FEBS 17329 FEBS Letters 391 (1996) 167–170

A structural tree for α-helical proteins containing α–α-corners and its application to protein classification

A.V. Efimov*

Institute of Protein Research, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia

Received 27 May 1996; revised version received 13 June 1996

Abstract A structural tree for α-helical proteins and domains including α–α-corners has been constructed. The α–α-corner is taken as a root structure of the tree. The larger protein structures are obtained by stepwise addition of α-helices to the root α–α-corner taking into account a restricted set of rules inferred from known principles of protein structure. The protein structures that can be obtained in this way are grouped into one structural class and those found in branches of the tree into subclasses.

Key words: Protein structure comparison; Stepwise folding; Structural motif; Structural similarity

1. Introduction

The structural motifs having unique overall folds and a unique handedness are of particular value in protein modelling and folding since they can be taken as starting structures in modelling and can be considered as nuclei in protein folding. Analysis shows that the larger protein structures can be obtained by a stepwise addition of α-helices and/or β-strands to the corresponding structural motifs taking into account simple rules. Several schemes of stepwise growth of the structural motifs have been constructed and published [1–5]. Each scheme shows possible pathways of growth of the corresponding motif, protein structures that can be obtained and different levels of structural similarity between them. Protein structures that can be obtained in accordance with a given scheme can be grouped into one structural class although their sequences may have no homology and their functions may differ.

This paper describes another representation of such a scheme called the structural tree. Proteins and domains containing the α–α-corner [2] are taken as an example. This structural tree includes more proteins and shows some novel pathways of growth of the α–α-corner as compared to the schemes published previously [2,4].

2. A structural tree for α-helical proteins and domains including α–α-corners

The α–α-corner is a structural motif formed by two α-helices adjacent along the polypeptide chain, packed approximately crosswise and connected by an interhelical loop. Variants of this motif were initially found in two protein families, 'E–F-hands' in the calcium-binding proteins [6] and 'helix–turn–helix' motifs in the DNA-binding proteins (for a review, see [7]). It was shown later that α–α-corners are widespread in both homologous and non-homologous proteins and occur practically always in one form in which the polypeptide chain nearly forms a turn of a left-handed superhelix in three dimensions [2]. The α–α-corner with a short connection has the $\alpha_m\gamma\alpha_L\beta\beta\alpha_n$-conformation of the polypeptide chain and the definite sequence pattern of the key hydrophobic, hydrophilic and glycine residues irrespective of whether it is found in homologous or non-homologous proteins [2]. Some small proteins and domains are merely composed of an α–α-corner and short irregular 'tails' [4]. All this taken together suggests that the α–α-corner represents a stable kind of fold which can adopt its unique structure per se.

The larger protein structures can be obtained by stepwise addition of α-helices and/or β-strands to the α–α-corner taking into account the following rules: (1) crossing of connections is prohibited [8]; (2) each α–α-corner of a growing structure should have its unique handedness; (3) an α-helix should be packed into the α-helical layer and a β-strand into the β-layer of a growing structure [5,9,10]; (4) the obtained structures should be compact.

In globular proteins, α-helices pack in one of three characteristic arrangements, aligned in parallel or antiparallel, orthogonal, or slanted (for details, see [10–13]). Taking into account these packing preferences of α-helices (rule 4) and restraints on the folding imposed by the structure of the loops (rules 1 and 2), one may conclude that there is a restricted set of structures that can be formed by two α-helices adjacent along the chain and connected by a short or medium-sized loop. These are α–α-hairpins, α–α-corners, L-shaped and V-shaped structures. Thus, addition of an α-helix to the α–α-corner can be done in different ways and results in formation of the structures shown in the bottom row of Fig. 1. Note that each structure in Fig. 1 can have both directions of the polypeptide chain but is drawn once to economize on space. Also for this reason, only the allowed structures observed in proteins are shown.

Thus, Fig. 1 represents a scheme of stepwise growth of an α–α-corner that can be called the structural tree for proteins including α–α-corners. There is an α–α-corner in the root of this structural tree. Arrows show different ways it grows. There are several levels (rows) in the structural tree. The bottom level (row) contains the structures obtained by an addition of one α-helix to the α–α-corner. The structures of the next level have two α-helices added, etc. In other words, each level contains structures composed of the same number of α-helices. As seen, the structural tree has several branches. The structures of the same branch have a higher level of structural similarity than those of different branches. All the structures

*Corresponding author. Tel./Fax: (7095) 924-0493.
E-mail: efimov@ipr.serpukhov.su

PII S0014-5793(96)00720-X

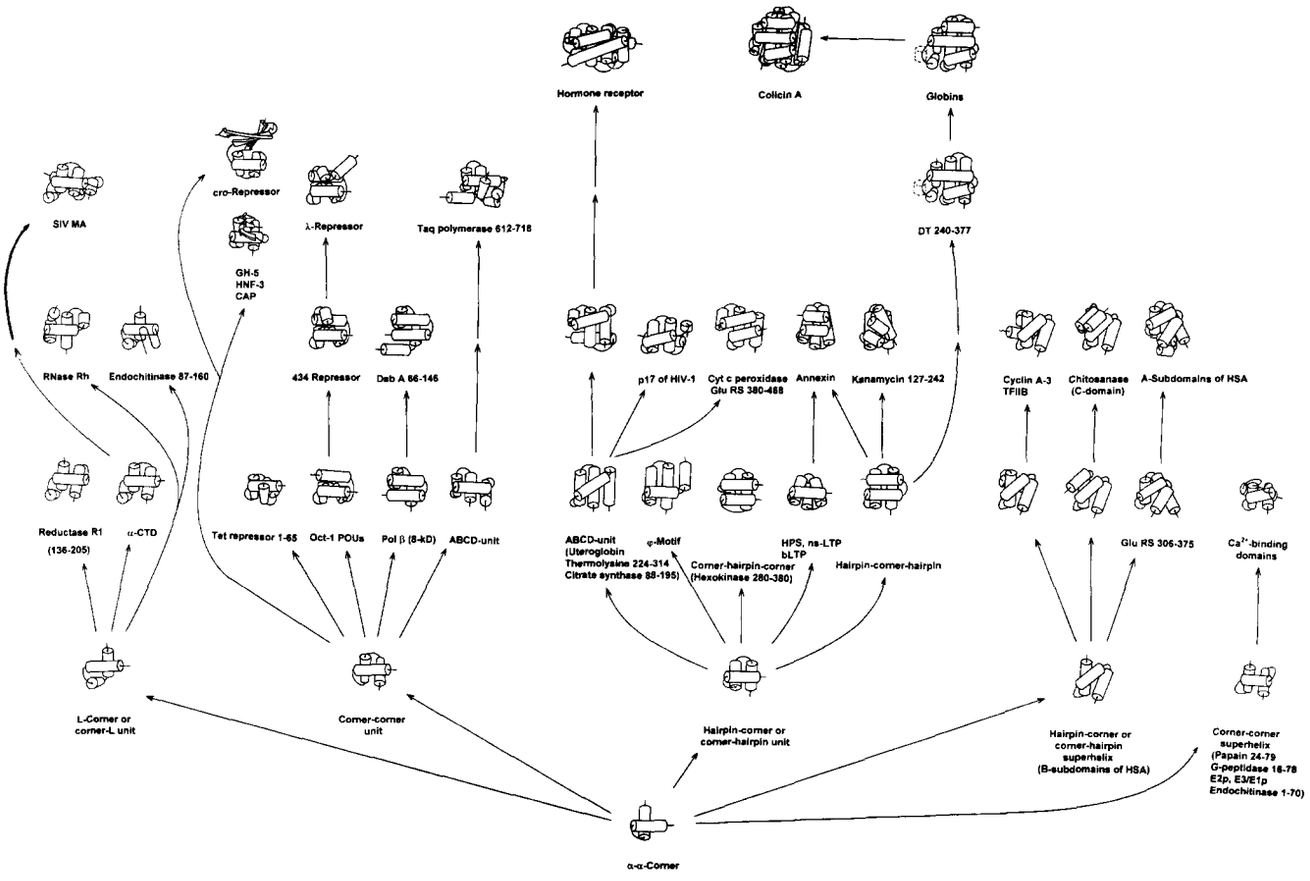

Fig. 1. A structural tree for proteins and domains including α–α-corners. Structural information is taken from the following papers: Endochitinase [14]; Hexokinase [15]; RNase Rh, ribonuclease Rh [17]; Cyt *c* peroxidase, cytochrome *c* peroxidase [18]; Reductase R1, ribonucleotide reductase protein R1 [20]; α-CTD, carboxyl-terminal domain of the RNA polymerase α-subunit [21]; SIV MA, simian immunodeficiency virus matrix antigen [22]; Tet repressor [23]; Oct-1 POUs, Oct-1 POU-specific domain [24]; Pol β (8 kDa), 8 kDa domain of rat DNA polymerase β [25]; 434 Repressor [26]; Dsb A, Dsb A protein [27]; GH-5, globular domain of histone H5 [28]; HNF-3, HNF-3/fork head DNA-recognition motif [29]; CAP, catabolite gene activator protein [30]; cro-Repressor [31]; λ-Repressor [32]; Taq polymerase [33]; Uteroglobin [34]; Thermolysin [35]; Citrate synthase [36]; HPS, hydrophobic protein from soybean [37]; ns-LTP, non-specific lipid-transfer protein [38]; bLTP, barley lipid-transfer protein [39]; p17 of HIV-1, p17 matrix protein of HIV-1 [40]; Glu RS, glutamyl-tRNA synthetase [41]; Annexin, human annexin V [42]; 'kanamycin', kanamycin nucleotidyltransferase [43]; DT, diphtheria toxin [44]; Globins [45]; Colicin A, pore-forming domain of colicin A [46]; Hormone receptor, ligand-binding domain of human retinoic acid receptor (RAR)-γ [47]; HSA, human serum albumin [48]; Papain [49]; G-peptidase, Zn^{2+}-containing D-alanyl-D-alanine peptidase [50]; E3/E1p, E3/E1p-binding domain of dihydrolipoamide acetyltransferase [51]; Cyclin A-3 [52]; TFIIB, transcription factor IIB [53]; Chitosanase [54]; Ca^{2+}-binding domains [6].

of different branches have a common structure located in the corresponding branching point. The higher a branching point is located in the tree, the higher the level of structural similarity between proteins of the corresponding branches observed.

It is possible to distinguish at least five subfamilies of these proteins. Each subfamily has its three-α-helical motif (located in the branching point) as the common fold. It is of interest that most proteins and domains containing the corner–corner unit have the function to bind DNA. Most proteins and domains within other subfamilies have different functions despite their structural similarity.

3. Some additional remarks on the commonly occurring folding units

As mentioned above, the α-helical structures shown in Fig. 1 can have both directions of the polypeptide chain. As an example, Fig. 2 represents the corner-L (Fig. 2a) and L-corner (Fig. 2b) units that have opposite directions of the polypeptide chains but the same overall fold. The corner-L unit occurs in endochitinase (120–160) [14], hexokinase (260–300) [15], transcription factor LEF-1 [16] and others. The L-corner unit is found, for example, in ribonuclease Rh (71–112) [17] and cytochrome *c* peroxidase (118–168) [18].

The structures shown in Fig. 2c,d have an overall fold that is reminiscent of the Greek letter φ. It is called the φ-motif. In one form of the φ-motif, pairs of helices AB, BC and CD form an L-shaped structure, an α–α-corner and an α–α-hairpin, respectively (Fig. 2c). In the other form (Fig. 2d), an α–α-corner is formed by helices A and B, an L-structure by helices B and C and an α–α-hairpin by helices C and D. There are 16 examples of different φ-motifs found in proteins of known structure [19].

Four-α-helical structures shown in Fig. 2f–i are variants of the same overall fold called the ABCD-unit. In this fold, helices B, C and D form a left-handed superhelix BCD and helix A is located in between helices B and D. For comparison, Fig. 2e represents the abcd-unit (a commonly occurring folding unit in β-proteins [1]) that has a right-handed super-

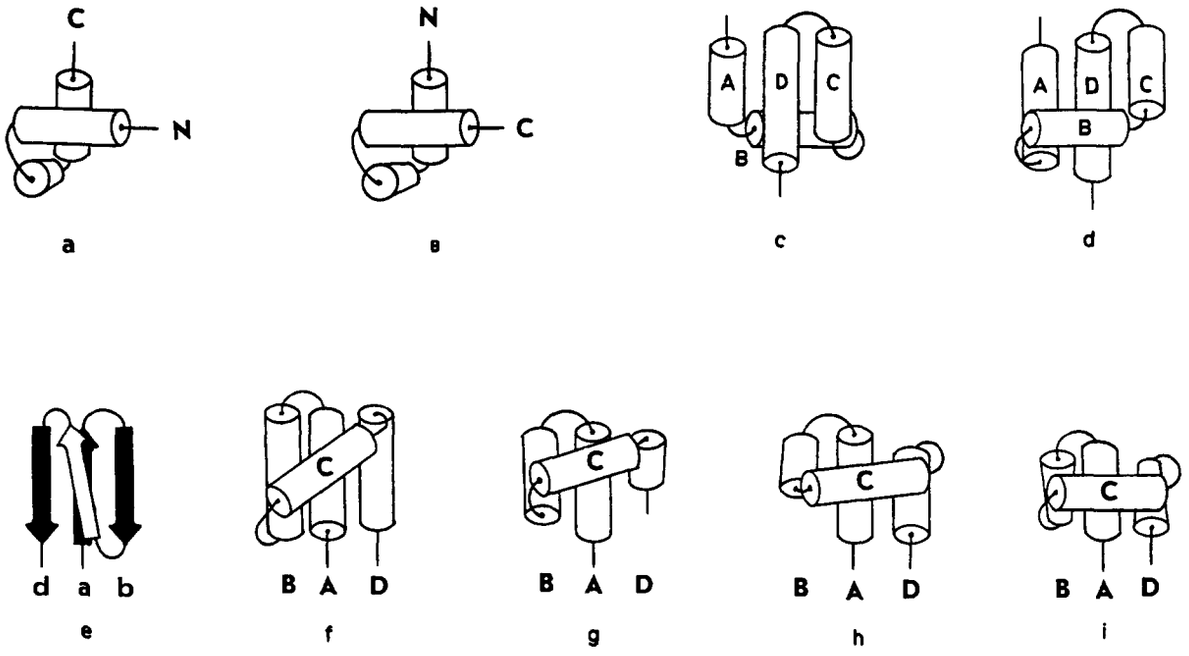

Fig. 2. A schematic representation of variants of some structural motifs. See the text for details.

helix bcd and strand a in between strands b and d. As seen, the overall folds of the abcd- and ABCD-units are rather similar but 'mirror-symmetrical' if segment conformations are ignored. A detailed analysis of these and other variants of the ABCD-unit including 22 examples from known proteins will be described elsewhere [19].

4. Discussion

The information obtained by analysis of the structural trees constructed by comparison of protein structures and possible folding pathways (Fig. 1) is of particular value in understanding the principles that govern the polypeptide chain folding (see also [1–5,55]). It can also be used for structural classification of proteins. Proteins and domains whose structures can be obtained by a stepwise addition of α-helices and/or β-strands to the same root motif can be grouped into one structural class. Proteins and domains found within branches of a structural tree can be considered as subclasses or subfamilies. Levels of structural similarity between different proteins can easily be observed by visual inspection. Within one branch, protein structures having a higher position in the tree include the structures located lower. Proteins and domains of different branches have the structure located in the branching point as the common fold, etc. It should be noted that two similar structures when superimposed can have a rather large value of the root-mean-square deviation since their α-helices and/or β-strands may be of different lengths and their connection regions may differ in length and conformation. Nevertheless, these structures can have very similar overall folds. From this point of view, this classification is rather different from those suggested by other authors [56–59].

Acknowledgements: This work was supported in part by Russian Foundation for Basic Research (RFFI Grant 95-04-11851a) and International Science Foundation (ISF Grants NKK000 and NKK300).

References

[1] Efimov, A.V. (1982) Mol. Biol. (in Russian) 16, 799–806.
[2] Efimov, A.V. (1984) FEBS Lett. 166, 33–38.
[3] Efimov, A.V. (1992) FEBS Lett. 298, 261–265.
[4] Efimov, A.V. (1994) FEBS Lett. 355, 213–219.
[5] Efimov, A.V. (1995) J. Mol. Biol. 245, 402–415.
[6] Kretsinger, R.H. and Nockolds, C.E. (1973) J. Biol. Chem. 248, 3313–3326.
[7] Pabo, C.O. and Sauer, R.T. (1984) Annu. Rev. Biochem. 53, 293–321.
[8] Lim, V.I., Mazanov, A.L. and Efimov, A.V. (1978) Mol. Biol. (in Russian) 12, 206–213.
[9] Levitt, M. and Chothia, C. (1976) Nature 261, 552–558.
[10] Efimov, A.V. (1977) Dokl. Akad. Nauk SSSR 235, 699–702.
[11] Crick, F.H.C. (1953) Acta Crystallogr. 6, 689–697.
[12] Chothia, C., Levitt, M. and Richardson, D. (1981) J. Mol. Biol. 145, 215–250.
[13] Richmond, T.J. and Richards, F.M. (1978) J. Mol. Biol. 119, 537–555.
[14] Hart, P.J., Pfluger, H.D., Monzingo, A.F., Hollis, T. and Robertus, J.D. (1995) J. Mol. Biol. 248, 402–413.
[15] Bennet, W.S., Jr. and Steitz, T.A. (1978) Proc. Natl. Acad. Sci. USA 75, 4848–4852.
[16] Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R. and Wright, P.E. (1995) Nature 376, 791–795.
[17] Kurihara, H., Nonaka, T., Mitsui, Y., Ohgi, K., Irie, M. and Nakamura, K.T. (1996) J. Mol. Biol. 255, 310–320.
[18] Su, X.-D., Yonetani, T. and Skoglund, U. (1994) FEBS Lett. 351, 437–442.
[19] Efimov, A.V. (1996) Mol. Biol. (in Russian), in press.
[20] Uhlin, U. and Eklund, H. (1994) Nature 370, 533–539.
[21] Jeon, Y.H., Negishi, T., Shirakawa, M., Yamazaki, T., Fujita, N., Ishihama, A. and Kyogoku, Y. (1995) Science 270, 1495–1497.
[22] Rao, Z., Belyaev, A.S., Fry, E., Roy, P., Jones, I.M. and Stuart, D.I. (1995) Nature 378, 743–747.
[23] Kisker, C., Hinrichs, W., Tovar, K., Hillen, W. and Saenger, W. (1995) J. Mol. Biol. 247, 260–280.

[24] Klemm, J.D., Rould, M.A., Aurora,R., Herr, W. and Pabo, C.O. (1994) Cell 77, 21–32.
[25] Pelletier, H., Sawaya, M.R., Kumar, A., Wilson, S.H. and Kraut, J. (1994) Science 264, 1891–1898.
[26] Mondragon, A., Subbiah, S., Almo, S.C., Drottar, M. and Harrison, S.C. (1989) J. Mol. Biol. 205. 189–200.
[27] Martin, J.L., Bardwell, J.C.A. and Kuriyan, J. (1993) Nature 365, 464– 468.
[28] Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L. and Sweet, R.M. (1993) Nature 362, 219–223.
[29] Clark, K.L., Halay, E.D., Lai, E. and Burley, S.K. (1993) Nature 364, 412–420.
[30] Weber, I.T. and Steitz, T.A. (1987) J. Mol. Biol. 198, 311–326.
[31] Anderson, W.F., Ohlendorf, D.H., Takeda, Y. and Matthews, B.W. (1981) Nature 290, 754–758.
[32] Pabo, C.O. and Lewis, M. (1982) Nature 298, 443–447.
[33] Kim, Y., Eom, S.H., Wang, J., Lee, D.-S., Suh, S.W. and Steitz, T.A. (1995) Nature 376, 612–616.
[34] Morize, I., Surcouf, E., Vaney, M.C., Epelboin, Y., Buehner, M., Fridlansky, F., Milgrom, E. and Mornon, J.P. (1987) J. Mol. Biol. 194, 725–739.
[35] Holmes, M.A. and Matthews, B.W. (1982) J. Mol. Biol. 160, 623–639.
[36] Remington, S., Wiegand, G. and Huber, R. (1982) J. Mol. Biol. 158, 111–152.
[37] Baud, F., Pebay-Peyroula, E., Cohen-Addad, C., Odani, S. and Lehmann, M.S. (1993) J. Mol. Biol. 231, 877–887.
[38] Shin, D.H., Lee, J.Y., Hwang, K.Y., Kim, K.K. and Suh, S.W. (1995) Structure 3, 189–199.
[39] Heinemann, B., Andersen, K.V., Nielsen, P.R., Bech, L.M. and Poulsen, F.M. (1996) Protein Sci. 5, 13–23.
[40] Matthews, S., Barlow, P., Boyd, J., Barton,G., Russell, R., Mills, H., Cunningham, M., Meyers, N., Burns, N., Clark, N., Kingsman, S., Kingsman, A. and Campbell, I. (1994) Nature 370, 666–668.
[41] Nureki, O., Vassylyev, D.G., Katayanagi, K., Shimizu,T., Sekine, S.-I., Kigawa, T., Miyazawa, T., Yokoyama, S. and Morikawa, K. (1995) Science 267, 1958–1965.
[42] Huber, R., Romisch, J. and Paques, E.P. (1990) EMBO J. 9, 3867–3874.
[43] Pedersen, L.C., Benning, M.M. and Holden, H.M. (1995) Biochemistry 34, 13305–13311.
[44] Choe, S., Bennet, M.J., Fujii, G., Curmi, P.M.G., Kantardjieff, K.A., Collier, R.J. and Eisenberg, D. (1992) Nature 357, 216–222.
[45] Perutz, M.F., Kendrew, J.C. and Watson, H.C. (1965) J. Mol. Biol. 13, 669–678.
[46] Parker, M.W., Postma, J.P.M., Pattus, F., Tucker, A.D. and Tsernoglou, D. (1992) J. Mol. Biol. 224, 639–657.
[47] Renaud, J.-P., Rochel, N., Ruff, M., Vivat, V., Chambon, P., Gronemeyer, H. and Moras, D. (1995) Nature 378, 681–689.
[48] He, X.M. and Carter, D.C. (1992) Nature 358, 209–215.
[49] Drenth, J., Jansonius, J.N., Koekoek, R. and Wolthers, B.G. (1971) Adv. Prot. Chem. 25, 79–116.
[50] Dideberg, O., Charlier, P., Dive, G., Joris, B., Frere, J.M. and Ghuysen, J.M. (1982) Nature 299, 469–470.
[51] Kalia, Y.N., Brocklehurst, S.M., Hipps, D.S., Appella, E., Sakaguchi, K. and Perham, R.N. (1993) J. Mol. Biol. 230, 323–341.
[52] Brown, N.R., Noble, M.E.M., Endicott, J.A., Garman, E.F., Wakatsuki, S., Mitchell, E., Rasmussen, B., Hunt, T. and Johnson, L.N. (1995) Structure 3, 1235–1247.
[53] Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G. and Burley, S.K. (1995) Nature 377, 119–128.
[54] Marcotte, E.M., Monzingo, A.F., Ernst, S.R., Brzezinski, R. and Robertus, J.D. (1996) Nature Struct. Biol. 3, 155– 161
[55] Ptitsyn, O.B. and Finkelstein, A.V. (1980) Q. Rev. Biophys. 13, 339–386.
[56] Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) J. Mol. Biol. 247, 536–540.
[57] Holm,L., Ouzounis, C., Sander, C., Tuparev, G. and Vriend, G. (1992) Protein Sci. 1, 1691–1698.
[58] Orengo, C.A., Flores, T.P., Taylor, W.R. and Thornton, J.M. (1993) Protein Eng. 6, 485–500.
[59] Rufino, S.D. and Blundell, T.L. (1994) J. Comput. Aided Mol. Design 8, 5–27.