ISSN 0006-2979, Biochemistry (Moscow), 2010, Vol. 75, No. 2, pp. 249-256. © Pleiades Publishing, Ltd., 2010.
Published in Russian in Biokhimiya, 2010, Vol. 75, No. 2, pp. 305-312.

A Novel Structural Motif and Structural Trees for Proteins Containing It

A. M. Kargatov and A. V. Efimov*

Institute of Protein Research, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia; fax: (495) 632-7871; E-mail: efimov@protres.ru

Received April 15, 2009

Abstract—In the present study, a novel structural motif that can be represented as a combination of the known βαβ-unit and ψ-motif is described and analyzed. In theory, there are four possible combinations of the motifs since each of them can exist in two forms, left-handed and right-handed. For this study, we have selected 140 nonhomologous proteins in which 158 combinations of such types have been found. The combination of the right-handed ψ-motif and the right-handed βαβ-unit has been shown to occur most often (87 cases out of 158) and the combination of the left-handed βαβ-unit and the left-handed ψ-motif does not occur at all. Three novel structural trees in which the commonly occurring combinations are taken as the root structures have been constructed.

DOI: 10.1134/S0006297910020161

Key words: βαβ-unit, modeling, ψ-motif, protein folding, structural classification

The structural motif is a super-secondary structure formed by two or more α-helices and/or β-strands that recurs many times within one protein molecule or often occurs in nonhomologous proteins. Structural motifs of a given type have the same number of α-helices and/or β-strands, the same or similar arrangement of them in the chain and in three dimensions, and the same overall fold. While many different structural motifs have been observed to recur within globular proteins, only some of the motifs exhibit unique handedness and a unique overall fold (for a review, see [1]).

Many small proteins and domains merely consist of structural motifs with unique overall folds [2]. All this taken together suggests that the structural motifs are relatively stable, can fold into a unique structure per se, and consequently, can act as nuclei in the folding of larger proteins. On the other hand, the structural motifs are very suitable to be taken as starting structures in protein modeling and as root structures in construction of structural trees. As analysis shows, the protein structures of many known large protein superfamilies can be obtained by a stepwise addition of secondary structural elements to the corresponding root structural motifs taking into account a restricted set of simple rules [3, 4]. If possible folding pathways are shown by lines that connect all the starting (root), intermediate, and final structures between each other, one common folding scheme can be obtained that is referred to as a structural tree. Structural trees can be used for solving several problems such as protein structure comparison, structural classification of proteins, protein folding and modeling, searching for all possible protein folds both known and unknown, etc.

In this study, we consider a novel structural motif that has been found in α/β-proteins as well as in (α+β)-proteins. In accordance with the structural classification [5], the α/β-proteins are composed of α-helices and β-strands that strongly alternate along the polypeptide chain and are arranged into the so-called Rossmann's folds [6]. The minimal commonly occurring folding unit in α/β-proteins is the βαβ-unit [6, 7]. The novel structural motif considered in this paper can be represented as a combination of the known ψ-motif [8, 9] and the βαβ-unit. Here we have analyzed all possible variants of this motif and constructed three novel structural trees for proteins containing them.

* To whom correspondence should be addressed.

METHODS OF INVESTIGATION

The combination of the βαβ-unit and the ψ-motif was revealed as a result of visual inspection and protein structure comparison of the Protein Data Bank entries (http://www.rcsb.org/pdb/) [10]. In total, 140 nonhomologous proteins were selected in which 158 combinations of the motifs were found. Possible homologies were revealed by the BLAST pairwise alignment (http://blast.ncbi.nlm.nih.gov/bl2seq/wblast2.cgi) [11]. PDB codes of proteins included in our database as well as the polypeptide chain regions forming the corresponding combinations are given in the table.

The structural trees were constructed taking into account the following general rules [3, 4]:

– overall folds of protein molecules and intermediate structures are taken into account and details of the structures are ignored. For space economy, only the pathways leading to known protein structures are represented;

– three commonly occurring combination of the ψ-motif and the βαβ-unit are taken as the root structures of the corresponding trees;

– larger protein and intermediate structures are obtained by a stepwise addition of β-strands and/or α-helices to the growing structure (in some cases, "ready building blocks", e.g. S-like β-sheets are added). At each step, the β-strand or α-helix nearest to the growing structure along the polypeptide chain is the first to be attached;

– α-helices and β-strands cannot be packed into one layer [12];

Proteins containing combinations of the βαβ-unit and ψ-motif and their PDB codes (regions of polypeptide chains forming these combinations are shown in parentheses)

Combinations of right-handed βαβ-unit and right-handed ψ-motif	Combinations of right-handed βαβ-unit and left-handed ψ-motif	Combinations of right-handed ψ-motif and left-handed βαβ-unit
1	2	3
1A8Y (32-99, 147-198, 249-316)	1AUO (5-48)	1D6T (30-86)
1ABA (2-69)	1AZW (23-66)	1DAR (506-570)
1CLI (101-165)	1B6G (35-80)	1FWK (33-86)
1DBF (42-97)	1C1D (23-82)	1J5E (14-66) – chain I
1DT9 (319-412)	1C4X (18-63)	1K47 (35-92)
1G7E (55-122)	1CR6 (246-289)	1KKH (35-106)
1G7O (2-53)	1CRL (99-154)	1MG7 (66-122)
1GH2 (25-83)	1DIN (18-61)	1NHI (234-293)
1GHH (2-53)	1EA5 (96-145)	1OYS (35-120)
1GWC (7-61)	1F0N (20-71)	1PIE (58-128)
1H75 (3-56)	1HLG (39-96)	1PKP (92-127)
1HLE (111-192) – chain A	1HPL (37-105)	1PVG (288-355)
1HYU (21-68, 119-176)	1HRD (71-128)	1RHY (16-63, 108-162)
1ILO (2-56)	1IMJ (17-67)	1RRE (610-667)
1IPA (47-101)	1IVY (32-98)	1UEK (37-82)
1J23 (6-40)	1JFR (41-87)	2GO3 (23-97)
1JWQ (92-145)	1JJF (63-125)	3CWV (236-306)
1K0D (114-170)	1JKM (103-157)	
1K0M (8-69)	1JMK (10-49)	
1LCP (373-448)	1JU3 (21-73)	
1LFW (182-426)	1L7A (67-115)	
1LJR (4-59)	1LCP (228-305)	
1M2D (4-68)	1LFW (70-150)	
1MGP (254-309)	1M33 (6-45)	
1NHY (5-53)	1MO2 (59-101)	
1OFU (260-314)	1PCA (46-108) – chain B	
1OYJ (7-61)	1PV1 (27-84)	
1PCA (201-270) – chain B	1QFM (449-502)	
1PMT (2-56)	1QLW (50-103)	
1Q8R (55-116)	1QO7 (94-147)	

Table (Contd.)

1	2	3
1QGV (25-85)	1R3D (4-48)	
1QFN (2-64)	1R3N (96-155)	
1QMH (189-243)	1T3B (79-122)	
1QU9 (71-126)	1TCA (20-66)	
1R3N (240-413)	1THT (18-68)	
1S3A (16-69)	1TK3 (522-578)	
1SA0 (316-379)	1TQH (10-48)	
1SEN (57-118)	1UFO (13-56)	
1T0A (91-158)	1VHE (50-211)	
1T4Z (13-74)	1VIX (58-170)	
1TTZ (3-52)	1VKH (19-71)	
1TUB (314-380)	1XMB (94-169)	
1U6T (2-70)	1Z6M (17-65)	
1UC7 (31-95)	2AFW (118-198)	
1V9W (33-106)	2B20 (177-232)	
1VIX (202-371)	2BZ1 (31-89)	
1VK3 (436-502)	2G9D (41-94)	
1VQV (78-136)	2NSM (39-106)	
1W5B (284-338)	2OR4 (349-421)	
1WIK (17-75)	2VEO (66-141)	
1WJK (18-72)	3BF7 (5-48)	
1XG8 (1-76)	3C5V (63-110)	
1XMB (207-387)	3SC2 (28-92)	
1YGY (348-418)		
1Z6N (56-117)		
1Z9H (101-155)		
1ZMA (29-91)		
2A2P (39-85)		
2A2R (3-57)		
2A4H (75-120)		
2AXO (45-120)		
2B5E (51-112, 162-206, 259-318, 397-456)		
2E5A (45-81)		
2EWC (63-119)		
2FA8 (7-54)		
2FNO (8-64)		
2FT1 (274-327)		
2FUG (78-135) – chain B		
2G9D (166-222)		
2GLZ (46-87)		
2HFD (36-99)		
2HLS (28-92, 140-202)		
2HQS (103-172) – chain C		
2IAF (32-120)		
2JNB (51-105)		
2NSM (199-290)		
2OR4 (461-538)		
2OS5 (2-63)		
2PK8 (30-76)		
2TRC (135-193) – chain P		

– crossing of connections [13] and formation of knots [14] are prohibited, but formation of the ϕ- and ψ-motifs is permitted;

– all the structural motifs (not only the root motifs) of the obtained structures should have the corresponding overall folds and handedness;

– in accordance with the principle of close packing, the obtained structures should be compact.

RESULTS AND DISCUSSION

As mentioned above, the novel structural motif is a combination of the two known motifs, the ψ-motif and the βαβ-unit. The βαβ-unit is a two-layer structure in which one layer is formed by two parallel β-strands and the other – by an α-helix. As a rule, in proteins βαβ-units occur as right-handed super-helices [6, 7], and left-hand-

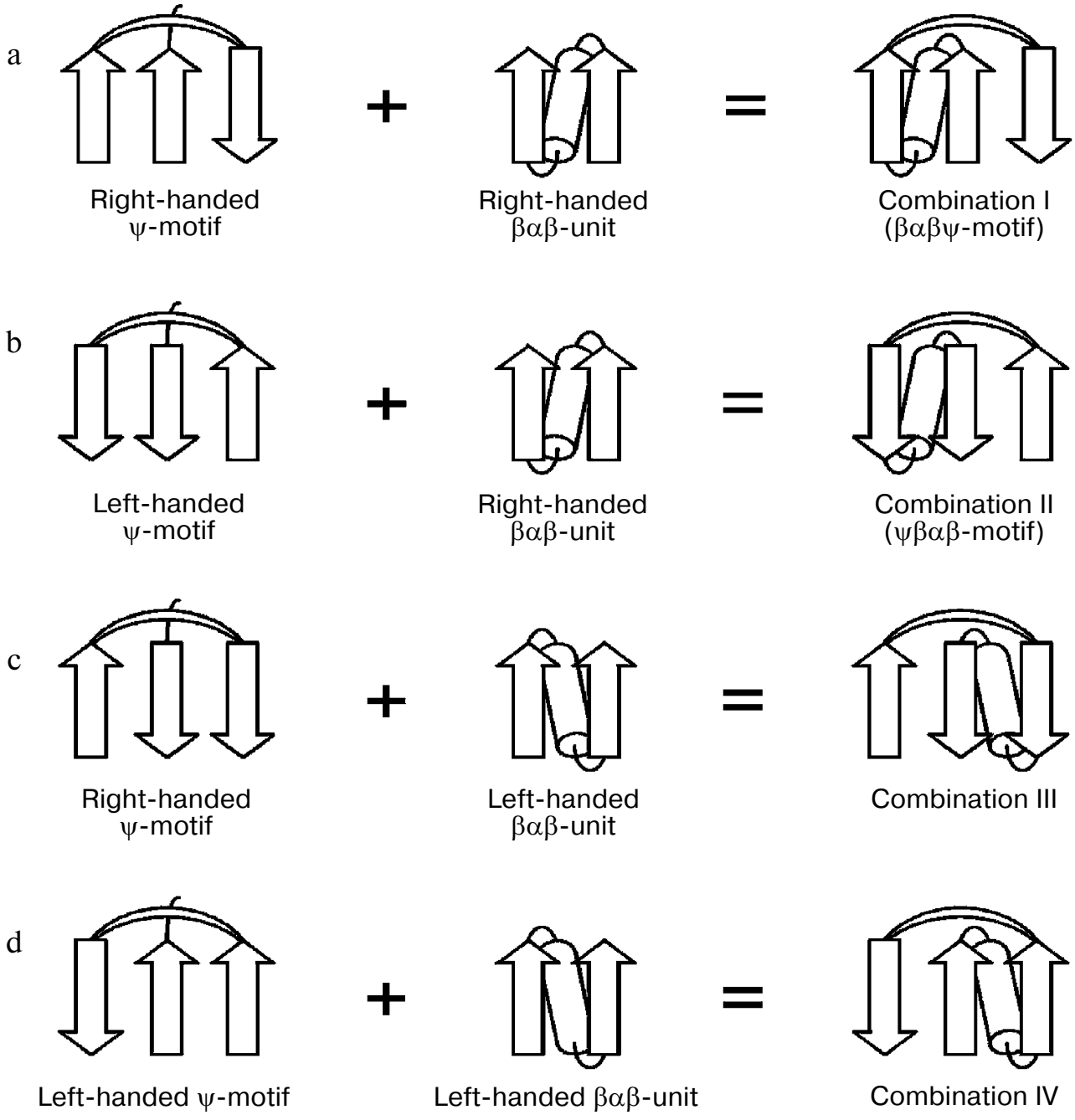

Fig. 1. Schematic representation of possible combinations of the βαβ-unit and ψ-motif: the right-handed βαβ-unit and right-handed ψ-motif (a), the right-handed βαβ-unit and left-handed ψ-motif (b), the left-handed βαβ-unit and right-handed ψ-motif (c), the left-handed βαβ-unit and left-handed ψ-motif (d). α-Helices are shown with cylinders, β-strands – with arrows directed from the N- to the C-ends, crossover loops – by double lines, and other loops – by single lines.

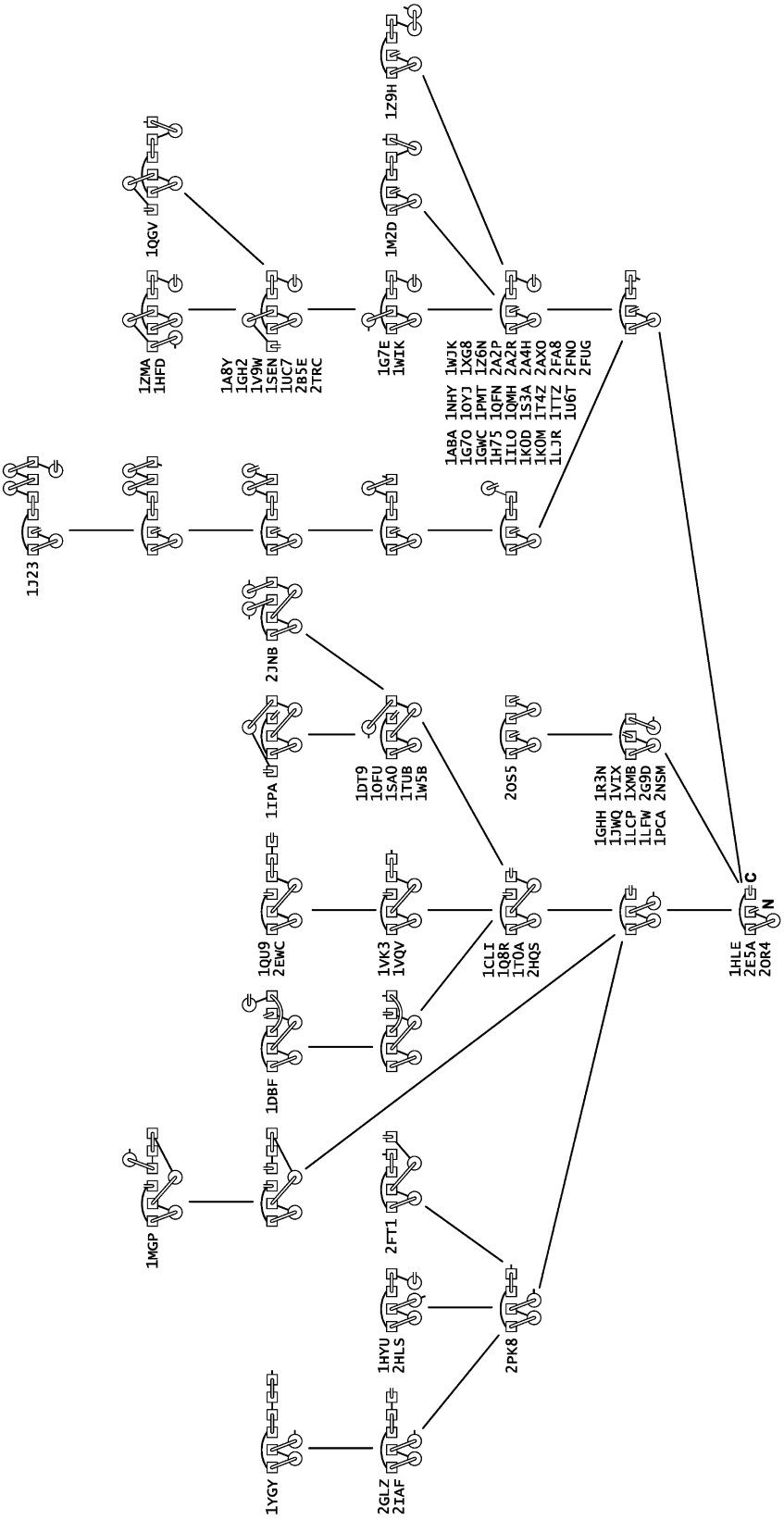

Fig. 2. Structural tree for proteins containing βαβψ-motifs. All the structures are oriented in a similar way and are viewed end-on with α-helices shown as circles and β-strands as rectangles. Near connections are shown by double lines and far connections by single lines. N and C are designations of the ends of the root structural motif. PDB codes show the final structures of the corresponding proteins and domains.

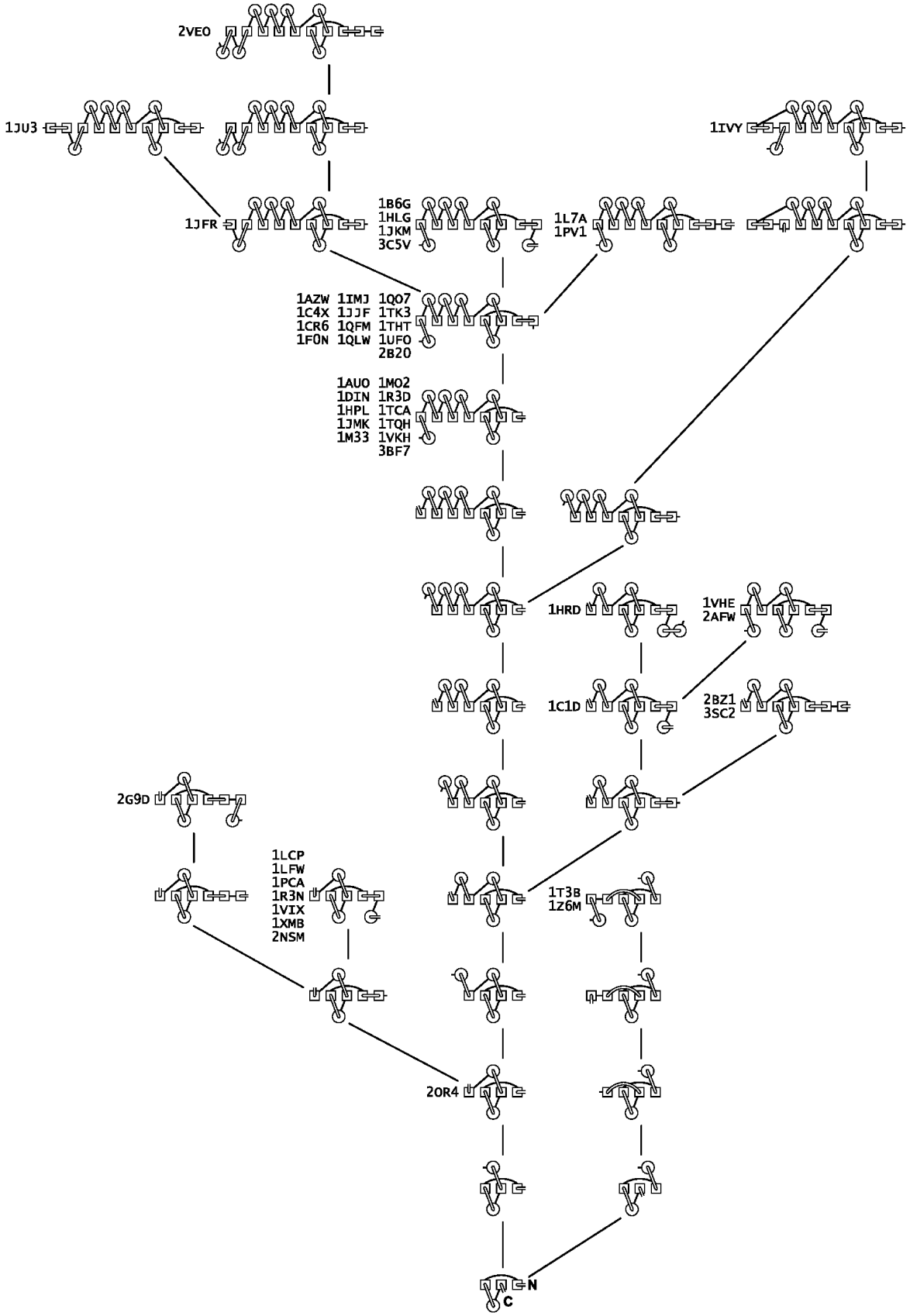

Fig. 3. Structural tree for proteins containing ψβαβ-motifs. Designations are the same as in Fig. 2.

ed βαβ-units occur rarely. The ψ-motif [8, 9] is formed by three β-strands packed in one β-sheet. It can be represented as a split β-hairpin that has an additional β-strand located between its two β-strands (Fig. 1). Here the loop that connects two edge β-strands of the ψ-motif will be referred to as a crossover loop (there is a similar crossover loop in the ϕ-motif [15]). The crossover loop crosses over the central β-strand or its extension of the ψ-motif so that together they form a cross-like structure. There are right-handed and left-handed ψ-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 ψ-motifs (Fig. 1, a and c) and in the anticlockwise direction in the left-handed ψ-motifs (Fig. 1, b and d).

Figure 1 represents four possible combinations of the ψ-motif and βαβ-unit. In proteins, frequencies of occurrence of these combinations are different and, in our database, they are as follows:

1) 87 (55%) combinations of the right-handed ψ-motif and right-handed βαβ-unit in 80 proteins;

2) 53 combinations of the left-handed ψ-motif and right-handed βαβ-unit (34%) in 53 proteins;

3) 18 combinations of the right-handed ψ-motif and left-handed βαβ-unit (11%) in 17 proteins;

4) no combinations of the left-handed ψ-motif and left-handed βαβ-unit at all.

Such frequencies of the combinations occurrence can be explained by the fact that the left-handed βαβ-units occur very rarely in proteins [6, 7] and because of they are unfavorable structures from the point of view of stereochemistry. In the double-psi β-barrel proteins and domains, the ψ-motif preferably occurs in one form [8, 9] (in accordance with our definition, as the right-handed ψ-motif; see Fig. 1). The ϕ-motif, which has a definite similarity to the ψ-motif, also occurs in the right-handed form (in 49 proteins out of 50) [15]. Nevertheless, it should be noted that although there are only 11% of the left-handed βαβ-units in the combinations, 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 such high frequencies of occurrence of the left-handed βαβ-units and left-handed ψ-motifs in the combinations are still poorly understood and should be investigated further.

In each combination shown in Fig. 1, two parallel β-strands are a part of both, the βαβ-unit and ψ-motif. It means that in the combinations the βαβ-unit and ψ-motif coexist as an indivisible structural motif composed of three β-strands and an α-helix. There are two types of the motifs (combinations), which have different distribution along the polypeptide chain of α-helical and β-structural regions that can be described by formulas βαββ (Fig. 1, a and d) and ββαβ (Fig. 1, b and c). The abCd-units have the same number and distribution along the chain of the α-helical and β-structural regions, which are also described as βαββ and ββαβ, but they have a different three-dimensional arrangement of their elements [1, 12]. To distinguish these commonly occurring motifs one from another, the combination in which the ψ-motif follows the βαβ-unit will be referred to as the βαβψ-motif (Fig. 1a) and the combination in which the βαβ-unit follows the ψ-motif as the ψβαβ-unit (Fig. 1b). It is of interest to note that the βαβψ- and ψβαβ-motifs have essentially the same overall fold if the polypeptide chain direction is ignored. Nevertheless the βαβψ-motif includes a right-handed ψ-motif and the ψβαβ-motif has a left-handed ψ-motif, and the first motif occurs in proteins more often than the other.

Figure 2 represents a structural tree for proteins containing βαβψ-motifs. The root motif of the tree is shown schematically below and possible pathways of its growth are shown with lines. The structure having a higher position in the tree is obtained by addition of a β-strand or an α-helix to the structure located lower in accordance with the abovementioned rules. If the structure has several possible pathways of growth it forms the so-called branching point of the tree. Proteins and domains of different branches have a structure located in the branching point as a common fold. Thus the structural tree shows possible pathways of growth as well as levels of structural similarity between proteins and domains.

Figures 3 and 4 represent two more structural trees for proteins containing ψβαβ-motifs and the combinations of left-handed βαβ-units and right-handed ψ-motifs, respectively. The construction and main features

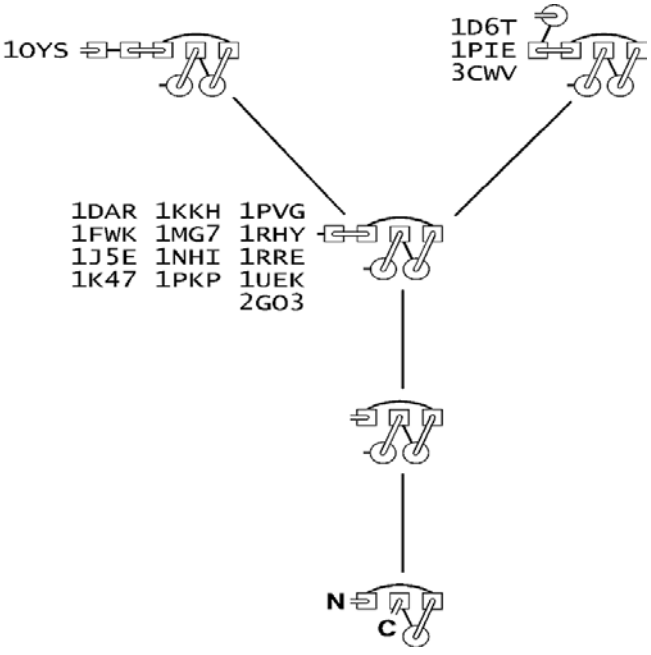

Fig. 4. Structural tree for proteins containing combinations of the right-handed ψ-motif and left-handed βαβ-unit. Designations are the same as in Fig. 2.

of these trees are very similar to that shown in Fig. 2. As seen, all the structures are oriented in a similar way and the root structural motifs are localized at the edges of two- or three-layer protein structures in most known proteins (shown with their PDB codes) except for some proteins and domains located in lower branches of the tree shown in Fig. 3. It should be noted that in proteins of other structural classes the root structural motifs have a strong tendency to be located at the edges of protein molecules as well [1, 3, 4]. As can be seen, the growth of the root βαβψ- and ψβαβ-motifs results in the formation of one, two, or more additional βαβ-units (Figs. 2 and 3). It means that most proteins and domains found within these structural trees are α/β-proteins and the remaining proteins can be classified as (α+β)-proteins in accordance with the classification by Levitt and Chothia [5]. Now taking into account the structural trees we can subdivide them into three subclasses, proteins containing βαβψ-motifs (Fig. 2), proteins containing ψβαβ-motifs (Fig. 3), and proteins containing combinations of left-handed βαβ-units and right-handed ψ-motifs (Fig. 4).

This work was supported in part by the Russian Foundation for Basic Research (grant No. 07-04-00659) and by the grant NSh-2791.2008.4.

REFERENCES

1. Efimov, A. V. (1994) *Structure*, **2**, 999-1002.
2. Efimov, A. V. (1994) *FEBS Lett.*, **355**, 213-219.
3. Efimov, A. V. (1997) *Proteins*, **28**, 241-260.
4. Efimov, A. V. (2004) *Uspekhi Biol. Khim.*, **44**, 109-132.
5. Levitt, M., and Chothia, C. (1976) *Nature*, **261**, 552-558.
6. Rao, S. T., and Rossmann, M. G. (1973) *J. Mol. Biol.*, **76**, 241-256.
7. Sternberg, M. J. E., and Thornton, J. M. (1976) *J. Mol. Biol.*, **105**, 367-382.
8. 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.
9. Castillo, R. M., Mizuguchi, K., Dhanaray, V., Albert, A., Blundell, T. L., and Murzin, A. G. (1999) *Structure*, **7**, 227-236.
10. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissing, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res.*, **28**, 235-242.
11. Tatusova, T. A., and Madden, T. L. (1999) *FEMS Microbiol. Lett.*, **174**, 247-250.
12. Efimov, A. V. (1995) *J. Mol. Biol.*, **245**, 402-415.
13. Lim, V. I., Mazanov, A. L., and Efimov, A. V. (1978) *Mol. Biol. (Moscow)*, **12**, 206-213.
14. Richardson, J. S. (1977) *Nature*, **268**, 495-500.
15. Efimov, A. V. (2008) *Biochemistry (Moscow)*, **73**, 23-28.