ISSN 0006-2979, Biochemistry (Moscow), 2010, Vol. 75, No. 10, pp. 1258-1263. © Pleiades Publishing, Ltd., 2010.
Published in Russian in Biokhimiya, 2010, Vol. 75, No. 10, pp. 1417-1423.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM10-107, September 26, 2010.

Structures Closed into Cycles in Proteins Containing 3β-Corners

E. A. Boshkova and 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 April 21, 2010

Abstract—In the present study, pathways of growth of protein structures represented in the structural tree for β-proteins containing 3β-corners are analyzed. It is shown that the frequency of occurrence of the completed structures of known proteins within branches of the tree is quite different. This means that allowed pathways of growth of protein structures are not equal and their usage is quite different. In most cases, addition of one or two β-strands nearest along the chain to the root 3β-corner (67%) or addition of three β-strands to the 3β-corner results in the formation of structures closed into cycles or barrels. Therefore, the pathways that result in closed structures are used most often in the first steps of growth of the root 3β-corner. Amino acid sequences coding for left-handed superhelices that close into cycles the 3β-corners are also analyzed. It is demonstrated that most crossover sites where the polypeptide chain passes from one β-layer to the other have one or two residues in sterically constrained α_L- or ε-conformations, which should be glycines or residues with flexible side chains in order to reduce the steric constraints.

DOI: 10.1134/S000629791010007X

Key words: protein structure modeling, protein folding, structural trees

The sequence of events in protein folding is one of the main determinants of unique protein tertiary structure. However, there are still no experimental methods to observe protein folding in real time. Theoretical modeling of protein folding based on construction and analysis of structural trees of proteins is a promising approach to solving the problem [1, 2]. The approach is based on the hypothesis that in the first step of protein folding a nucleus is formed, and then the remaining part of the molecule or domain is folded around it. In modeling, the structural motif having a unique overall fold and handedness is taken as the starting structure or the root structure of the tree. Other α-helices and/or β-strands are added to the root structural motif step-by-step in accordance with a restricted set of stereochemical rules inferred from known principles of protein structure. At each step there are several pathways of structure growth, but the number of allowed pathways is limited since the rules drastically reduce it. A general scheme that represents the root structural motif, all the intermediate and completed structures connected by lines showing allowed pathways of structure growth is referred to as the structural tree.

In this study, we analyzed the pathways of growth of structures in the structural tree for β-proteins containing 3β-corners [3, 4] and found the following.

1. The allowed pathways of growth of protein structures are not equal and their usage is quite different.

2. The first steps of growth of the root 3β-corner, i.e. addition to it of one or two β-strands nearest along the chain, lead to the formation of structures closed into cycles in most cases (67%).

These results are of particular value in both protein folding and modeling as the selected pathway of structure growth determines to a great extent the polypeptide chain fold of the completed structure.

METHODS OF INVESTIGATION

The 3β-corner is a structural motif that can be represented as a Z-like triple-stranded 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 [3]. The 3β-corners are common in proteins. To date we have compiled a database of 720 proteins containing 3β-cor-

* To whom correspondence should be addressed.

ners (among them 224 are nonhomologous, in which there are 253 3β-corners). This is a substantially larger set of protein structures than that presented in the corresponding structural tree constructed in 1997 [4]. Proteins were manually selected from Protein Data Bank (http://www.rcsb.org/pdb/). Protein structures were visually examined using the RasMol molecular graphics program [5]. Possible homologies were revealed by the BLAST 2 SEQUENCES program (http://blast.ncbi.nlm.nih.gov/bl2seq/wblast2.cgi) [6]. Taking into account the new database, we have constructed an updated structural tree for proteins containing 3β-corners. The updated tree includes 720 proteins (among them 224 are nonhomologous). Note that the tree of 1997 [4] contained 59 proteins including homologous ones. The updated tree is based on the tree of 1997 and incorporates it, but it has several additional novel branches and contains more known proteins. A computer version of the tree has been constructed that will be accessible at the web-site http://strees.protres.ru/ (Institute of Protein Research, Russian Academy of Sciences) where six structural trees for other protein classes are now accessible. This structural tree was used for comparative analysis of the pathways of growth of protein structures. The frequencies of occurrence of the closed and open structures were calculated taking into account the numbers of known nonhomologous proteins found within the corresponding branches of the tree.

RESULTS AND DISCUSSION

Figure 1 shows a schematic representation of the 3β-corner as well as open and closed structures obtained by addition to it of one or two β-strands nearest along the chain. For comparison, "top" (Fig. 1e) and end-on (Fig. 1j) views of the abcd-unit [1] are shown. The abcd-unit is closed into a cycle by the right-handed bcd-superhelix formed by β-strands b, c, and d. Analogous right-handed superhelices can be observed in the structure shown in Fig. 1c if an imaginary crossover strand would connect the C-end of β-strand 1 and the N-end of β-strand 3′, as well as in the structure shown in Fig. 1d if an imaginary crossover strand would connect the C-end of β-strand 2 and the N-end of β-strand 4. Similarity between these structures and the abcd-unit can easily be observed if one compares their simplified representations shown in Fig. 1, h-j. The structure shown in Fig. 1c will be referred to

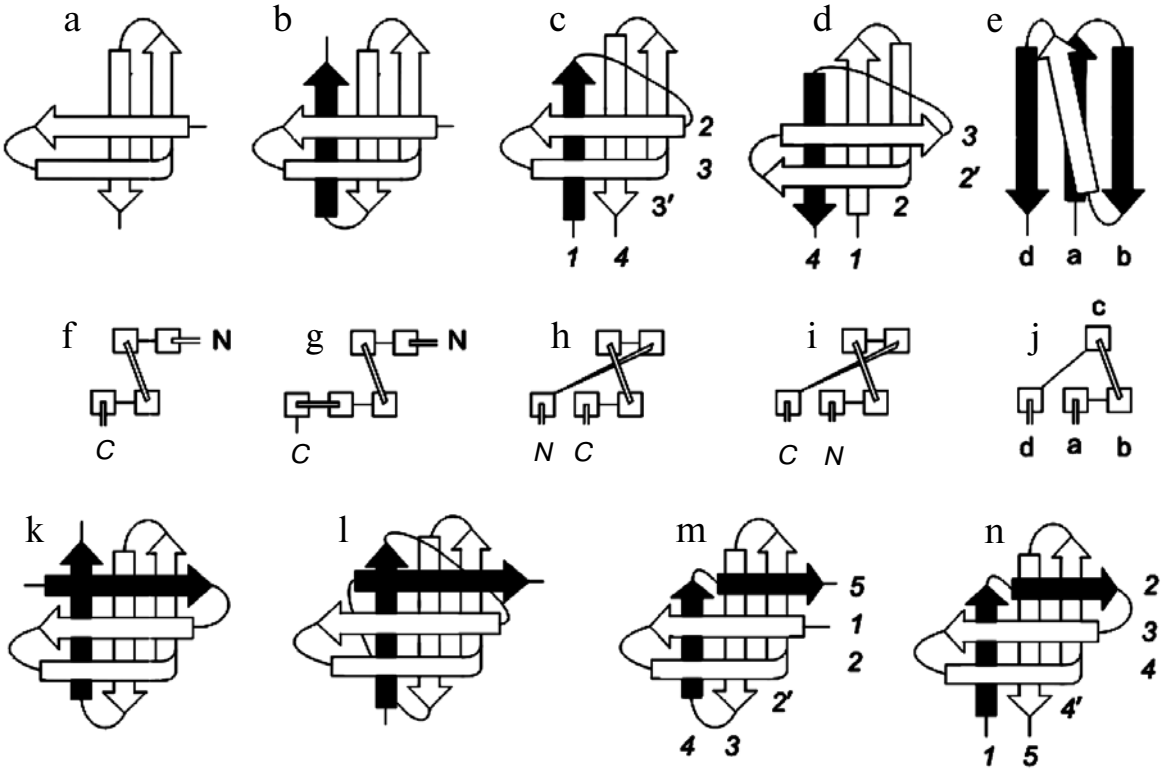

Fig. 1. Schematic representation of the 3β-corner (a) as well as open (b, k) and closed (c, d, l, m, n) structures obtained by addition to the 3β-corner of one (upper row) and two (bottom row) β-strands. On the right (e) the abcd-unit [1] is shown for comparison. β-Strands are shown with arrows directed from the N- to C-ends, and β-strands added to the 3β-corner are shown with black arrows, and loops are shown with single lines. In the middle row there are simplified representations of the structures of one upper row as they look from one side (f-j). Here β-strands are shown as rectangles, near connections by double lines, and far connections by single lines. N and C are designations of the ends. See also the text.

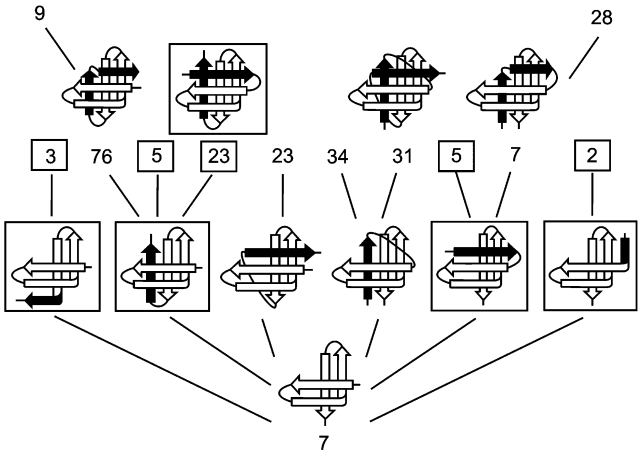

Fig. 2. A fragment of the structural tree for proteins containing 3β-corners and the frequencies of occurrence of known proteins in the corresponding branches. The structures are represented in a way similar to those in Fig. 1. The open structures and their frequencies of occurrence are framed. The unframed numbers show the frequencies of occurrence of the closed structures.

here as the 3β-corner closed at the N-end, and that in Fig. 1d will be referred as the 3β-corner closed at the C-end. In the structure shown in Fig. 1l, the 3β-corner is closed at both the N- and C-ends. The structures having the SH3-like fold (Fig. 1m) [7] and the GroES-like fold (Fig. 1n) [8] include right-handed superhelices of the (β+S+β)-type [9], which make these folds closed. In the former case (Fig. 1m), the right-handed superhelix is formed by β-strand 2, the S-like β-sheet (strands 2′, 3, and 4), and β-strand 5. In the second case (Fig. 1n), the superhelix is formed by β-strand 1, the S-like β-sheet (strands 2, 3, and 4), and β-strand 4′. In the literature the SH3-like and GroES-like folds are usually called β-barrels because their β-strands are closed into cylinders by a system of hydrogen bonds. Note that in the closed structures the N- and C-end are adjacent and the N- and C-terminal β-strands are cross-stitched by hydrogen bonds.

In the open structures shown in Fig. 1 (b, g, k), the β-strands are added to the 3β-corner without passing of the polypeptide chain from one β-layer to the other. In these open structures, the N- and C-terminal β-strands are located in different β-layers and do not form hydrogen bonds between each other. In other open structures (see Fig. 1 in [4]), the polypeptide chain passes from one β-sheet to the other while bending by 90° to form a right-handed superhelix, but their N- and C-terminal β-strands are located apart.

We have analyzed the pathways of growth of the root 3β-corner up to the completed structures of known proteins and calculated the frequencies of occurrence of the closed and open structures in our database of proteins containing 3β-corners. We have found that the number of the completed structures of known proteins within branches of the tree is quite different. This means that the pathways are not equal and their usage is different. The fragment of the structural tree shown in Fig. 2 represents the structures of the first level obtained as a result of addition of one β-strand to the root 3β-corner and some structures of the second level as well as their frequencies of occurrence in the corresponding known proteins. As seen, 35% of 3β-corners (88 3β-corners out of 253 non-homologous ones) are closed at the first step of growth when the β-strand nearest to the 3β-corner along the chain is added to it at the C- or N-end (compare with Fig. 1, c and d). In the second step, 31 out of these 88 closed 3β-corners are closed twice, at the C- and N-ends (Fig. 1l). Thirty-two percent of 3β-corners are closed when the second β-strand is added to the root 3β-corner so as to form SH3-like or GroES-like folds (Fig. 1, m and n). Fifteen percent of 3β-corners are closed when the third β-strand is added. In total, 82% of 3β-corners form closed structures after three steps of their growth.

Thus, the pathways that result in the closed structures are predominantly used at the first steps of growth of the root 3β-corner. One of the most probable reasons of this is that the closed structures are more cooperative and higher barriers should be overcome to unfold them compared with their open analogs.

As mentioned above, the central β-strand of the 3β-corner forms half a turn of the right-handed superhelix when passing from one β-sheet to the other (see Figs. 1a and 3a) [3]. Such superhelices occur most often in proteins with the orthogonal β-sheet packings [10], including proteins containing 3β-corners [3, 4], S-like β-sheets [11]

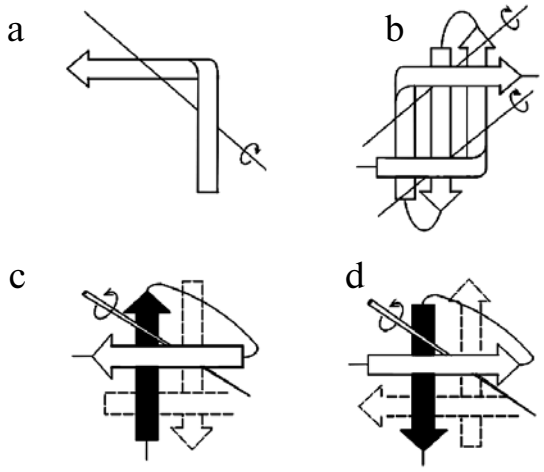

Fig. 3. Handedness of crossovers of the polypeptide chain from one β-sheet to the other in β-proteins with orthogonal β-sheet packings. a) Schematic representation of the right-handed superhelix that is formed by the central β-strand of the 3β-corner [3], the β-bend [10], β-strands of the β-β-corner [12], etc. b) The right-handed (β+S+β)-superhelix [9]. c, d) Left-handed superhelices that close into cycles with 3β-corners at the N- and C-ends. Imaginary axes of the superhelices are shown as straight lines (a, b) and bars (c, d). Other designations are as in Fig. 1.

a

ββββββββββαααααααααααα$\alpha_r\gamma\alpha_L$ββ-------βββββββββ

```
1.  1v6z:A  -------Gvlplretrhlve--vlRar-----vGdrFtVf----  13-38
2.  1uhe:A  -------Itidedlak------laKlr-----eGmkVeIv----  26-48
3.  1vle:S  --yKyWiMrvnsidae------arGik-----nGdlIrAyn---  767-794
4.  1wlf:A  -------Lhlprrlva------qlHll-----qNqaIeV-----  29-50
5.  1e32:A  ------vVslsqpkmd------elQlf-----rgdtVlLk----  38-60
6.  1mbm:A  ----eVvVltashvvG------raNma--------tLkI-----  31-51
7.  1vfs:A  --tHlAlVpagy----------adGip-rnasGrgpVlV-----  280-305
8.  1y12:A  ---LsFtKyidkstpnlmmac-ssGkh--------yPqAkLtIr  56-87
9.  1arb:A  ------yFltah----------hcGmgtastaasivVy------  51-72
10. 1zbo:A  -GvLrLrVfe-iryldmvrrciadGse-------fgVvVl----  17-47
11. 1efc:A  kFeSeVyIls-k----------deGgrhtpffkgyrPqFyF---  303-332
```

b

βββββββββ----------$\beta\beta\beta_P\alpha_L\beta_P$ βββββββ

```
1.  1kzl:A  --gFaMkIeapqildd-------chtGd sIaV--     22-44
2.  1t62:A  ----aTcSsldiykmeeeql---pkaGq yDiIl-     1050-1075
3.  1uhe:A  -----ItIdedlaklaK------lreGm kVeIvD     26-48
4.  2ar1:A  ----tSpWdGvrnyaarnnmra-msvGd kVlFyH     33-61
5.  1v6z:A  -------GvlplretrhlvevlRarvGd rFtVf-     13-38
6.  1vly:A  ---WlLaGsasrl----------peaGe dLeLk-     249-268
7.  1yln:A  --gCrFiTpplgkt---------yqvGd lVaLeI     167-189
8.  2f9h:A  -----ViQsfpekdqvt------laeGd hLkI--     40-60
9.  1o65:A  -----LsTdGltesn--------vymGd iFrW--     98-116
10. 1wmm:A  ----vWgVpkkhkntlsr-----vkpGd kLvIyV     21-45
11. 1zun:A  sDaFdAmLvwmaeep--------mlpGk kYdIk-     332-356
12. 1ep3:B  --iFeMvLkGtlvdem-------dlpGq fLhL--     20-42
13. 1w0j:A  ----iArVhGLrn----------vqaEe mVeF--     38-55
14. 2ey4:C  ----fLiVrtnwv----------pslNd rVv---     15-31
15. 1oru:B  -----VaVsGmpdlts-------lkeGs rIi---     95-113
16. 1e0t:A  ------mVavtyegfttd-----lsvGn tVlV--     105-125
17. 1vle:S  yKyWiMrVnsidaearG------iknGd lIrAyN     767-794
18. 1wlf:A  -----LhLprrlvaqlHl------lqNq aIeVa-     29-50
19. 1cr5:A  ----vAaVspnd----------fpnNi yIiI--     44-60
20. 1gtr:A  ----lViKqgfaepslkd-----avaGk aFqF--     496-518
21. 1k0h:A  -------PslfvrtDevrq----lrrGd tLtI--     63-83
22. 1s04:A  ---KiEgRlydekrrq-------ikpGd vIsF--     21-42
23. 1te7:A  ---TiTiRdesesh---------fktGd vLrVgr     22-43
24. 1mbm:A  -------Clawtts---------gdsGs aVvQ--     111-126
25. 1sqr:A  ----vMiIkpldVnsreeas---kliGr lVlW--     21-45
```

Fig. 4. Structural alignment of amino acid sequences coding for left-handed superhelices closing 3β-corners at the N-ends. Each column is headed by a symbol, α, β, α_L, γ, showing the conformation of the residues in it; a dash shows that the residues in this column have various conformations. PDB codes of proteins are listed on the left and residue numbers of the sequences are shown on the right. Residues having α_L-conformations as well as inside residues of β-strands are shown in capital letters. a) Examples containing α-helices in the crossover regions. b) Examples having $\beta\beta\beta_p\alpha_L\beta_p$-arches.

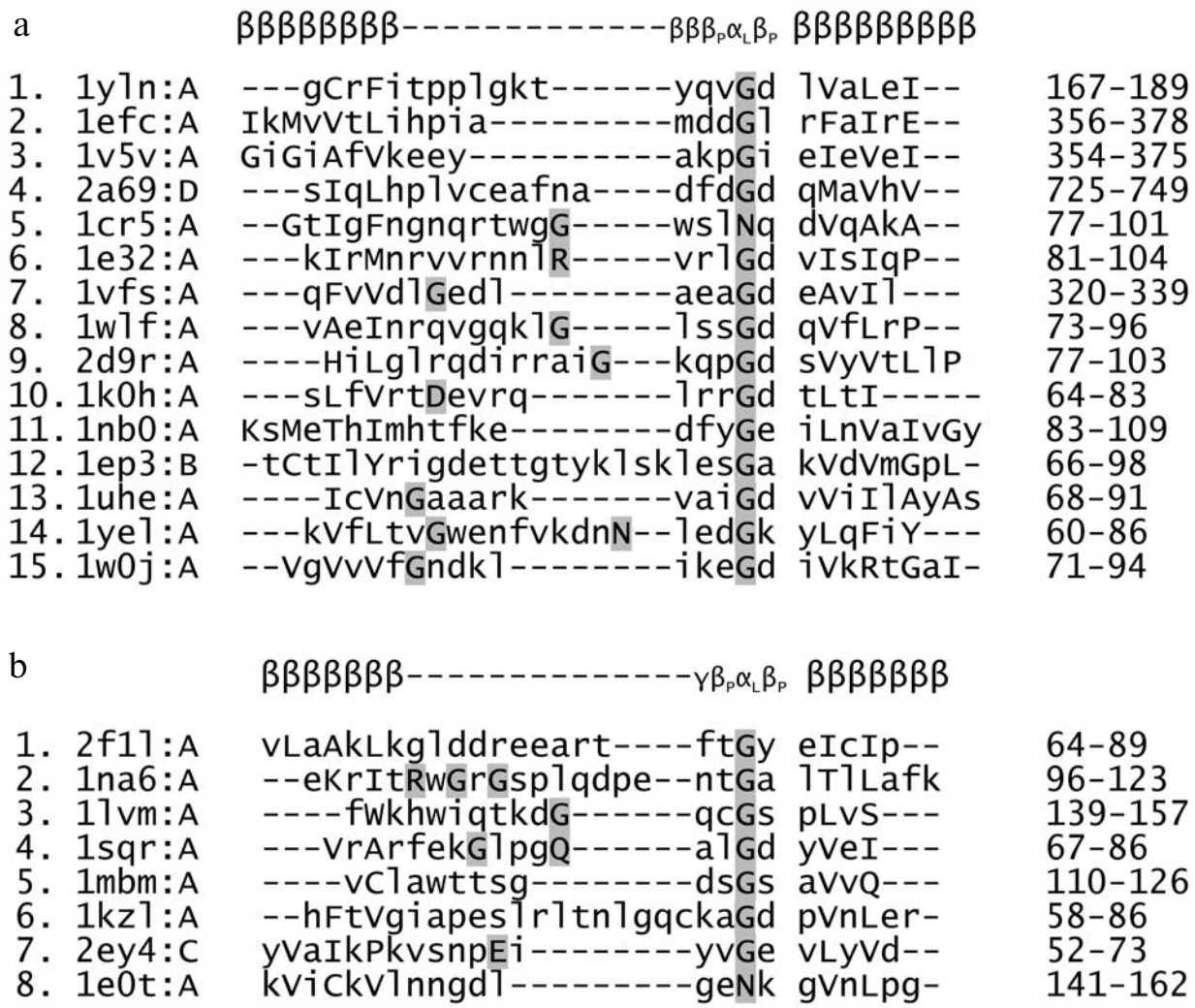

Fig. 5. Structural alignment of amino acid sequences coding for left-handed superhelices that close 3β-corners at C-ends. a) Examples having $\beta\beta\beta_p\alpha_L\beta_p$-arches. b) Examples having $\gamma\beta_p\alpha_L\beta_p$-structures. Designations are as in Fig. 4.

(Figs. 1m, 1n, and 3b), and β-β-corners [12]. Analysis shows that at the bending site, i.e. at the site where the polypeptide chain passes from one β-sheet to the other, it can adopt a characteristic conformation forming what is called the β-bend [10], or another standard structure with βαββ-, ββα$_L$β-, βαγβ-, βααγβ-, or βεβ-conformation [12, 13] (γ-conformation corresponds to a region with $\varphi = -90 \pm 30°$, $\psi = 0 \pm 30°$; and ε-conformation to $\varphi = 110 \pm 30°$, $\psi = -170 \pm 20°$). Each of these small standard structures provides a 90° bend and a crossover of the polypeptide chain from one β-layer to the other.

In contrast, addition of the β-strand nearest along the chain to the 3β-corner so as to form a closed structure (Fig. 1, c and d) results in the formation of a left-handed superhelix at the crossover site (Fig. 3, c and d). We have analyzed the left-handed superhelices in our database and found that most crossovers of the polypeptide chain from one β-sheet to the other (84% of 3β-corners closed at N-ends and 75% of 3β-corners closed at C-ends for the non-homologous proteins) have one, two, or more residues in the sterically constrained α$_L$- or ε-conformations. More than 80% of these α$_L$- and ε-positions are occupied by glycines, and the remaining 20% are occupied by residues with flexible side chains. It can be concluded that the formation of left-handed superhelices in the closed 3β-corners results in steric constraints at the crossover sites, and to reduce them in proteins the α$_L$- and ε-positions are occupied by glycines or residues with flexible side chains and cannot be occupied by Pro, Thr, or bulky hydrophobic residues.

As a rule, the polypeptide chain regions forming the bending site or the crossover of the chain from one β-sheet to the other in the left-handed superhelices are longer than in the right-handed ones. In order to find the features of amino acid sequences coding for the left-handed superhelices, we performed structural alignment

of nonhomologous superhelices subdivided into several groups with similar structures. The structural alignment was performed by hand. Amino acid residues occupying equivalent positions in the superhelices and having the same conformations are arranged column-wise (Fig. 4). Each column is headed by the symbol showing the conformation of the residues. For example, all the C-terminal α_L-residues of α-helices fall into one column headed by α_L, the N-terminal residues of α-helices into another column, the inside and outside β-residues into the corresponding columns, etc. In many cases the crossover region consists of an α-helix and one or two small standard structures (Fig. 4a). Note that C-terminal residues of α-helices having the α_L-conformations are glycines or residues with flexible side chains. However, in most left-handed superhelices the crossover region can be represented as a combination of several small standard structures. The β-β-arch occurs most often. This standard structure has been observed earlier in β-proteins with the aligned β-sheet packings in which the β-β-arch provides a crossover of the chain from one β-sheet to the other [14]. Its shortest variant consists of five residues and has $\beta\beta_p\beta_p\alpha_L\beta$-conformation ($\beta_p$-conformation corresponds to the polyproline helix region on the Ramachandran plot). In the left-handed superhelices that close 3β-corners, such β-β-arches have $\beta\beta\beta_p\alpha_L\beta_p$-conformations (Figs. 4b and 5a), which is slightly different from that mentioned above. Note that the majority of the first β-positions of these β-β-arches are occupied by hydrophobic or nonpolar residues and most α_L-positions by glycines. Figure 5b shows some examples of the left-handed superhelices containing small structures with $\gamma\beta_p\alpha_L\beta_p$-conformations in which α_L-positions are also occupied by glycines.

Figures 4 and 5 represent examples of elements of the crossover regions that occur most often in the left-handed superhelices, although the number of possible combinations of the small standard structures in such regions is rather large. This variety is one of the main difficulties in studying the relationship between the structure and the amino acid sequence of the crossover regions. Nevertheless, as can be seen in Figs. 4 and 5, the amino acid sequences of the crossover regions should have glycines or residues with flexible side chains in particular positions. Moreover, there should be fulfilled necessary conditions of the formation of β-strands flanking the crossover region, in part, their inside positions pointed to the hydrophobic core should be occupied by hydrophobic residues and those pointed outside by both hydrophobic and hydrophilic residues.

It should be noted in conclusion that structures closed into cycles are widespread in proteins. In many cases, superhelices of different types close the structures into cycles. For example, these are right-handed βαβ-superhelices in α/β-proteins [15] and (α+β)-proteins containing abCd-units (see the corresponding structural motifs and trees in [2]), superhelices bcd in abcd-units [1], (β+S+β)-superhelices in β-proteins with the orthogonal β-sheet packings [4, 9, 11], as well as the superhelices considered in this paper. It seems likely that in protein folding the formation of closed structures is a mechanism for obtaining cooperative and stable structures and the folding pathways that result in the closed structures are predominantly used.

This work was supported in part by the Russian Foundation for Basic Research (grant No. 10-04-00727) and the Federal Agency for Science and Innovations (project No. 02.740.11.0295).

REFERENCES

1. Efimov, A. V. (1982) *Mol. Biol. (Moscow)*, **16**, 799-806.
2. Efimov, A. V. (1997) *Proteins*, **28**, 241-260.
3. Efimov, A. V. (1992) *FEBS Lett.*, **298**, 261-265.
4. Efimov, A. V. (1997) *FEBS Lett.*, **407**, 37-41.
5. Sayle, R. A., and Milner-White, E. Y. (1995) *Trends Biochem. Sci.*, **20**, 374-376.
6. Tatusova, T. A., and Madden, T. L. (1999) *FEMS Microbiol. Lett.*, **174**, 247-250.
7. Masacchio, A., Noble, M., Pauptit, R., Wierenga, R., and Saraste, M. (1992) *Nature*, **359**, 851-855.
8. Hunt, J. F., Weaver, A. J., Landry, S. J., Gierasch, L., and Deisenhofer, J. (1996) *Nature*, **379**, 37-45.
9. Efimov, A. V. (1993) *FEBS Lett.*, **334**, 253-256.
10. Chothia, C., and Janin, J. (1982) *Biochemistry*, **21**, 3955-3965.
11. Efimov, A. V. (1998) *FEBS Lett.*, **437**, 246-250.
12. Efimov, A. V. (1991) *FEBS Lett.*, **284**, 288-292.
13. Efimov, A. V. (1986) *Mol. Biol. (Moscow)*, **20**, 250-260.
14. Efimov, A. V. (1987) *FEBS Lett.*, **224**, 372-376.
15. Rao, S. T., and Rossmann, M. G. (1973) *J. Mol. Biol.*, **76**, 241-256.