ISSN 0006-2979, Biochemistry (Moscow), 2011, Vol. 76, No. 13, pp. 1385-1390. © Pleiades Publishing, Ltd., 2011.
Published in Russian in Uspekhi Biologicheskoi Khimii, 2011, Vol. 51, pp. 25-36.

REVIEW

Structures Closed into Cycles in Globular Proteins

A. V. Efimov

Institute of Protein Research, Russian Academy of Sciences, ul. Institutskaya 4, 142290 Pushchino, Moscow Region, Russia; fax: (495) 514-0218; E-mail: efimov@protres.ru

Received January 26, 2011
Revision received February 15, 2011

Abstract—Different types of structures closed into cycles are widespread at all the levels of structural organization of proteins. β-Hairpins, triple-stranded β-sheets, and βαβ-units represent simple structural motifs closed into cycles by systems of hydrogen bonds. Secondary closing of these simple motifs into larger cycles by means of different superhelices, split β-hairpins, or SS-bridges results in formation of complex structural motifs such as abcd-units, φ-motifs, five- and seven-segment α/β-motifs, etc. At the level of tertiary structure many proteins and domains fold into structures closed into cylinders. Apparently, closing the motifs and domains into cycles and cylinders results in formation of more cooperative and stable structures as compared with open ones, and this may be the reason for high frequencies of occurrence of the motifs in proteins.

DOI: 10.1134/S0006297911130025

Key words: βαβ-unit, β-sheet, handedness, split β-hairpin, structural motif, superhelix

To date, the basic structural principles of protein molecules are known and widely used in studying the structure and function of proteins. Nevertheless, the problem of protein folding remains one of the central problems in biochemistry and molecular biology to be solved. This means that a more detailed analysis of the known principles of protein structure as well as a search for unknown principles and novel structural determinants of proteins should be made. In connection with these needs, the studying of polypeptide structures closed into cycles and their role in protein folding seems to be of particular value.

Different types of structures closed into cycles are widespread at all the levels of structural organization of proteins. α-Helical turns and β-turns are the simplest structures formed by 4-6 amino acid residues closed into cycles by hydrogen bonds. A β-hairpin represents a larger structure (on average, 10-20 residues) closed into cycles many times by a network of hydrogen bonds. At the level of super-secondary structures secondary closing of β-hairpins, triple-strand β-sheets, and βαβ-units results in formation of more complex structural motifs with unique overall folds and the definite handedness, such as abcd- and abCd-units, five- and seven-segment α/β-motifs, φ- and ψ-motifs, etc. At the level of tertiary structure many proteins and domains fold into structures closed into cylinders. In this paper, both the abovementioned and other cyclic structures are described, and their role in protein folding is discussed.

SIMPLE CYCLIC STRUCTURES IN PROTEINS

In globular proteins, the polypeptide chain folds over upon itself many times so that at the sites of different turns, and in the loop regions, it forms irregular but well-defined structures [1]. Small standard structures providing for chain reversal by 180° are called β-turns. They consist of 4-6 amino acid residues closed into cycles by hydrogen bonds (Fig. 1) [1, 2]. In α-helices, there is a saturated network of hydrogen bonds in which the CO-group of residue *i* forms a hydrogen bond with the NH-group of residue *i* + 4 closing into a 13-atom cycle. In 3_{10} helix, the CO-group of residue *i* forms a hydrogen bond with the NH-group of residue *i* + 3 closing into a 10-atom cycle. In β-hairpins, there is a saturated network of hydrogen bonds between the β-strands where every hydrogen bond forms the corresponding cycle (Fig. 2). Triple-strand S- and Z-like β-sheets can be represented as combinations of two β-hairpins and, consequently, there is a set of cycles formed by hydrogen bonds in each of them.

β-Hairpins can be either right- or left-turned (Fig. 2) depending on whether the second β-strand runs on the right or left relative to the first one when viewed from the same side (e.g. as viewed from the hydrophobic core). Similarly, there are two forms of triple-strand β-sheets having simple up-and-down topology, S- and Z-like β-sheets. One of the two forms of β-hairpins or triple-strand

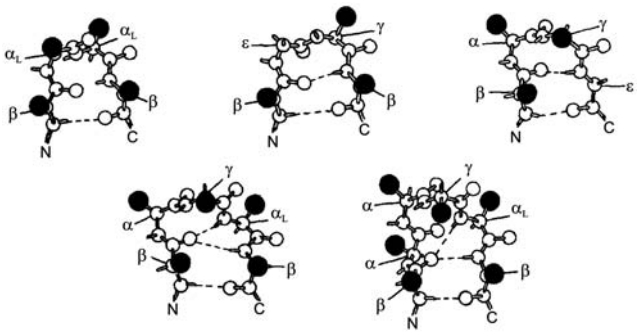

Fig. 1. Structures of β-turns closed into cycles with hydrogen bonds (shown by dashed lines). Structures are shown with ball-and-stick models. Side chains are shown with black balls. Conformations of residues are shown with Greek letters.

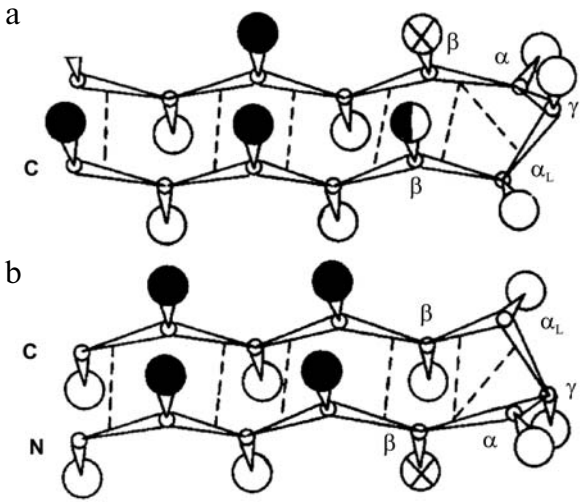

Fig. 2. Schematic representation of right- (a) and left-turned (b) β-hairpins having saturated networks of hydrogen bonds (dashed lines).

β-sheets is selected at the level of super-secondary structures of higher order that include the β-sheets or β-hairpins. Most often, secondary closing of β-hairpins and triple-strand β-sheets into larger cycles by means of different superhelices results in such selection.

SUPER-SECONDARY STRUCTURES CLOSED INTO CYCLES WITH SUPERHELICES

Apparently, the superhelix formed by the βαβ-unit (Fig. 3a) occurs in proteins most often. Rao and Rossmann [3] were the first to observe and describe the βαβ-unit and analogous structures later called Rossmann's folds. Sternberg and Thornton [4] have shown that the βαβ-unit forms the right-handed superhelix in the large majority of observed examples in both homologous and nonhomologous α/β-proteins. In the βαβ-unit, the β-strands form a parallel β-sheet and the α-helix is located in the other layer so that the overall fold of the polypeptide chain represents a large cycle (a right-handed superhelix) stitched by hydrogen bonds between the β-strands. The βαβαβ-unit (Fig. 3d) that is also a commonly occurring folding unit in α/β-proteins has two turns of the right-handed superhelix. There are seven βαβ-units (and consequently seven turns of the right-handed superhelix) closed in a large cycle, the so-called $(\alpha/\beta)_8$-barrel or cylinder (Fig. 3f), in the family of α/β-barrel proteins.

Secondary closing of the βαβ- and βαβαβ-unit into a higher order cycle can be performed in other ways, by means of a split βαβ-unit or a split β-hairpin. In the split βαβ-unit, there are one, two, or more additional β-strands in the β-layer between the β-strands of the βαβ-unit. Secondary closing of the βαβ- and βαβαβ-unit by means of a split βαβ-unit results in formation of the five-segment (Fig. 3b) or seven-segment α/β-motif (Fig. 3e), correspondingly (see also Fig. 4). These structural motifs have unique overall folds and are widespread in the corresponding families of the three-layer α/β-proteins [5, 6]. Another example of a super-secondary structure that can be obtained as a result of secondary closing of the βαβ-unit by means of a split β-hairpin is shown in Fig. 3c [7].

Secondary closing of the β-hairpin by means of a split βαβ-unit results in formation of the abCd-unit (Fig. 4, c and h) that is a commonly occurring folding unit in (α+β)-proteins [5, 8, 9]. There are two forms of the abCd-unit, one has a direct chain orientation (Fig. 4c) and the other a reverse one (Fig. 4h). Their topologies can be described as ββαβ and βαββ, correspondingly. The

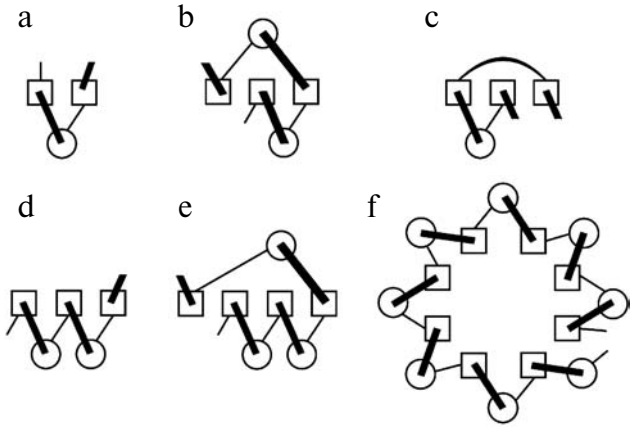

Fig. 3. Structural motifs in α/β-proteins. The structures are viewed end-on with α-helices shown as circles and β-strands as rectangles. The near connections are shown by double lines and the far connections by single lines. a) βαβ-unit; b) five-segment α/β-motif; c) combination of βαβ-unit and ψ-motif; d) βαβαβ-unit; e) seven-segment α/β-motif; f) α/β-barrel. See also the text.

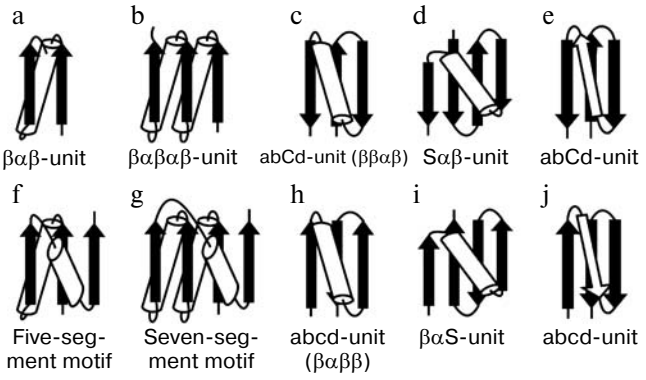

Fig. 4. Structural motifs closed into cycles with the right-handed βαβ- or bcd-superhelices. α-Helices are shown as cylinders and β-strands as arrows directed from the N- to C-ends.

first form of the abCd-unit includes the right-turned β-hairpin when viewed from the hydrophobic core and the other one the left-turned β-hairpin.

The Sαβ- and βαS-units shown in Fig. 4 (d and i) are obtained as a result of secondary closing of the S-like β-sheet by means of a split βαβ-unit. They are similar to the abCd-unit having one additional β-strand in the β-layer. The Sαβ-unit has the S-like β-sheet at the N-end and the βαS-unit at the C-end. It should be noted that the Z-like β-sheet cannot be included in a similar super-secondary structure. Secondary closing of the Z-like β-sheet could be performed by a left-handed βαβ-superhelix, but such structures have not been found in proteins [10].

In most bilayer β-proteins with the aligned β-sheet packing there is a four-stranded super-secondary structure denoted as the abcd-unit [8, 11]. In the abcd-unit, the β-hairpin formed by strands a and b is closed into a large cycle by superhelix bcd (Fig. 4, e and j). Superhelix bcd is analogous to the βαβ-superhelix in the abCd-unit and always occurs in the right-handed form in proteins. The abcd- and abCd-units have the same overall fold of the chain but different conformations of regions c and C. The abcd-unit having the direct chain orientation (Fig. 4e) includes the right-turned β-hairpin, and that with the reverse chain orientation the left-turned β-hairpin (Fig. 4j).

Figure 5 shows that the S-like β-sheet can be included into superhelices of the βSβ- and βSα-types [10]. The overall fold of the chain in these structures can be represented as a right-handed superhelix if the S-like β-sheet is replaced by one imaginary strand. In the so-called β-barrel proteins, the structures are closed into cylinders by means of the βSβ-superhelices (see, e.g. SH3-like and GroES-like folds [12, 13] and H-subunit [14]). In OB-folds [15] and neurophysin [16], the structures are closed by means of the βSα-superhelices. Note that analogous superhelical structures including the Z-like β-sheets do not occur in proteins [10].

There are two ways of including the Z-like β-sheet into cyclic structures. One way is formation of a complex variant of the abcd-unit in which region c is a Z-like β-sheet when viewed from the hydrophobic core [8, 10]. The other way is shown in Fig. 6. At the first step, a 3β-

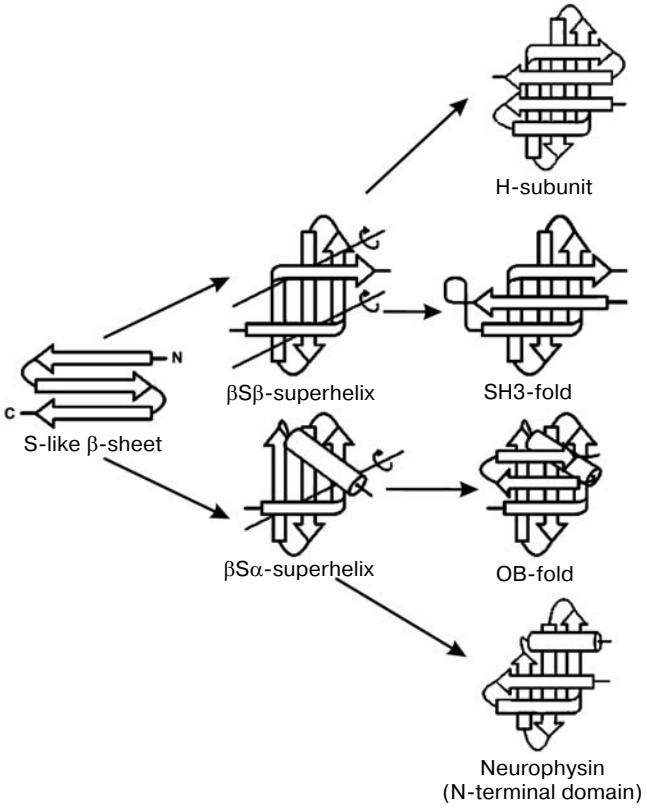

Fig. 5. Structures closed into cycles with the βSβ- or βSα-superhelices. Imaginary polypeptide chain rotation axes on transition from one layer to another are shown by straight lines with circular arrows. On the right, examples of real folds in proteins containing such superhelices.

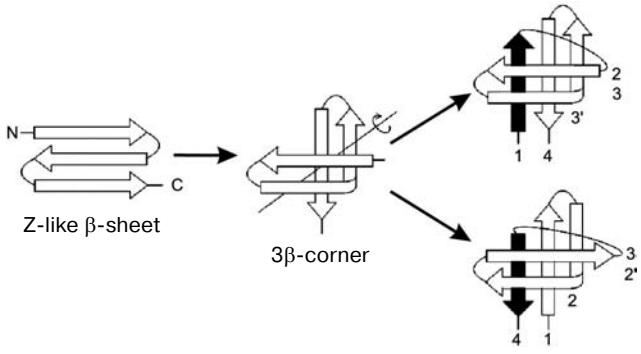

Fig. 6. A scheme of including the Z-like β-sheet into closed structures. In the first step, a 3β-corner is formed (rotation axis is shown as on Fig. 5), and then additional β-strands (black arrows), which close β-corners into cycles, are attached to it. Figures designate the order of β-strands beginning from N-end.

corner is formed. The 3β-corner is a structural motif that can be represented as a Z-like β-sheet folded upon itself so that its two β-hairpins are packed approximately orthogonally in different layers, and the central β-strand is bent by 90° when passing from one layer to the other to form a half-turn of the right-handed superhelix [17]. An addition of a β-strand to the 3β-corner at the N- or C-end results in formation of a right-handed superhelix that closes the structure into a cycle similar to superhelix bcd closing the abcd-unit. In the upper structure (Fig. 6), the superhelix can be observed if an imaginary crossover strand would connect the C-end of strand 1 and the N-end of strand 3′, and in the bottom structure if an imaginary crossover strand would connect the C-end of strand 2 and the N-end of strand 4 (for details, see [18]).

SUPER-SECONDARY STRUCTURES CLOSED INTO CYCLES WITH SPLIT β-HAIRPINS

Secondary closing of the β-hairpin into a cycle by means of a split β-hairpin results in formation of a structural motif denoted as the φ-motif [19]. The simplest φ-motif is formed by three adjacent β-strands connected by loops and packed in one β-sheet so that its overall fold resembles the Greek letter φ (Fig. 7). There are two types of φ-motifs – the hairpin-strand type where a β-strand follows a β-hairpin (Fig. 7, a and c) and the strand-hairpin type in which a β-hairpin follows a β-strand (Fig. 7, b and d). The loop, which connects two edge β-strands and crosses over the central β-strand or its extension, is referred to here as the crossover loop. There are right-handed and left-handed φ-motifs. When viewed from the crossover loop, the polypeptide chain runs from the N- to the C-end in the clockwise direction in the right-handed φ-motifs (Fig. 7, a, b and e), and in the anticlockwise direction in the left-handed φ-motifs (Fig. 7, c and d). In the right-handed φ-motifs, the β-hairpins formed by strands 1 and 2 (Fig. 7, a and e) and strands 2 and 3 (Fig. 7b) as well as the split β-hairpins formed by strands 2 and 3 (Fig. 7, a and e) and strands 1 and 2 (Fig. 7b) are right-turned, but in the left-handed φ-motifs the corresponding β-hairpins are left-turned (Fig. 7, c and d). Analysis shows that in proteins φ-motif occur predominantly in the right-handed form [19].

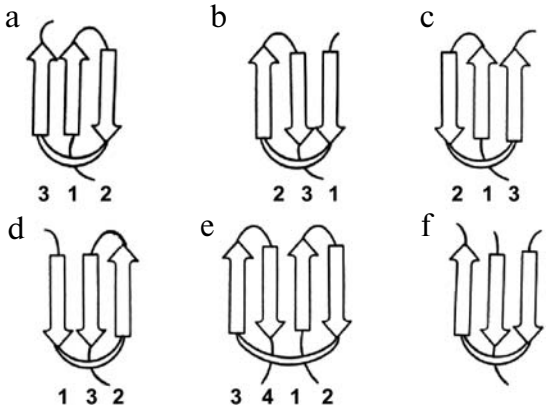

Fig. 7. Schematic representation of different variants of the φ-motif (a-e) and the ψ-motif (f). 1-4 are numbers of β-strands in the chain starting from the N-end.

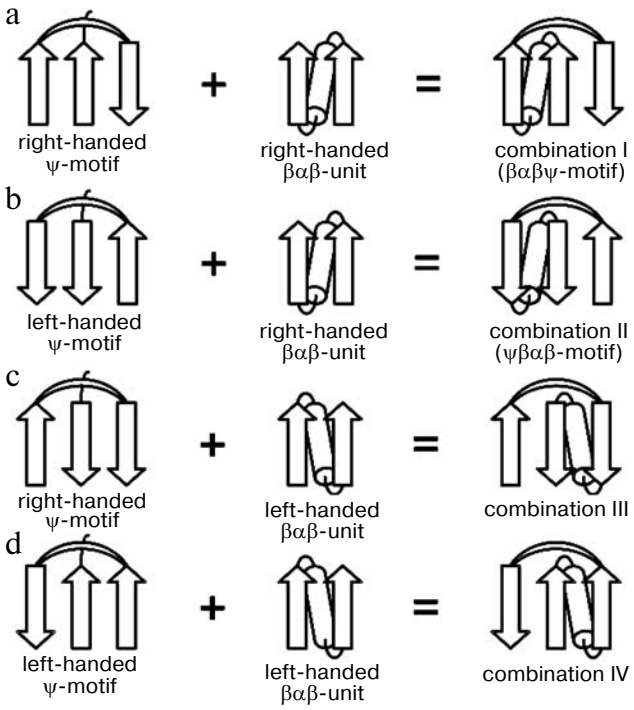

Fig. 8. Schematic representation of possible combinations of the βαβ-unit and ψ-motif.

Figure 7f shows a schematic representation of the ψ-motif [20, 21] that has both similarity to the φ-motif and essential differences. The main difference between the φ- and ψ-motifs is that the central β-strand of the ψ-motif is not connected by loops with the other β-strands and can be formed by a rather distant region of the polypeptide chain. In β-proteins, the ψ-motif occurs predominantly in the right-handed form [20, 21], in which the split β-hairpin is right-turned when viewed from the crossover loop (Fig. 7f). Note that in α/β-proteins a marked part of the observed ψ-motifs are left-handed [7].

As mentioned above (see Fig. 3c), secondary closing of the βαβ-unit into a larger cycle can be performed by a split β-hairpin. This results in formation of a super-secondary structure that can be represented as a combination of the βαβ-unit and the ψ-motif. Figure 8 shows four possible combinations of the βαβ-unit and the ψ-motif. Analysis shows that the combination of the left-handed ψ-motif and the left-handed βαβ-unit does not

occur in known proteins. Combinations of the right-handed ψ-motif and the left-handed βαβ-unit occur in 11% of the observed cases, combinations of the left-handed ψ-motif and the right-handed βαβ-unit occur in 34% of the cases, and combinations of the right-handed ψ-motif and the right-handed βαβ-unit occur in 55% of the cases [7]. Such frequencies of occurrence of the combinations can be explained by the fact that the left-handed βαβ-unit are unfavorable structures from the point of view of stereochemistry and occur very rarely in proteins (less than 1%, see, e.g. [3, 4]). Note that there are 11% of the left-handed βαβ-units in the combinations, and this is rather much as compared with that in total α/β-proteins. The frequency of occurrence of the left-handed ψ-motifs (34%) in the combinations is also much higher than in other protein classes. The reasons for this are still poorly understood and should be investigated further.

STRUCTURE OF β-HAIRPINS CLOSED INTO CYCLES WITH SS-BRIDGES

In the preceding sections different structures closed into cycles by hydrogen bonds have been considered. In this section, structures closed into cycles by covalent bonds, the so-called SS-bridges, are described and analyzed. The fact that closing of the polypeptide chain regions into cycles by SS-bridges results in formation of cyclic structures with definite handedness is of particular interest. It was shown recently [22] that among 118 nonhomologous β-hairpins in which SS-bridges are formed by cysteines located opposite to each other in the neighboring β-strands, 110 β-hairpins are left-turned and only eight are right-turned when viewed from the side where SS-bridges are situated (Fig. 9c).

In the other type of β-hairpins studied in this work the SS-bridge is formed by two cysteines, one of which is located in the β-strand and the other in the loop juxtaposed to the β-hairpin at the N- or C-terminus (Fig. 9, d and e). In total, 228 structures of the loop-hairpin type have been found in nonhomologous proteins and in 211 of them the polypeptide chain forms a left-handed superhelix and a left-turned β-hairpin (Fig. 9d). In most structures of the hairpin-loop type, a similar left-handed superhelix is also formed (Fig. 9e). The high frequency of occurrence of the left-handed superhelices is determined to a great extent by the arrangement of secondary structural elements in higher-order structures in which the superhelices are included. Most often these are the abcd- and abCd-units and the 3β-corners.

Thus, structures closed into different cycles are widespread at all the levels of structural organization of proteins. Apparently cyclic structures are more cooperative and stable as compared with open ones, and this may

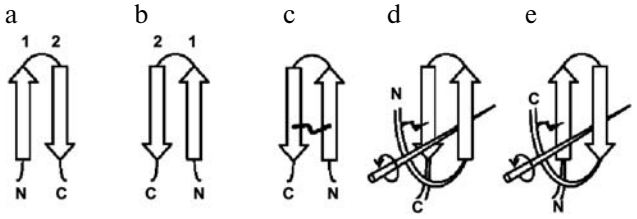

Fig. 9. A schematic representation of a right-turned (a) and left-turned (b) β-hairpin, a left-turned β-hairpin closed into a cycle with an interstrand SS-bridge (c), and left-handed superhelices formed by β-hairpins and loops juxtaposed to them at the N- (d) or C-terminus (e). Imaginary axes of the superhelices are shown as straight bars.

be the reason for their high frequencies of occurrence in proteins. Secondary closing of β-hairpins, triple-strand β-sheets, and βαβ-units into larger cycles by means of different superhelices, split β-hairpins, and SS-bridges is of particular value because this results in formation of more complex structural motifs having unique overall folds and unique handedness (see also [23]). Note that each structure of higher order includes only one of the two forms of β-hairpins (the right- or left-turned one) or triple-strand β-sheets (the S- or Z-like β-sheet). It can be suggested that the structural motifs having unique overall folds can act as nuclei or "ready-made" building blocks in protein folding or can be taken as the starting structures in protein modeling [5, 6]. Recent theoretical modeling of protein folding based on construction and analysis of structural trees showed that the folding pathways resulting in the closed structures are predominantly used [18].

REFERENCES

1. Efimov, A. V. (1986) *Mol. Biol.*, **20**, 250-260.
2. Venkatachalam, C. M. (1968) *Biopolymers*, **6**, 1425-1436.
3. Rao, S. T., and Rossmann, M. G. (1973) *J. Mol. Biol.*, **76**, 241-256.
4. Sternberg, M. J. E., and Thornton, J. M. (1976) *J. Mol. Biol.*, **105**, 367-382.
5. Efimov, A. V. (1994) *Structure*, **2**, 999-1002.
6. Efimov, A. V. (1997) *Proteins*, **28**, 241-260.
7. Kargatov, A. M., and Efimov, A. V. (2010) *Biochemistry* (Moscow), **75**, 249-256.
8. Efimov, A. V. (1982) *Mol. Biol.*, **16**, 799-806.
9. Gordeev, A. B., and Efimov, A. V. (2009) *Mol. Biol.*, **43**, 521-526.
10. Efimov, A. V. (1993) *FEBS Lett.*, **334**, 253-256.
11. Gordeev, A. B., Kondratova, M. S., and Efimov, A. V. (2008) *Mol. Biol.*, **42**, 323-326.
12. Masacchio, A., Noble, M, Pauptit, R., Wierenga, R., and Saraste, M. (1992) *Nature*, **359**, 851-855.
13. Hunt, J. F., Weaver, A. J., Landry, S. J., Gierash, L., and Deisenhofer, J. (1996) *Nature*, **379**, 37-45.

14. Deisenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H. (1985) *Nature*, **318**, 618-624.
15. Murzin, A. G. (1993) *EMBO J.*, **12**, 861-867.
16. Chen, L., Rose, J. P., Breslow, E., Yang, D., Chang, W.-R., Furey, W. F., Jr., Sax, M., and Wang, B.-C. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 4240-4244.
17. Efimov, A. V. (1992) *FEBS Lett.*, **298**, 261-265.
18. Boshkova, E. A., and Efimov, A. V. (2010) *Biochemistry* (Moscow), **75**, 1258-1263.
19. Efimov, A. V. (2008) *Biochemistry* (Moscow), **73**, 23-28.
20. Suguna, K., Bott, R. R., Padlan, E. A., Subramanian, E., Sheriff, S., Cohen, G. H., and Davies, D. R. (1987) *J. Mol. Biol.*, **196**, 877-900.
21. Castillo, R. M., Mizuguchi, K., Dhanaraj, V., Albert, A., Blundell, T. L., and Murzin, A. G. (1999) *Structure*, **7**, 227-236.
22. Brazhnikov, E. V., and Efimov, A. V. (2010) *Mol. Biol.*, **44**, 529-534.
23. Efimov, A. V. (2010) *BBRC*, **399**, 412-415.