ISSN 0006-2979, Biochemistry (Moscow), 2008, Vol. 73, No. 1, pp. 23-28. © Pleiades Publishing, Ltd., 2008.
Published in Russian in Biokhimiya, 2008, Vol. 73, No. 1, pp. 29-35.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM07-187, December 16, 2007.

Structural Trees for Proteins Containing φ-Motifs

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 June 15, 2007

Abstract—In the present study, a novel structural motif of proteins referred to as the φ-motif is considered, and two novel structural trees in which the φ-motif is taken as the root structure have been constructed. 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. Construction of the structural trees and modeling of folding pathways have shown that all structures of the protein superfamilies can be obtained by stepwise addition of α-helices and/or β-strands to the root φ-motif taking into account a restricted set of rules inferred from known principles of protein structure. The structural trees are a good tool for structure comparison, structural classification of proteins, as well as for searching for all possible protein folds and folding pathways.

DOI: 10.1134/S0006297908010033

Key words: protein structure similarity, classification, modeling, protein folding

The structural tree for proteins is a scheme that includes all the intermediate and final three-dimensional structures that can be obtained by stepwise addition of secondary structural elements to the root (starting) structure. Possible folding pathways are shown by lines that connect all the structures between each other giving one structural tree. Secondary structural elements are added to the growing structures in accordance with a set of rules inferred from known principles of protein structure. The structural motif having a unique overall fold is taken as the root structure of the tree.

Structural trees are good and promising tools for solving several problems such as protein structure comparison, structural classification of proteins, mechanisms of protein folding, searching for all possible protein folds both known and unknown, etc.

Structural trees for seven large protein superfamilies—β-proteins containing abcd-units, 3β-corners, S-like β-sheets; two-layer (α+β)-proteins containing abCd-units; three-layer α/β-proteins containing five- and seven-segment α/β-motifs; and α-proteins containing α-α-corners—were constructed earlier [1-4]. In this study, two novel structural trees for proteins containing φ-motifs have been constructed and analyzed. Two variants of the φ-motif are taken as root structures of the trees.

METHODS OF INVESTIGATION

The φ-motif has been revealed as a result of visual inspection and protein structure comparison of the Protein Data Bank entries (http://www.rcsb.org/pdb/) and protein structures described in the original papers. In total, fifty proteins containing φ-motifs have been found (their PDB codes are 1BW4, 2CR5, 2CAE, 2J7N, 1N1O, 1NSO, 2RSP, 2MIP, 1AW8, 1FDO, 2NAP, 1G8J, 1EU1, 2IVF, 2ENG, 1OA9, 1NYC, 5P21, 1EGA, 1M7B, 1LAN, 1HMJ, 1NQT, 1B7G, 1GD1, 1S95, 1UTE, 2CZR, 2TMN, 1JAP, 1JAO, 1BUV, 1KAP, 1AGJ, 1A3A, 1QLD, 1ADM, 1BKN, 1EI1, 2GO3, 1KKH, 1FI4, 1HK7, 1QME, 2FD6, 1Q6U, 1FD9, 1FKJ, 1C9H, and 1E44).

The structural trees have been constructed taking into account the following general rules [1, 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;

– a variant of the φ-motif is taken as the root structure of the tree;

– larger protein and intermediate structures are obtained by stepwise addition of β-strands and/or α-helices to the growing structure (in some cases, "ready building blocks", e.g. β-β-hairpins or α-β-units 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 [5], so the next β-strand should be packed into a β-layer and the next α-helix into an α-layer of the growing structure;

– crossing of connections [6] and formation of knots [7] 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 handedness and overall folds;

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

RESULTS AND DISCUSSION

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. 1). There are two types of φ-motifs—the hairpin-strand type where a β-strand follows a β-hairpin (Fig. 1, a and c) and the strand-hairpin in which a β-hairpin follows a β-strand (Fig. 1, b and d). Figure 1e represents a split φ-motif, which has one additional β-strand packed in between the β-strand and the β-hairpin so that there are two β-hairpins in this structure.

A distinctive feature of the φ-motif is the loop, which connects two edge β-strands and crosses over the central β-strand or its extension so that together they form a cross-like structure. The loop provides a reverse turn of the polypeptide chain as well as its transition from one edge of the φ-motif to the other. This loop will be 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. 1, a, b, and e) and in the anticlockwise direction in the left-handed φ-motifs (Fig. 1, c and d). In the right-handed φ-motifs, the β-hairpins formed by strands *1* and *2* (Fig. 1, a and e) and strands *2* and *3* (Fig. 1b) as well as the split β-hairpins formed by strands *2* and *3* (Fig. 1, a and e) and strands *1* and *2* (Fig. 1b) are right-turned, but in the left-handed φ-motifs, the corresponding β-hairpins are left-turned (Fig. 1, c and d). It should be noted that strands *1* and *3* are packed parallel to each other in all the φ-motifs.

Figure 1f shows a schematic representation of the ψ-motif described previously [8, 9], which has both similarity to the φ-motif and essential differences. Both the motifs have the crossover loop that crosses over the central β-strand or its extension. When viewed from the crossover loop, the split β-hairpin formed by the two edge β-strands of the ψ-motif is right-turned as well as in the φ-motifs shown in Fig. 1 (a, b, and e). In this form, the ψ-motif occurs in the double-psi β-barrels [8, 9] and in proteins containing φ-motifs. In some other proteins, there are ψ-motifs having a left-turned split β-hairpin (unpublished data). The essential 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, while the φ-motif is formed by three adjacent β-strands connected by loops into one cooperative structure.

Figure 2 shows some examples of complex φ-motifs from known protein structures. As seen, they have the same overall fold despite the fact that their crossover loops can have different lengths and conformations, e.g. an α-helical conformation or both α-helical and irregular conformations. The β-hairpin of the φ-motif can be folded orthogonally upon itself so as to form the so-called β-β-corner [10] as observed in the double-psi β-barrels [9] (see also in Fig. 3 three branches in the left part of the tree).

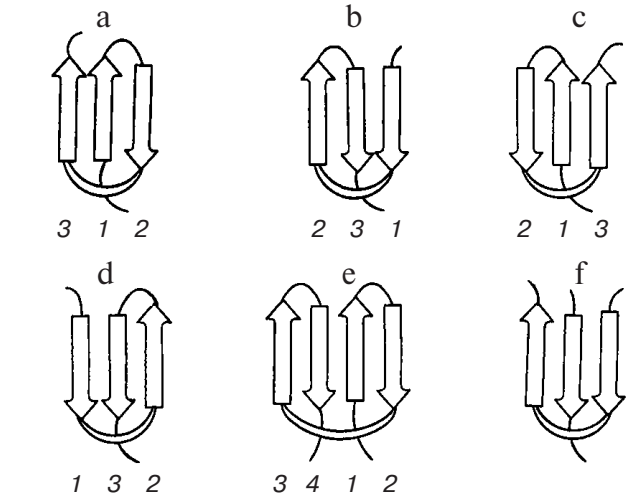

Fig. 1. Schematic representation of different variants of the φ-motif (a-e) and the ψ-motif (f). β-Stands are shown with arrows directed from the N- to the C-ends, crossover loops by double lines and other loops by single lines. *1-4* are numbers of β-strands in the chain starting from the N-end.

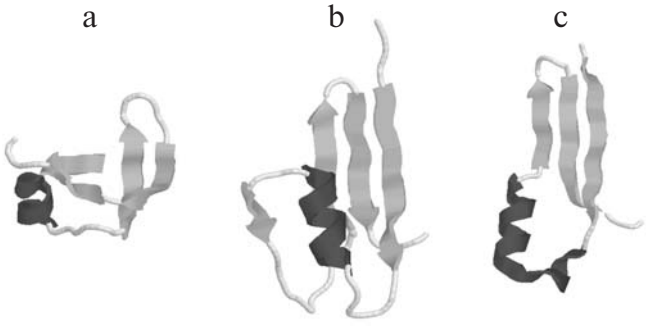

Fig. 2. Examples of complex φ-motifs in known proteins. a) A φ-motif found in aspartate decarboxylase (1AW8, region 26-65) in which the β-hairpin forms the β-β-corner; b) a φ-motif in which the long crossover loop consists of an α-helix and an irregular region (1EGA, 4-64); c) a φ-motif in which the crossover loop forms an α-helix (1NYC, 1-47). The figures are drawn with the program RasMol [11].

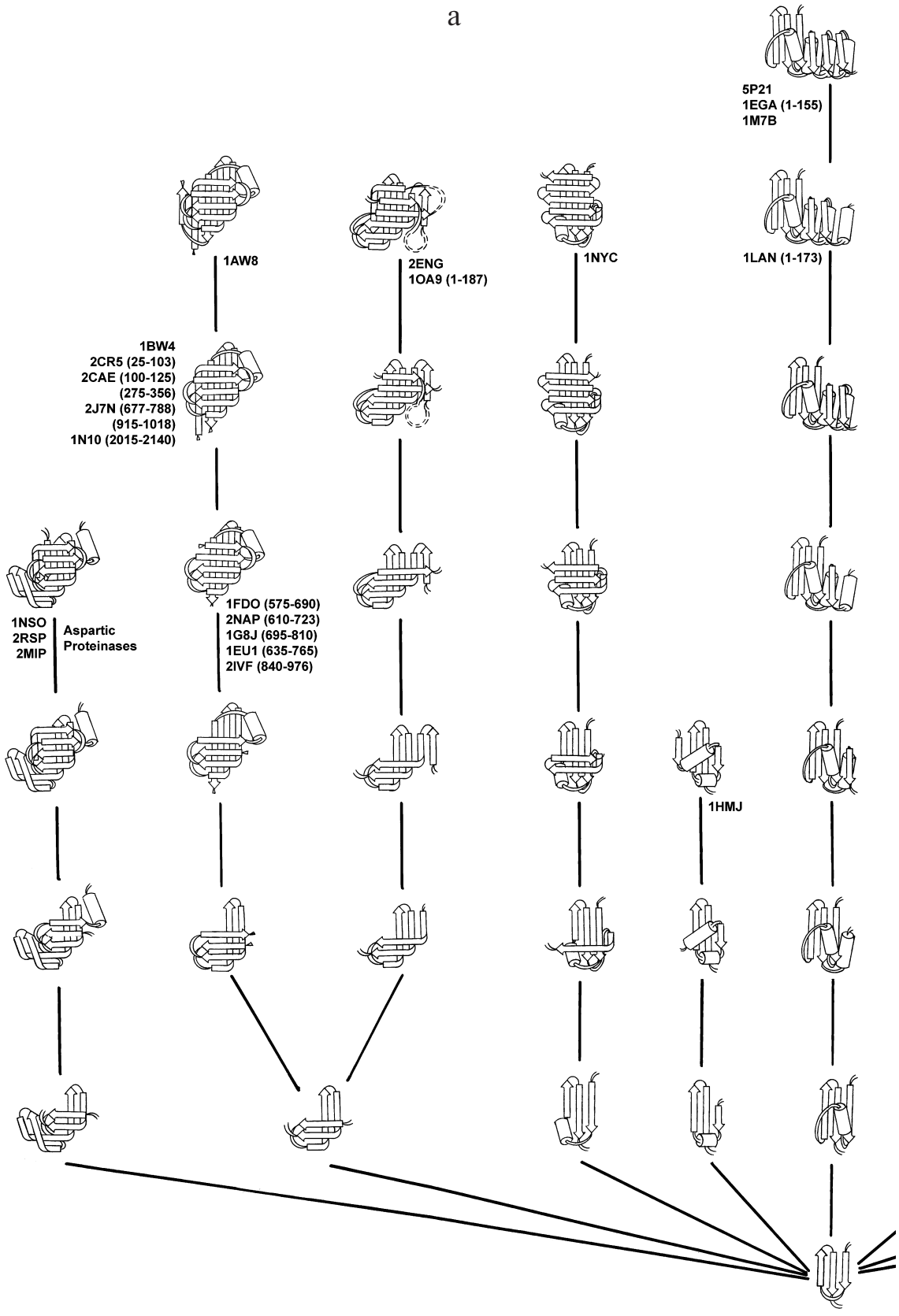

Fig. 3. a, b) Structural tree for proteins containing φ-motifs of the strand-hairpin type. Panels (a) and (b) are the left and right halves of the tree. β-Strands are shown with arrows directed from the N- to the C-ends, α-helices by cylinders and irregular regions and loops by single and double lines. c) Structural tree for proteins containing φ-motifs of the hairpin-strand type.

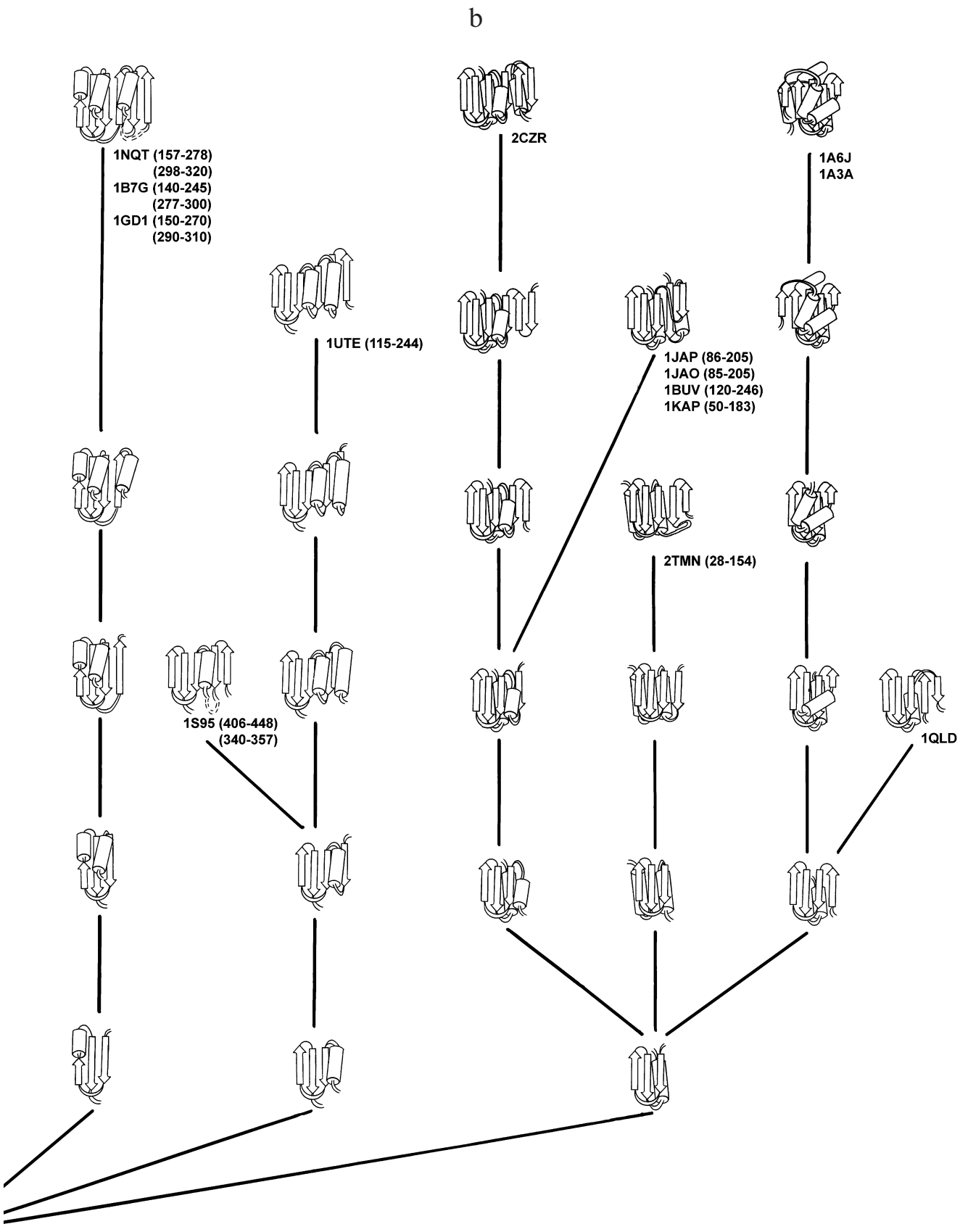
b
1NQT (157-278)
(298-320)
1B7G (140-245)
(277-300)
1GD1 (150-270)
(290-310)
1UTE (115-244)
1S95 (406-448)
(340-357)
2CZR
1JAP (86-205)
1JAO (85-205)
1BUV (120-246)
1KAP (50-183)
2TMN (28-154)
1A6J
1A3A
1QLD

Fig. 3. (Contd.)

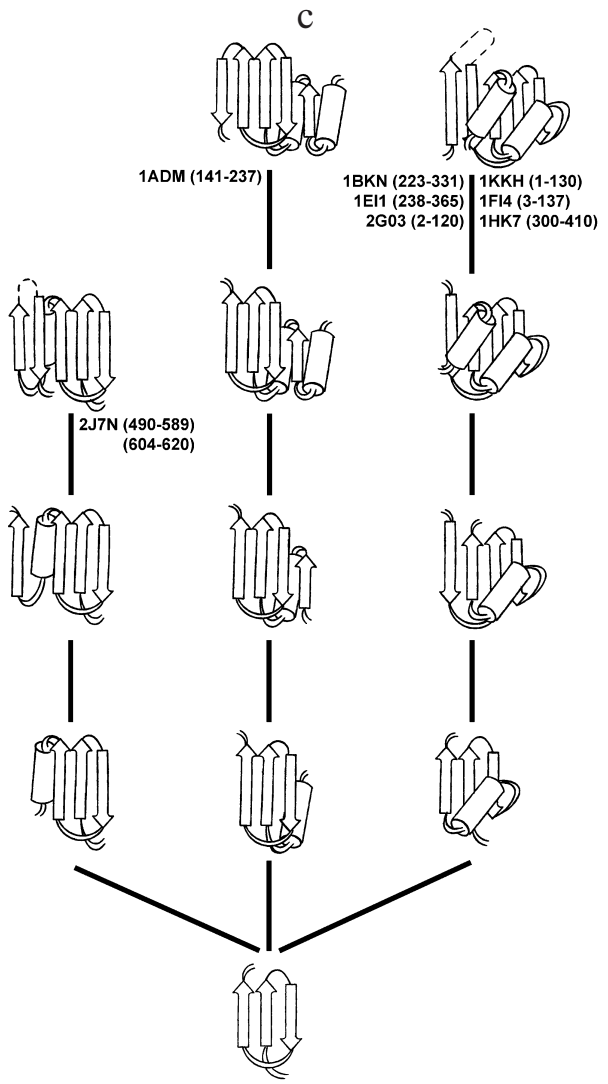

Fig. 3. (Contd.)

Analysis shows that different variants of the φ-motif have different frequencies of occurrence in proteins. Among 50 proteins containing φ-motifs (see the database in the preceding section), only one protein (PDB code 1QME) has two left-handed φ-motifs (regions 657-690 and 715-750). This means that in proteins φ-motifs occur predominantly in the right-handed form. Among the right-handed φ-motifs, the strand-hairpin type occurs in 36 proteins, the hairpin-strand type in eight proteins, and the split φ-motif in five proteins. The reasons for different frequencies of occurrence of the φ-motifs are still poorly understood and will be further investigated.

Figure 3 (a and b) represents a structural tree for proteins containing φ-motifs of the strand-hairpin type constructed in accordance with the rules. The root φ-motif of the tree is shown schematically beneath. More complex φ-motifs having additional secondary structural elements are shown in the bottom row of the tree. In the central branches of the tree, the φ-motifs have additional α-helices in loops. In the three branches on the left, the β-hairpins of the φ-motifs are folded upon themselves into β-β-corners. In aspartic proteinases, the crossover loops are so long that additional β-hairpins are formed in these regions (these β-hairpins can be very flexible as in 1NSO and 2RSP). It is interesting to note that in these three branches, addition of β-strands to the growing structures results in formation of ψ-motifs. Figure 3c shows a structural tree for proteins containing φ-motifs of the hairpin-strand type. This tree is very similar to that shown in Fig. 3 (a and b); however, it represents three novel pathways of growth of the φ-motif.

As can be seen, there are several rows or levels in the structural trees. The structures of a higher row are obtained by addition of a β-strand or an α-helix to the structures located in the preceding row below. As mentioned above, each structural tree has several branches. Within one branch, structures having a higher position in the tree include the structures located lower. Structures of different branches have a common fold located in the branching point. The higher a branching point is located in the tree, the higher the level of structural similarity between proteins and domains of the corresponding branches. Note that structural similarity of proteins does not mean that they have similar functions, so in one branch there can be proteins and domains having both similar and different functions.

Proteins and domains found within one structural tree can be grouped into one structural class or superfamily, proteins and domains containing φ-motifs. Proteins and domains found within branches of the tree can be considered as subclasses or subfamilies. This classification is based on similarity of overall folds and modeled folding pathways of proteins and domains. In this classification, amino acid sequences, functions, and homology of proteins are not taken into account, so it is different from classifications suggested by other authors [12-14] in which this information is used.

Thus, taking into account all the data presented above, it is possible to suppose that the φ-motif can fold into its unique structure itself independently of the remaining part of the molecule. The structural trees presented in Fig. 3 demonstrate that the other structural elements of protein molecules can be attached to the growing structures step by step in accordance with the general rules. All this together has led us to a hypothesis that the φ-motif can act as a nucleus in protein folding, and the pathways of its stepwise growth can be considered as possible folding pathways of the proteins. In any case, the structural trees are a good tool for theoretical modeling of protein folding and can be used in experimental studies of the problem.

I thank Dr. E. V. Brazhnikov and A. M. Kargatov for help in drawing the figures.

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

REFERENCES

1. Efimov, A. V. (1997) *Proteins*, **28**, 241-260.
2. Efimov, A. V. (1997) *FEBS Lett.*, **407**, 37-41.
3. Efimov, A. V. (1998) *FEBS Lett.*, **437**, 246-250.
4. Efimov, A. V. (2004) *Uspekhi Biol. Khim.*, **44**, 109-132.
5. Efimov, A. V. (1995) *J. Mol. Biol.*, **245**, 402-415.
6. Lim, V. I., Mazanov, A. L., and Efimov, A. V. (1978) *Mol. Biol.* (Moscow), **12**, 206-213.
7. Richardson, J. S. (1977) *Nature*, **268**, 495-500.
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., Dhanaraj, V., Albert, A., Blundell, T. L., and Murzin, A. G. (1999) *Structure*, **7**, 227-236.
10. Efimov, A. V. (1991) *FEBS Lett.*, **284**, 288-292.
11. Sayle, R., and Milner-White, J. (1995) *Trends Biochem. Sci.*, **20**, 374-376.
12. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) *J. Mol. Biol.*, **247**, 536-540.
13. Holm, L., and Sander, C. (1996) *Science*, **273**, 595-602.
14. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., and Thornton, J. M. (1997) *Structure*, **5**, 1093-1108.