## STRUCTURAL AND FUNCTIONAL ANALYSIS
## OF BIOPOLYMERS AND THEIR COMPLEXES

# Relationship between Structure and Amino Acid Sequence of Strongly Twisted and Coiled β-Hairpins in Globular Proteins

**E. A. Boshkova, E. V. Brazhnikov, and A. V. Efimov***

*Institute of Protein Research, Russian Academy of Sciences, Pushchino, Moscow oblast, 142290 Russia*
*\*e-mail: efimov@protres.ru*
Received October 2, 2015; in final form, November 14, 2015

**Abstract**—β-Hairpins are widespread in proteins, and it is possible to find them both within β-sheets and separately. In this work, a comparative analysis of amino acid sequences of β-strands within strongly twisted β-hairpins from different structural protein subclasses has been conducted. Strongly twisted and coiled β-hairpin generates in the space a right double helix out of β-strands that are connected by a loop region (connections). The frequencies of amino acid residues on the internal (concave) and external (convex) surfaces of strongly twisted β-hairpins have been determined (220 β-hairpins from nonhomologous proteins were studied). The concave surface of these β-hairpins is mainly generated by hydrophobic residues, while the convex surface by hydrophilic residues; accordingly, the alternation of hydrophobic internal and hydrophilic external residues is observed in their amino acid sequences. Amino acid residues of glycine and alanine (especially in places of the largest twisting of the strands) were anomalously frequently found in internal positions of strongly twisted and coiled β-hairpins. It was established that internal positions never contain the proline residues, while external positions in the twisting region contain them in a relatively large amount. It was demonstrated that at least one amino acid residue in $\alpha_L$- or ε-conformation is required for generation of relatively short (up to 7 amino acid residues) connection. As a rule, these positions are occupied by glycines. Thus, not only the alternation of hydrophobic and hydrophilic amino acid residues, but also the presence of one or two glycine residues in the connection region and the excess of glycines and alanines in the places of the largest strand twisting on the concave surface, as well as the presence of prolines on the convex surface, are required to generate a strongly twisted and coiled β-hairpin.

*Keywords*: twisted β-hairpin, β-turn, conformation, stereochemical analysis, alignment, amino acid sequence

**DOI:** 10.1134/S0026893316050046

## INTRODUCTION

β-Hairpin is one of the simplest and most frequently found structural motifs in proteins. It is generated by two adjacent by the chain β-strands that are folded on each other and connected by a loop region (connection) that results in the generation of antiparallel β-structure enclosed by hydrogen bonds. β-Hairpins are found in structural globular proteins both in isolated form and within other structural motifs, e.g., abcd- and abCd-units, S- and Z-shaped β-sheets, ββ-hairpins and 3β-corners, etc. β-Hairpins can be right or left if looking on from the one selected side (for example, from the side of hydrophobic nucleus). In the right hairpins, the second strand by the chain is located on the right relative to the first one; whereas in the left hairpins, the opposite is observed. Most β-hairpins and β-sheets have twisted, not flat structures, which resembles the right propeller (if one is looking along the β-strands) [1]. The twisting degree of β-sheets in proteins is different, but at the average dihedral angle between neighboring β-strands is close

to −20° [1, 2]. In strongly twisted β-sheets, β-strands should be twisted as well as coiled (in order to generate a large contact surface without damaging the hydrogen bond system) [2, 3]. As demonstrated by the analysis, many β-hairpins in proteins have this strongly twisted and coiled structure, which can be presented as a kind of right double helix with a concave and convex surface. A distinctive feature of such right double helixes is that they are always generated by right β-hairpins (if one looking from the side of the concave surface) [4]; i.e., the second strand by the chain will always be located to the right relative to the first strand (Fig. 1a). Left β-hairpins do not generate such right double helices and they are not found in proteins in strongly twisted form [4]. However, it should be noted that both right and left β-hairpins, which can be flat or weakly twisted, are widespread in proteins. The structure of one of the proteins with strongly twisted and coiled β-hairpin obtained by means of RasMol program [5] is presented on Fig. 1b.
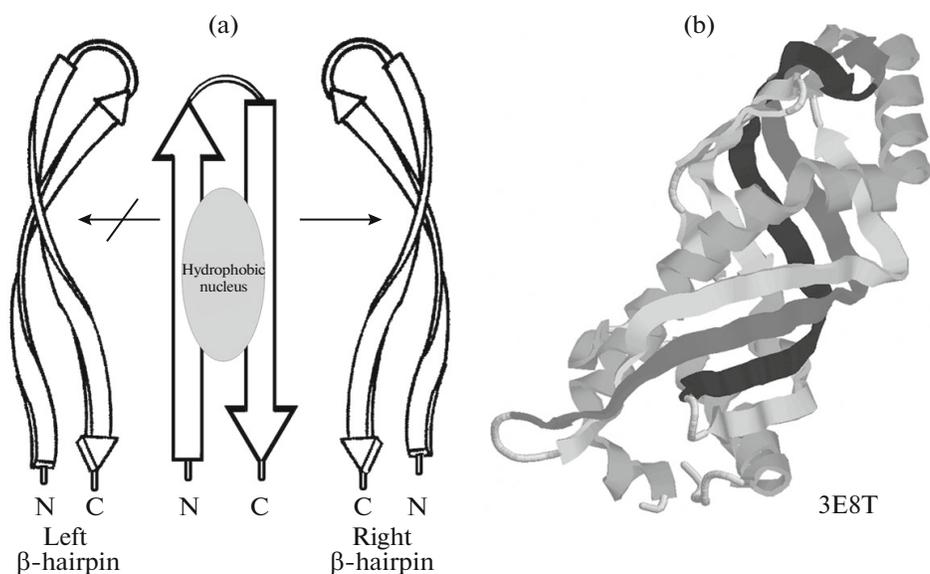
**Fig. 1.** (a) Schematic image of structure of strongly twisted and coiled β-hairpins (explanations in the text). (b) Band model of the protein containing right strongly coiled β-hairpin (PDB-code 3e8t, transport *Takeout-like protein 1* from Epiphyas postvittana). Strongly twisted β-hairpin is highlighted in dark color.

Long β-hairpins can be folded on themselves, generating ββ-corner, and these hairpins are also right if looking from the concave surface [4]. Strongly twisted and coiled β-hairpins are most frequently found in β-barrels, toxins, inhibitors, SH3-domains, and Wrap- proteins.

In the present study, we conducted a detailed analysis of the primary structure of strongly twisted and coiled β-hairpins. It was demonstrated that glycines, alanines, valines, isoleucines, leucines, phenylalanines, and tyrosines are mainly located on hydrophobic internal hairpin surface; in addition to hydrophilic and charged amino acid residues, all proline residues (available in the sequence) are located on the external surface. The types of connections most typical of strongly twisted β-hairpins and amino acid residues required for their development were determined. The data obtained will be useful both for predicting the spatial structure of the protein based on the primary structure and for constructing artificial proteins.

## EXPERIMENTAL

**The proteins for the study** containing twisted and coiled β-hairpins were selected from the Protein Data Bank (PDB) by means of the structural protein classification (PCBOST) developed in our group (available at http://strees.protres.ru/) [6], from the SCOP database [7], or directly from the PDB server using key words. Nonhomologous proteins (protein domains) were selected for the analysis; checking for the homology was conducted using Blast 2 sequences program for pairwise alignment (http://blast.ncbi.nlm.nih.gov/Blast.cgi) [8]. According to recommendations of the

program developer, those proteins were considered nonhomologous, for which the index, which takes into account both the complete identity of the sequence regions and substitutions on similar amino acids, was less than 50 points during the alignment of amino acid sequence of the studied domain relative to each of the sequences of other analyzed domains. β-Hairpins were visually determined by means of the RasMol program. Only β-hairpins with structures close to canonical (strongly twisted and coiled determined with a high resolution that contain no continuous fractures or loopings and have with β-strands at least 5 amino acid residues) were selected for the study. The length and conformation of the connection between β-strands was not considered during the selection.

Strongly twisted β-hairpins from the structural subclass of α + β-proteins (which were called "Wrap-proteins") that we isolated was the main object of the study [9]. The proteins and domains related to this subclass consist of strongly twisted and coiled β-sheet, on the internal (concave) surface of which one or two α-helixes is packed. These proteins are only united by the structural similarity but not homology and include inhibitors of different enzymes, antibiotics, transcription regulators, etc. The band model of the structure of one member from the Wrap-proteins with the PDB code 3E8T is demonstrated in Fig. 1b.

A total of 80 strongly twisted β-hairpins from Wrap-proteins, as well as 36 β-hairpins from the proteins with β-barrel structure, 33 from toxins, 21 from inhibitors, 23 from SH3- and SH3-like domains, and another 27 twisted and coiled β-hairpins from the proteins of other structural subclasses were selected from the database. In total, 220 strongly twisted and coiled
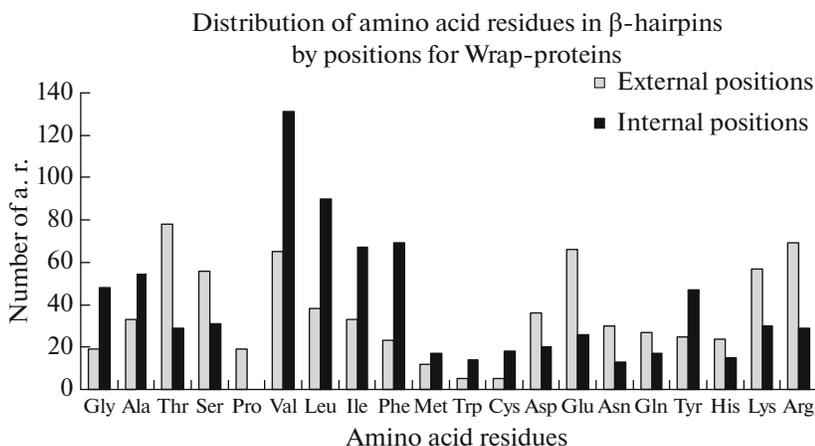
**Fig. 2.** Frequencies of different amino acid residues in internal and external positions of strongly twisted and coiled β-hairpins within Wrap-proteins; a.r., amino acid residue.

β-hairpins from nonhomologous proteins were selected and studied. β-Hairpins both within β-sheets and single were in the sample.

The analysis of the amino acid composition of β-strands on the external (convex) and internal (concave) surfaces of strongly twisted β-hairpins was conducted separately for each of the protein subclasses. The classification of each β-strand position on an external or internal surface was visually determined on the skeletal model of the studied protein molecule in the RasMol program. Subsequently, the number of each of the 20 amino acid residues in all β-strand of twisted β-hairpins in external and internal positions was calculated. Data are presented as histograms discussed below.

In the case of Wrap-proteins, the alignment and analysis of the amino acid sequence and conformation of connections between β-strands were additionally conducted. The alignment was carried out manually taking into account data on the conformation of each of the amino acid residues and the presence of hydrogen bonds. The MolMol was used to calculate angles φ and ψ [10]. The conformation of amino acids was designated according to the nomenclature suggested in 1986 [11]. The connections or connection regions with the same conformation were aligned. Their amino acid composition in each position was subsequently analyzed.

## RESULTS

The histogram of amino acid distribution on the internal (concave) and external (convex) surfaces of β-hairpins from Wrap-proteins is presented in Fig. 2. It can be seen that mainly glycines, alanines, valines, leucines, isoleucines, and phenylalanines are located on the internal surface. The differences in the ratio of hydrophobic and hydrophilic amino acid residues are expected (internal positions are apriory directed

towards the hydrophobic nucleus of strongly twisted β-hairpin). However, the sharp prevalence of glycine residues in internal positions attracts attention. During the alignment of amino acid sequences of twisted β-hairpins, it can be seen that the glycine residues are concentrated approximately in the middle of β-strands, i.e., in the region of maximal bend (see below). As demonstrated by a theoretical stereochemical analysis of strongly twisted and coiled β-hairpins [4], the twisting should inevitably cause tension, which results in specific amino acid selection. Small β-bends are usually generated due to twisting in the central part of β-strands. Glycine is the most conformationally flexible residue. The filling of internal positions by glycine or alanine residues removes the stereochemical tension.

Hydrophilic and charged amino acid residues prevail on the external side. It is remarkable that no prolines were found on the internal side, while they are present in large amounts on the external side. The presence of proline on the concave surface of β-hairpins would damage the system of hydrogen bonds. In the external positions, proline develops a chain break, which facilitates twisting. It is possible that the twisting of long β-hairpins at a suitable sequence for this amino acid provides the development of denser hydrophobic nucleus.

A similar distribution of amino acid residues on the concave and convex surfaces was found for strongly twisted β-hairpins and in other protein classes (Figs. 3a−3d). No difference in the content of glycine amino acid residues in internal and external positions was detected for proteins from the group of toxins; however, the strict prohibition on the proline content in the internal positions was established (typical for all strongly twisted β-hairpins).

The analysis of histograms demonstrates the similarity between the properties of strongly twisted and

(a)
Distribution of amino acid residues in β-hairpins
by positions (beta-barrels)

(b)
Distribution of amino acid residues in β-hairpins
by positions (SH3- and SH3-like domains)

(c)
Distribution of amino acid residues in β-hairpins
by positions (inhibitors)

(d)
Distribution of amino acid residues
in β-hairpins by positions (toxins)
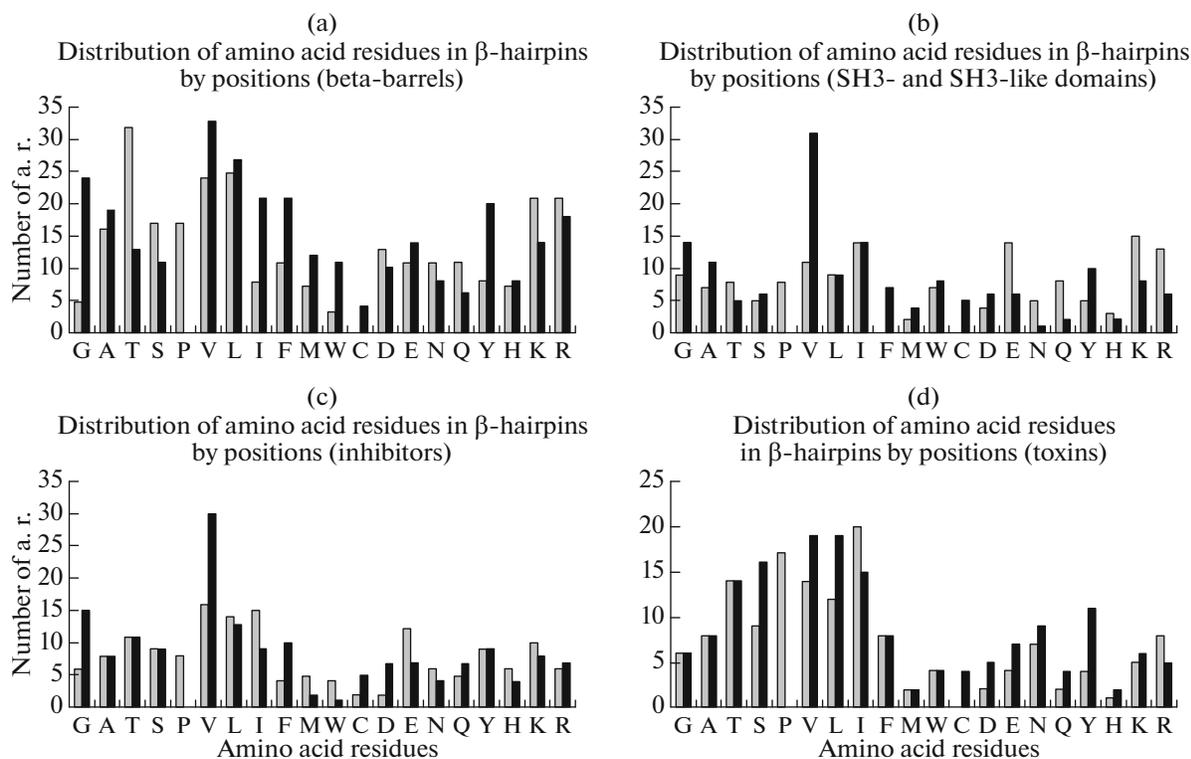
**Fig. 3.** Frequencies of different amino acid residues in internal and external positions of strongly twisted and coiled β-hairpins within (a) beta-barrels, (b) SH3-like domains, (c) inhibitors, and (d) toxins. External positions of β-hairpins are designated on histograms by a grey color; internal positions, by a black color.

coiled β-hairpins from the proteins of difference classes. This allowed one to construct a total diagram of the amino acid distribution by the surfaces of β-hairpins for all studied proteins (Fig. 4). It is obvious that the observed differences are not inherent to a special class of proteins; they are typical of the most strongly twisted and coiled β-hairpin. These properties are weakly dependent on the structural surrounding.

Distribution of amino acid residues
in β-hairpins by positions (all proteins from database)

□ External positions
■ Internal positions
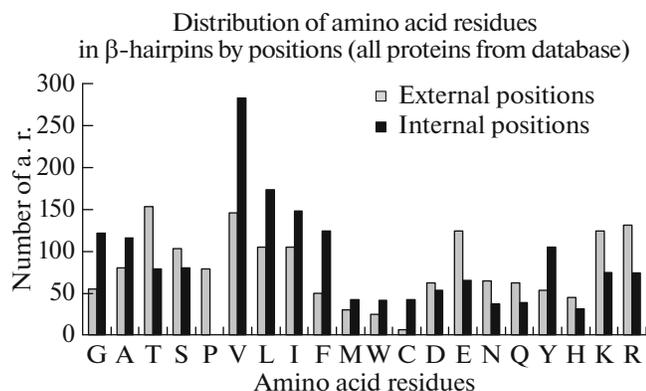
**Fig. 4.** Frequencies of different amino acid residues in internal and external positions of all 220 studied strongly twisted and coiled β-hairpins.

Previously, we described a similar character of amino acid selection for external and internal positions for β-hairpins within the structural 3β-corner motif, where the central of three β-strands is bent at 90° and passes into the orthogonal layer [12]. This bend also results in a strong stereochemical tension; therefore, glycine and alanine amino acid residues in internal positions of central β-strand and the complete absence of the proline residues are required for its development. The selection of amino acid sequence required for generation of 3β-corner is even stricter than for strongly twisted β-hairpins within other structural motifs (the glycine content in internal positions of central β-strand exceeds its content in external positions by more than nine times) [13].

In order to analyze the connection structures between β-strands, amino acid sequence of strongly twisted β-hairpins from Wrap-proteins were manually aligned one below the other so that either internal or external residues fell in each vertical column. Subsequently, β-hairpins were distributed to the groups with similar length and conformation of the connections determined by the value of φ and ψ angles, which are included in their composition of amino acid residues, and aligned by the conformation of the connections. Residues with similar conformations were located in each vertical column (indicated in the upper line). The

example of the alignment of amino acid sequences of twisted β-hairpins is demonstrated in Fig. 5, where the most widespread conformations of short connections are given as follows: $\beta\alpha_L\alpha_L\beta$-, $\beta\gamma\gamma\alpha_L\beta$-, $\beta\epsilon\gamma\beta$-, and $\beta\gamma\gamma\epsilon\beta$, where $\alpha_L$-conformation has the range of torsion angle values $\varphi \approx 70 \pm 30°$, $\psi \approx 10 \pm 30°$; $\gamma$-conformation, $\varphi \approx -90 \pm 30°$, $\psi \approx 0 \pm 30°$; $\epsilon$-conformation, $\varphi \approx 100 \pm 30°$, $\psi \approx -170 \pm 30°$. A detailed description of these and other small standard structures frequently found in irregular protein regions is presented in works [11, 16, 17].

As a result of the analysis, it was found that amino acid residues in $\alpha_L$- and $\epsilon$-conformations (mainly occupied by glycines) are present in both long and short connections. It was empirically established that the presence of at least one amino acid residue in $\alpha_L$- or $\epsilon$-conformations is obligatory for the short connections (with a length up to 7 amino acid residues). This conformation was found most frequently at the entrance to the second strand along the polypeptide chain. Thus, the presence of glycines both in the region of β-hairpin bend and in the connection is required to generate strongly twisted and coiled β-hairpin.

## DISCUSSION

The interaction between the amino acid sequence and spatial structure of the protein remains an intriguing and unresolved problem of molecular biology. It was established more than 40 years ago that the distribution along the chain of hydrophobic and hydrophilic residues plays a key role in the coding of α-helixes and β-strands. Based on this, prediction methods of the secondary structure of the proteins were developed that had rather high accuracy [14, 15]. Later, it became clear that the secondary structure generation also depends on other factors, first of all, on the structural context, i.e., on interactions with other protein-chain participants. The mutual location of α-helixes and/or β-strands in the space is also determined by their interactions with other secondary structure elements, including with connections that connect them [16].

The structural motifs of two α-helices or two β-strands connected by connections (such as, for example, α- and β-hairpins, αα-corners, ββ-arcs, etc.) are the simplest objects for studying all types of interactions between the elements and the interconnection between the structure and amino acid sequence. It was previously demonstrated that each such structural motif has a location along the chain of key hydrophobic, hydrophilic, and glycine residues that is specific only to it [16, 17]. It should be especially noted that the left and right α-hairpins that have a similar lengths and connection conformations have different locations of the key residues along the chain [18, 19]. The same was demonstrated for left and right



**Fig. 5.** Structural alignment of amino acid regions from Wrap-proteins encoding strongly twisted and coiled β-hairpins with short connections. Amino acid residues on internal positions of β-hairpin are underlined. Glycine is designated by the G letter. Conformations of residues in columns are indicated from above by β, αL, γ, ε symbols. Protein PDB codes are indicated in left column.

β-hairpins [17]. In the present work, a significant difference between right strongly twisted and coiled β-hairpins and right flat or weakly twisted β-hairpins was demonstrated. As was previously demonstrated, in the sequence of β-hairpins, the presence of strictly defined location of key hydrophobic, hydrophilic, and glycine residues in the chain (i.e., a certain pattern) inherent to each type of β-hairpin is a necessary condition for generating usual β-hairpins (flat and weakly twisted) [16, 17]. In this work, we demonstrated that, in addition to the implementation of these necessary

conditions, the presence of additional glycine residues and other small residues (alanine, serine) on concave surfaces and prolines on external (convex) surfaces is required for generating strongly twisted and coiled β-hairpins.

Thus, the unique spatial structure of strongly twisted and coiled β-hairpins (right double helixes always generated by right β-hairpins) and the above-described interconnection between the structure and amino acid sequence allows us to hypothesize that these β-hairpins can fold by themselves independently of the other part of the chain and can be embryos or ready structural blocks during the protein folding.

## ACKNOWLEDGMENTS

## REFERENCES

1. Chothia C. 1973. Conformation of twisted β-pleated sheets in proteins. *J. Mol. Biol.* **75**, 295−302.

2. Chothia C. 1983. Coiling of β-pleated sheets. *J. Mol. Biol.* **163**, 107−117.

3. Nishikawa K., Scheraga H.A. 1976. Geometrical criteria for formation of coiled-coil structures of polypeptide chains. *Macromolecules.* **9**, 395−407.

4. Efimov A.V. 1991. Structure of coiled β−β-hairpins and β−β-corners. *FEBS Lett.* **284**, 288−292.

5. Sayle R.A., Milner-White E.J. 1995. RASMOL: Biomolecular graphics for all. *Trends Biochem. Sci.* **20**, 374−376.

6. Gordeev A.B., Kargatov A.M., Efimov A.V. 2010. PCBOST: Protein classification based on structural trees. *Biochem. Biophys. Res. Commun.* **397**, 470−471.

7. Murzin A.G., Brenner S.E., Hubbard T. and Chothia C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536−540.

8. Tatusova T.A., Madden T.L. 1999. Blast 2 sequences: A new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174**, 247−250.

9. Boshkova E.A, Gordeev A.B, Efimov A.V. 2014. A novel structural tree for wrap-proteins, a subclass of (α+β)-proteins. *J. Biomol. Struct. Dyn.* **32**, 222−225.

10. Koradi R., Billeter M., Wuthrich K. 1996. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51−55.

11. Efimov A.V. 1986. Standard polypeptide chain conformations in irregular protein regions. *Mol. Biol.* (Moscow). **20**, 250−260.

12. Efimov A.V. 1992. A novel super-secondary structure of β-proteins. A triple-strand corner. *FEBS Lett.* **298**, 261−265.

13. Efimov A.V., Boshkova E.A. 2014. Two mechanisms of protein folding: A theoretical analysis (review article). *Russ. J. Bioorg. Chem.* **40** (6), 612 −619.

14. Schiffer M., Edmundson A.B. 1967. Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophys. J.* **7**, 121−135.

15. Lim V.I. 1974. Algorithms for prediction of α-helical and β-structural regions in globular proteins. *J. Mol. Biol.* **88**, 873−894.

16. Efimov A.V. 1993. Standard structures in proteins. *Progr. Biophys. Mol. Biol.* **60**, 201−230.

17. Efimov A.V. 1993. Patterns of loop regions in proteins. *Curr. Opin. Struct. Biol.* **3**, 379−384.

18. Efimov A.V. 1991. Structure of α−α-hairpins with short connections. *Protein Eng.* **4**, 245−250.

19. Brazhnikov E.V., Efimov A.V. 2001. Structure of α−α-hairpins with short connections in globular proteins. *Mol. Biol.* (Moscow). **35** (1), 89−97.

*Translated by A. Barkhash*