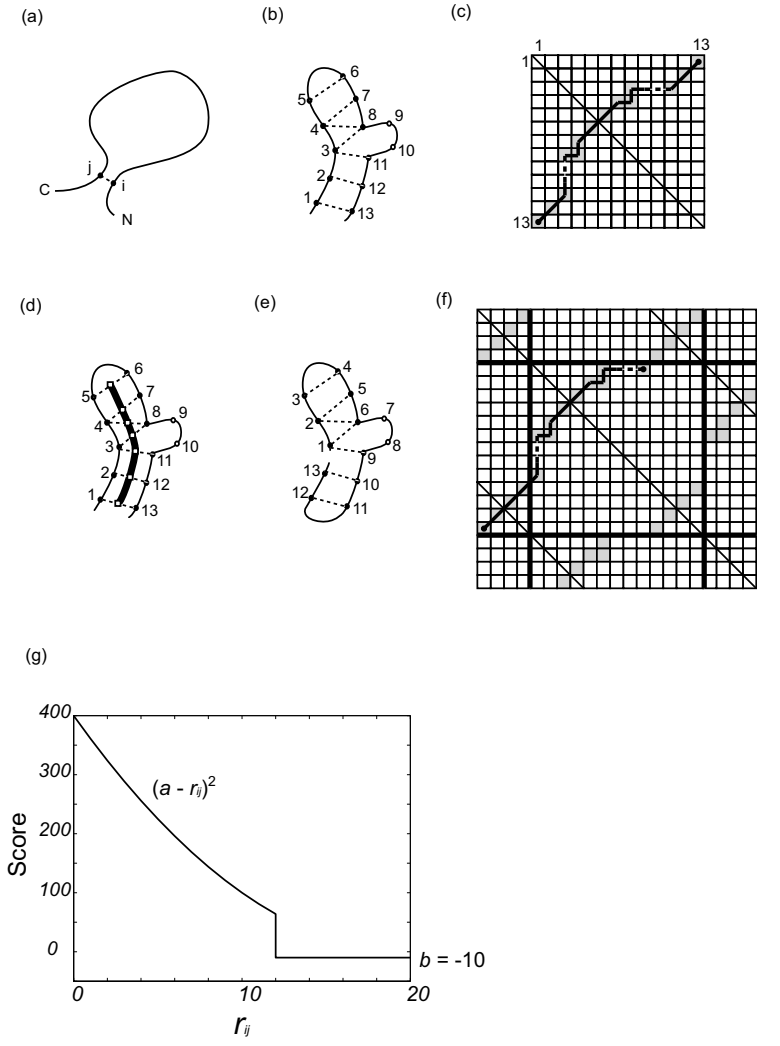

Fig. 1. a: Schematic representation of the closed region. The ratio of the closed region is calculated to be $(j-i+1)/n$, where n is the total number of residues in the domain. b: Schematic representation of the zipper-like contact pattern. Black circles and dotted lines indicate pairs of Cα atoms in the contacting residues. White circles represent the atoms in the insertion. c: The distance matrix of the protein shown in b. Shaded boxes indicate the contacting residues, and the thick line is the optimal path calculated by the DP algorithm. d: The string representation of the zipper-like contact pattern, which was obtained by the optimal path in c. e: A schematic example of a circular permutated zippered structure. Naturally occurring examples are seen in Fig. 4c,e. f: The distance matrix with the periodic boundary to detect the circular permutated zippered structure. g: The score function used in the dynamic programming algorithm to detect the zippered structure.

3. Results and discussion

3.1. Ring structure

According to the definition of the ring structure and the zippered contact pattern, we carried out a classification of the 95 β folds. The results of the classification are summarized in Fig. 3. Detailed descriptions of the classification are available at http://www.tsurumi.yokohama-cu.ac.jp/bioinfo/beta/.

We found 87 (92%) folds with the ring structure. The other eight folds are classified into 'non-ring' structures. This number changes when a different value of the threshold is used. However, Fig. 2a clearly shows the dominance of the ring structure in β proteins. It was also confirmed that the way of selecting a representative from each fold scarcely affects the result. We calculated the values of R_{ring} for the 277 families in the 95 all-β folds (the first entry of each family was taken as the representative). It was found that 264 (95%) families were classified as 'ring', and the distribution (Fig. 2c) is very similar to that of the 95 folds (Fig. 2a).

The 'non-ring' structures are the folds of β-helix (two domains), meander-like (five domains), and helix–meander hybrid (one domain). The structural feature of 'non-ring' is found in the large fraction of the short-range contacts compared with 'ring'. This is seen in the quantitative measure of the relative contact order (RCO, the normalized average of the sequence separation for each residue contact) [15]. The average values of RCO are 0.17 for the 87 domains in 'ring', and 0.09 for the eight domains in 'non-ring'. This means that β proteins with a large fraction of long-range contacts (large RCO) tend to have the ring structure.

3.2. Zippered structure

In the proteins with the ring structure, we found 57 folds (66% of 87 'ring' folds) with the zipper-like contact pattern (57 'zippered' and 30 'non-zippered' in Fig. 3). Incidentally, 165 'zippered' families (63% of 264 'ring' families) were found. Fig. 2b and Fig. 2d show the distribution of R_{zipper} for the 87 folds and the 264 families, respectively.

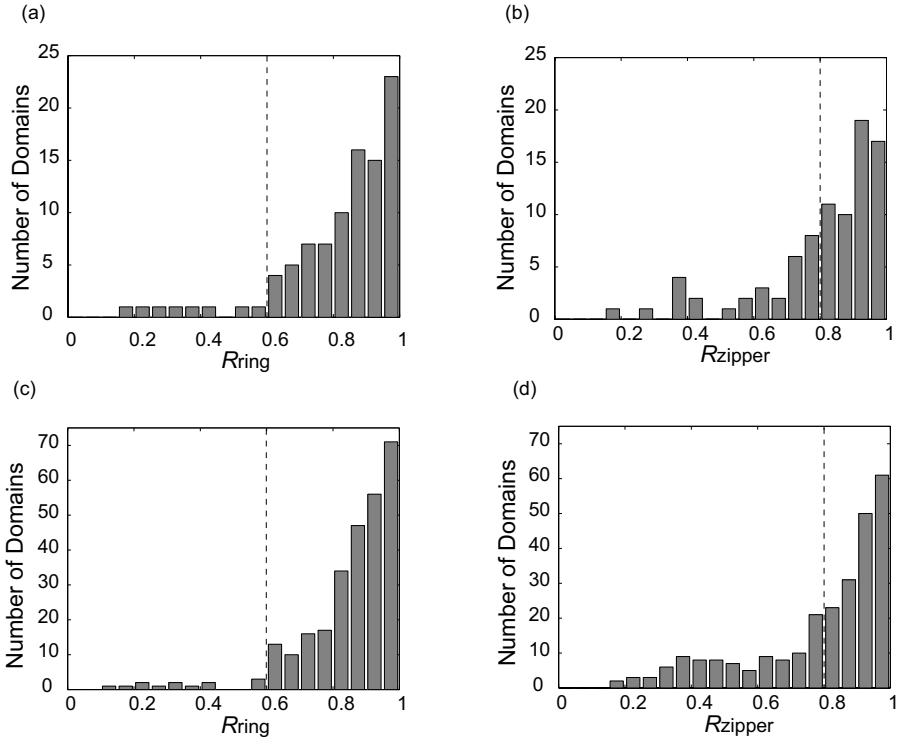

Fig. 2. a: Distribution of R_{ring} for the 95 all-β folds. R_{ring} is the number of residues in the closed regions normalized by the total residue numbers in each domain. b: Distribution of R_{zipper} for the 87 folds with 'ring' structure. R_{zipper} is the number of residues involved in the contact along the optimal path normalized by the number of residues in the closed region. c: Distribution of R_{ring} for the 277 families classified in all-β folds. d: Distribution of R_{zipper} for the 264 families classified as 'ring'.

Typical examples of 'zippered' are shown in Fig. 4a,c,e, and one 'non-zippered' is shown in Fig. 4g. The most typical example of 'ring/zippered' may correspond to a jelly-roll motif (Fig. 4a). The zippered structure is characterized in the long path of the distance map as shown in Fig. 4b,d,f. On the other hand, the non-zippered structure has only fragmented paths (Fig. 4h). The average RCOs for 'zippered' and 'non-zippered' are 0.20 and 0.12, respectively. This means that the zippered structures have a larger fraction of long-range interactions than the non-zippered structures.

3.3. Shape of ring/zippered structure

The zippered structures are further classified by the shape of the zipper, which is depicted by a string made up of the center points of the corresponding Cα atoms belonging to the zipper, followed by the smoothing procedure of Priestle's algorithm [16] (Fig. 1d). The strings were categorized by visual inspection into 'helical' (43 domains, e.g. Fig. 4a,c) and 'non-helical' (14 domains, e.g. Fig. 4e). 'Helical' structures are further classified according to the handedness of the helical region, into 28 'right-handed' (e.g. Fig. 4a) and 15 'left-handed' (e.g. Fig. 4c). The summary of the classification is given in Fig. 3.

The difference in structural features between 'helical' and 'non-helical' is seen in the distance map (Fig. 4b,d,f); 'helical' has a dominant sub-optimal path reflecting the helical symmetry, while 'non-helical' does not. In Fig. 4c, an example of the circular permutated zipper structures is shown. Circular permutation does not show any preference in the classification; 27 circular permutated zipper structures are classified into 12 'right-handed', seven 'left-handed', and eight 'non-helical'. Our classification does not correlate with the well-known classification of β protein such as β-sandwich and β-barrel [6]; 57 'zippered' structures contain 26 β-sandwiches and 22 β-barrels, according to the annotation of SCOP.

3.4. Classification in terms of RCO

In Fig. 5, the distribution of RCO is plotted for the three classes of β proteins, namely 'non-ring', 'ring/non-zippered' and 'ring/zippered'. As mentioned above, their distributions are separated; their average values of RCO are 0.09, 0.12 and 0.20 for 'non-ring', 'ring/non-zippered' and 'ring/zippered', respectively. It is worth mentioning that the upper limit of RCO for all the protein folds in SCOP, belonging to not only all-β, but also to all-α, α+β, and α/β, was found to be 0.33 for 1png_1, which is the 'PNGase F-like' fold classified in 'ring/zippered/helical/right-handed'. It is also noted in Fig. 5 that the number of folds increases with ascending order of RCO, i.e. 'non-ring', 'ring/non-zippered' and

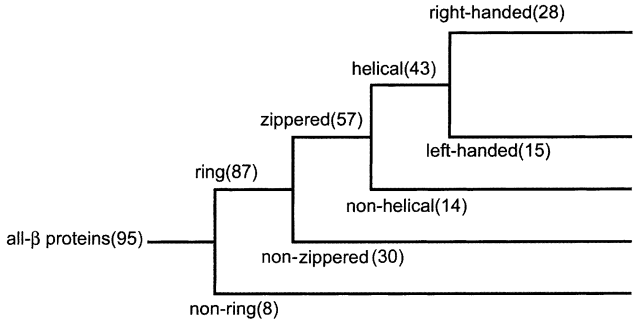

Fig. 3. Summary of the classification. The number in parentheses denotes the number of domains in each category.

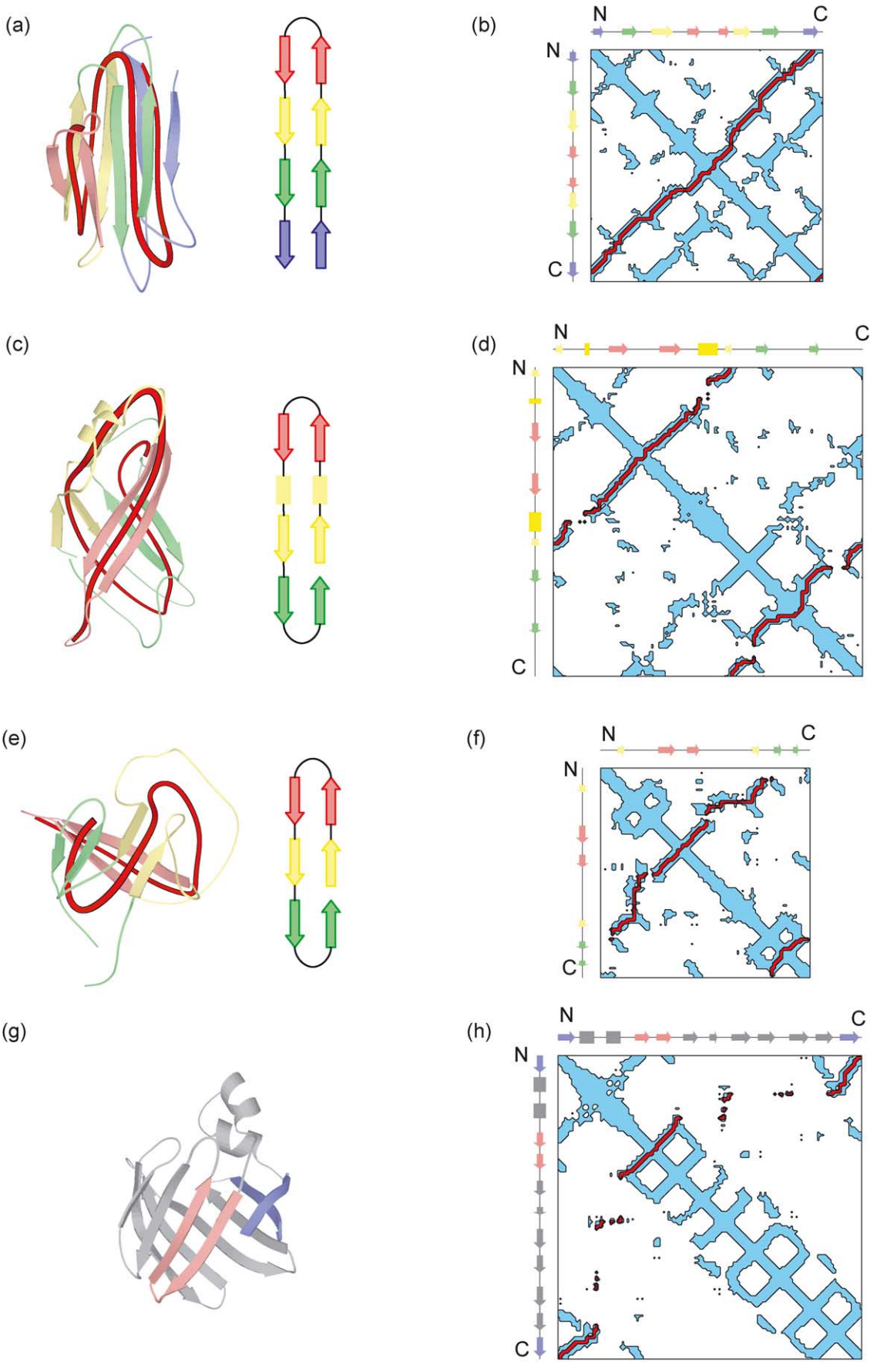
(a)
(b)
N
C
N
C
(c)
(d)
N
C
N
C
(e)
(f)
N
C
N
C
(g)
(h)
N
C
N
C

Fig. 4. Molscript [20] pictures of typical examples of each category, and their corresponding distance matrices with the contacting residues in light blue and the optimal path in red indicating the zippered structure. Strings in red depict the shape of the zippered structure (see Fig. 1f). Corresponding segments are colored in the order red, yellow, green and blue. a,b: 'Ring/zippered/helical/right-handed' (spermadhesin CUB domain, 1sfp_ _ (in SCOP domain ID). c,d: 'Ring/zippered/helical/left-handed' (EV matrix protein, 1es6a_ (44–194). e,f: 'Ring/zippered/non-helical' (oncogene products, 1a1x_ _). g,h: 'Ring/non-zippered' (lipocalins, 1ifc_ _).
←

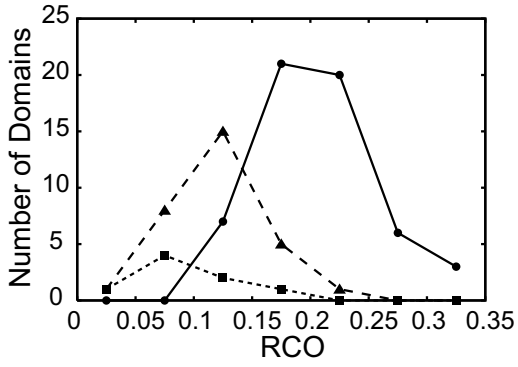

Fig. 5. Distributions of RCOs for 'non-ring' (dotted line), for 'ring/non-zippered' (broken line), and for 'ring/zippered' (solid line), whose average values are 0.09, 0.12, and 0.20, respectively.

'ring/zippered'. In fact, 'non-ring', consisting mostly of short-range contacts, can contain only two patterns of folds, meander and β-helix. On the other hand, in 'ring/zippered', due to the variety in the shape of the zipper (Fig. 4), a large structural variety is realized.

It is seen in Fig. 5 that almost no β fold other than the ring and zippered structures is allowed in all-β proteins with large RCO values (>0.2), or with a large fraction of long-range native contacts. This observation leads us to the following conjecture. There is a certain mechanism inherent in the ring and zippered structures to form long-range native contacts easily in the process of folding through a specific pathway. Other types of β fold with a similar RCO, if any, cannot fold into the native structure in a reasonable time scale.

Folding pathways leading to the ring and zippered structures can be considered in a number of ways. Ring closure in an early stage of folding would reduce the chain entropy to accelerate the subsequent process of folding. The folding path of zipping a zipper-like structure would reduce the conformational space to be searched and speed up folding [17]. The folding from a long hairpin structure may be one of the plausible pathways [18]. An analogy of the supercoiling of double-stranded DNA [19] is another attractive scenario. A combination of ring closure and the local twisting power generates a variety of higher-order three-dimensional structures of the DNA chain. Likewise, in β proteins, the ring structure and the twisting tendency of each β-strand would be enough to form a helical zippered structure with no help from specific long-range interactions for pairing β-sheets.

Acknowledgements: This work was supported by a grant from the MEXT to K.K. and A.K. R.K. acknowledges the support from JSPS.

References

[1] Pearl, F.M., Lee, D., Bray, J.E., Buchan, D.W., Shepherd, A.J. and Orengo, C.A. (2002) Protein Sci. 11, 233–244.
[2] Lo Conte, L., Ailey, B., Hubbard, T.J., Brenner, S.E., Murzin, A.G. and Chothia, C. (2000) Nucleic Acids Res. 28, 257–259.
[3] Orengo, C.A., Jones, D.T. and Thornton, J.M. (1994) Nature 372, 631–634.
[4] Brenner, S.E., Chothia, C. and Hubbard, T.J. (1998) Proc. Natl. Acad. Sci. USA 95, 6073–6078.
[5] Richardson, J.S. (1981) Adv. Protein Chem. 34, 167–339.
[6] Chothia, C., Hubbard, T., Brenner, S., Barns, H. and Murzin, A. (1997) Annu. Rev. Biophys. Biomol. Struct. 26, 597–627.
[7] Chothia, C. and Finkelstein, A.V. (1990) Annu. Rev. Biochem. 59, 1007–1039.
[8] Finkelstein, A.V. and Ptitsyn, B.O. (1987) Prog. Biophys. Mol. Biol. 50, 171–190.
[9] Sternberg, M.J.E. and Thornton, J.M. (1977) J. Mol. Biol. 110, 269–283.
[10] Richardson, J.S. (1976) Proc. Natl. Acad. Sci. USA 73, 2619–2623.
[11] Efimov, A.V. (1997) Proteins 28, 241–260.
[12] Zhang, C. and Kim, S.H. (2000) Proteins 40, 409–419.
[13] Ruczinski, I., Kooperberg, C., Bonneau, R. and Baker, D. (2002) Proteins 48, 85–97.
[14] Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. 48, 443–453.
[15] Plaxco, K.W., Simons, K.T. and Baker, D. (1998) J. Mol. Biol. 277, 985–994.
[16] Priestle, P.J. (1988) J. Appl. Crystallogr. 21, 572–576.
[17] Dill, A.K., Fiebig, M. and Chan, H.S. (1993) Proc. Natl. Acad. Sci. USA 90, 1942–1946.
[18] Ptitsyn, O.B. and Finkelstein, A.V. (1980) Q. Rev. Biophys. 13, 339–386.
[19] Olson, W.K. and Zhang, P.S. (1991) Methods Enzymol. 203, 403–432.
[20] Kraulis, P.J. (1991) J. Appl. Crystallogr. 24, 946–950.