Volume 224, number 2, 372–376 November 1987

Pseudo-homology of protein standard structures formed by two consecutive β-strands

A.V. Efimov

Institute of Protein Research, Academy of Sciences of the USSR, 142292 Pushchino, Moscow Region, USSR

Received 17 September 1987

Protein standard structures formed by two consecutive β-strands connected by short loops are considered in this paper. A stereochemical analysis of each standard structure has been performed to determine the necessary conditions which must be fulfilled in the amino acid sequence encoding the standard structure. It is shown that amino acid sequences coding for the same standard structure in different proteins have practically the same order of hydrophobic, hydrophilic and glycine residues. The results of the stereochemical analysis are confirmed by a large number of examples from known protein structures.

β-Strand; β-Turn; β-β-Hairpin; Standard structure; Amino acid sequence

1. INTRODUCTION

Recently, a number of protein standard structures such as α-α-corners [1], α-β-hairpins [2] and β-α-hairpins [3] and a number of standard structures in irregular regions of proteins [4–6] have been revealed and described. Standard structures of the same type have a very similar arrangement of α-helices and/or β-strands and the fold of the polypeptide chain irrespective of whether they occur in homologous proteins. Two superimposed standard structures of the same type from different proteins practically coincide and the residues occupying equivalent positions have very similar conformations (signifying that angles ϕ and ψ of every pair of structurally equivalent residues fit the same region on a Ramachandran plot).

Here, standard β-β-hairpins and β-β-arcs (the β-β-arc is formed by two connected β-strands localized in different layers of a protein sandwich structure) and the features of amino acid sequences coding for them are considered.

Correspondence address: A.V. Efimov, Institute of Protein Research, Academy of Sciences of the USSR, 142292 Pushchino, Moscow Region, USSR

2. STEREOCHEMISTRY OF STRUCTURES FORMED BY TWO CONSECUTIVE β-STRANDS

Structures formed by two consecutive β-strands connected by a loop can be of two classes. Structures of the first class are those in which polypeptide chains fold into β-β-hairpins. There are some families of β-β-hairpins which differ in length and conformation of the loops [5]. β-β-Hairpins can be right- or left-handed depending on whether the second β-strand runs on the right or left, relative to the first when viewed from the hydrophobic core. In the second class, the β-strands lie in different layers of a protein sandwich structure and do not form hydrogen bonds. They look like arcs and will be described here as β-β-arcs. In this paper, only standard structures having short loops will be considered.

For the description of a polypeptide chain conformation, the method suggested in [6] will be used. A residue conformation will be denoted as α, α_L, β, β_P or γ. This means that a residue, denoted by a symbol, has ϕ,ψ angles, corresponding to a right-handed helix, left-handed helix, β-structure, polyproline helix regions or a region with ϕ =

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

$-90 \pm 30°$, $\psi = 0 \pm 30°$ on a Ramachandran plot, respectively.

A schematic representation of a standard β-β-hairpin with the five-residue $\beta\alpha\gamma\alpha_L\beta$-turn is shown in fig.1a. Side chains pointing to the hydrophobic core are designated by solid circles and must be hydrophobic in accordance with the theory of secondary structure [7]. Some of these side chains can be hydrophilic if they are situated at the edge of a β-sheet or at the ends of β-strands and are accessible to water molecules. The first residue of a $\beta\alpha\gamma\alpha_L\beta$-turn must not be very hydrophobic and the fourth (α_L-) must be glycine or a hydrophilic residue with a flexible side chain (for details see [6]). The remaining residues of this β-β-hairpin can be hydrophilic or hydrophobic. In fig.1a a right-handed β-β-hairpin is represented (its upper side is hydrophobic). The left-handed β-β-hairpin can have practically the same conformation of the polypeptide chain but hydrophobic side chains must be situated on the opposite (bottom) side.

Two β-strands can be connected by a six-rsidue $\beta\alpha\alpha\gamma\alpha_L\beta$-turn to form a β-β-hairpin having the $\beta_m\beta\alpha\alpha\gamma\alpha_L\beta\beta_n$-conformation, where m and n denote the numbers of residues in the first and second β-strands, respectively. The first residue of the $\beta\alpha\alpha\gamma\alpha_L\beta$-turn must not be very hydrophobic and the fifth (α_L-) must be glycine [6]. Hydrophobic side chains of the β-β-hairpin must be arranged so as to form the hydrophobic surface. These β-β-hairpins can also be right- and left-handed.

There are more complicated β-β-hairpins when one of the β-strands has a β-bulge [8]. An example of such a β-β-hairpin is shown in fig.1b. In this example there is a β-bulge having a standard $\beta\alpha\beta\beta$-conformation and the β-strands are connected by the $\beta\alpha\gamma\alpha_L\beta$-turn. As can be observed, the side chains of the two middle residues of the β-bulge ($\alpha\beta$ residues) point to the same side and must be hydrophobic if the hydrophobic core is from above. Consequently, amino acid sequences, encoding β-β-hairpins with and without a β-bulge, must differ in the order of hydrophobic residues (cf. fig.1a,b). In some cases β-bulges can have a $\beta\beta_P\beta_P\beta$-conformation. As a rule, one of the β_P-residues here is proline.

In fig.1c a schematic representation of a β-β-arc having a $\beta_m\beta\beta_P\beta_P\alpha_L\beta\beta_n$-conformation is shown. The side chain of the first residue of the $\beta\beta_P\beta_P\alpha_L\beta$-loop (it is also the last residue of the first β-strand) is hidden in the hydrophobic core and must be hydrophobic. The side chain of the fifth loop residue (also the first residue of the second β-strand) is partially hidden and can be hydrophobic or hydrophilic. The fourth (α_L-) residue of the loop must be glycine. Prolines in the second or third (β_P-) positions facilitate the formation of this β-β-arc, but their presence is not necessary. Side chains of the β-strands hidden in the hydrophobic core must be hydrophobic (shown by solid circles).

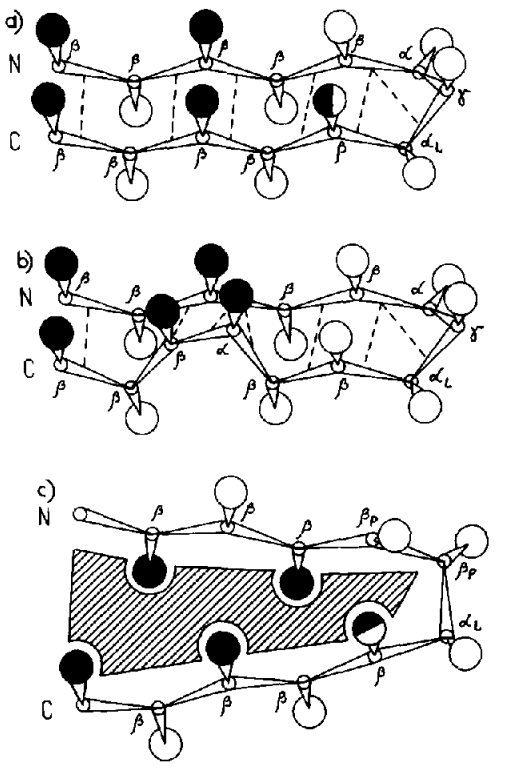

Fig.1. A schematic representation of some standard structures formed by two consecutive β-strands. (a) A β-β-hairpin with the $\beta\alpha\gamma\alpha_L\beta$-turn; (b) a β-β-hairpin having the $\beta\alpha\gamma\alpha_L\beta$-turn and the $\beta\alpha\beta\beta$-bulge; (c) a β-β-arc with the $\beta\beta_P\beta_P\alpha_L\beta$-loop (the hatched area denotes the hydrophobic core). C_α-atoms and side chains of residues are shown by small and large circles, respectively. See also the text.

3. PSEUDO-HOMOLOGY OF AMINO ACID SEQUENCES CODING FOR STANDARD β-β-HAIRPINS AND β-β-ARCS

An alignment of amino acid sequences coding

for β-β-hairpins with $\beta\alpha\gamma\alpha_L\beta$-turns is shown in fig.2. Each column contains structurally similar residues and is headed by a symbol showing the conformation of the residues in it. Hydrophobically invariant residues (they are hydrophobic in most of the proteins) are encircled. Indeed, some residues in the columns containing hydrophobic in-

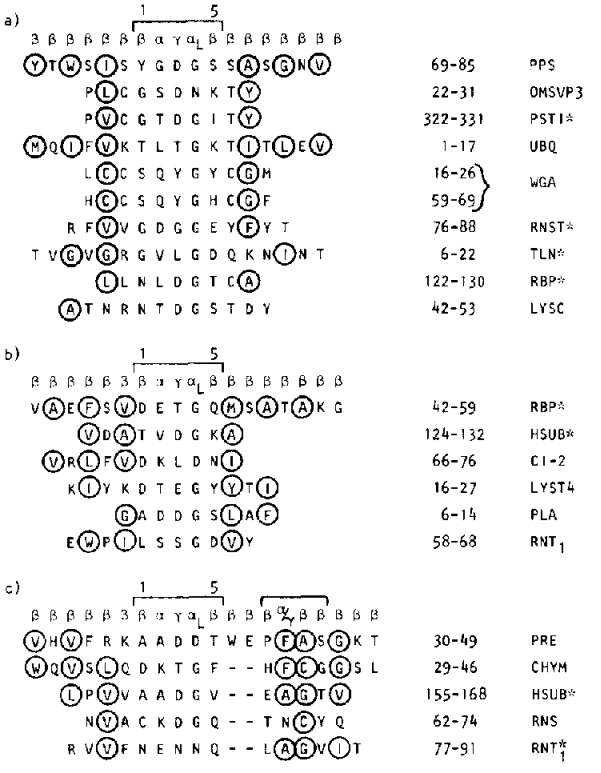

Fig.2. Alignment of the amino acid sequences coding for right-handed (a) and left-handed (b) β-β-hairpins with the $\beta\alpha\gamma\alpha_L\beta$-turns and β-β-hairpins having the $\beta\alpha\gamma\alpha_L\beta$-turns and the $\beta\alpha\beta\beta$-bulges (c). Regions of the sequences and the proteins containing them are listed on the right: PPS, penicillopepsin [9]; OMSVP3, ovomucoid [10]; PSTI, pancreatic secretory trypsin inhibitor [11]; UBQ, ubiquitin [12]; WGA, wheat germ agglutinin [13]; RNST, ribonuclease St [14]; TLN, thermolysin [15]; RBP, retinol-binding protein [16]; LYSC, chicken lysozyme [17]; HSUB, H-subunit of the photosynthetic reaction center [18]; CI-2, chymotrypsin inhibitor 2 [19]; LYST4, phage T_4 lysozyme [20]; PLA, plastocyanin [21]; RNT_1, ribonuclease T_1 [22]; PRE, prealbumin [23]; CHYM, chymotrypsin [24]; RNS, ribonuclease S [25]. Proteins indicated by asterisks are those for which conformations of residues are established taking into account C_α-atom stereo plots only.

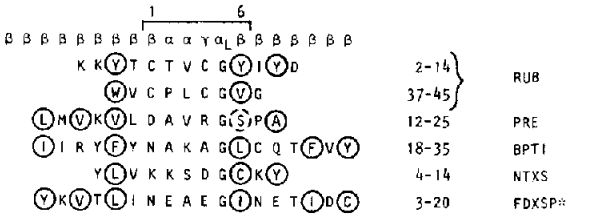

Fig.3. Alignment of the amino acid sequences coding for right-handed β-β-hairpins with the $\beta\alpha\alpha\gamma\alpha_L\beta$-turns. RUB, rubredoxin [26]; BPTI, bovine pancreatic trypsin inhibitor [27]; NTXS, scorpion neurotoxin [28]; FDXSP, *S. platensis* ferredoxin [29]. Other designations as in fig.2.

variants are hydrophilic, but in the corresponding proteins these residues are accessible to water molecules. As can be seen, most of the first (β-) residues of the $\beta\alpha\gamma\alpha_L\beta$-turns are hydrophilic or small and the α_L-positions are occupied by glycines (~80%) or hydrophilic residues (Asp, Asn). Hydrophobically invariant residues occupy those positions which must be hydrophobic in accordance with the results of the stereochemical analysis (cf. figs 1a and 2a, figs 1b and 2c). Superpositions of the hydrophobic invariants in the right-handed (fig.2a) and left-handed (fig.2b) $\beta_m\beta\alpha\gamma\alpha_L\beta\beta_n$-hairpins do not coincide, as should have taken place according to the definition, since they have hydrophobic surfaces on the opposite sides.

Some examples of right-handed β-β-hairpins

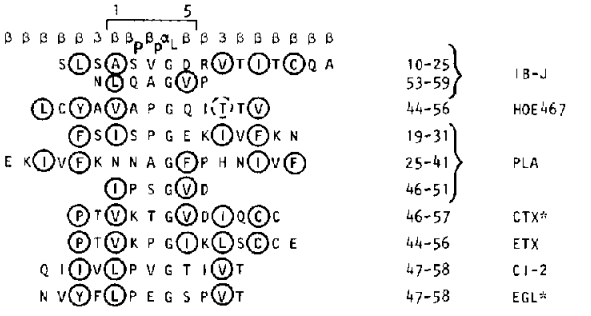

Fig.4. Alignment of the amino acid sequences coding for β-β-arcs with the $\beta\beta_P\beta_P\alpha_L\beta$-loops. IB-J, immunoglobulin (Bence-Jones protein) [30]; HOE467A, α-amylase inhibitor [31]; CTX, cobra toxin [32]; ETX, erabutoxin [33]; EGL, eglin [34]. Other designations as in figs 2 and 3.

with the six-residue $\beta\alpha\alpha\gamma\alpha_L\beta$-turns are shown in fig.3. Here, the first (β-) residues of the turns are also hydrophilic or small and the α_L-residues are glycines. A superposition of the hydrophobic invariants clearly shows which positions of these β-β-hairpins must be hydrophobic.

An alignment of the amino acid sequences coding for β-β-arcs with the shortest loops is represented in fig.4. Practically all the first residues of the $\beta\beta_P\beta_P\alpha_L\beta$-loops are hydrophobic and the α_L-residues are glycines. In many cases one of the β_P-positions (second or third) is occupied by proline. A superposition of the hydrophobic invariants is in good agreement with the positions to be hydrophobic according to the results of the stereochemical analysis (fig.1c).

Thus, each of the standard structures considered has a strictly definite order of hydrophobic, hydrophilic and glycine residues, irrespective of whether or not the structures are taken from homologous proteins. This similarity in amino acid sequences, coding for the same type of structures from non-homologous proteins, is not homology in the usual sense, but is pseudo-homology. Pseudo-homology of standard structures is widespread in proteins (see also [1–3]) and results from fulfillment of the main stereochemical requirements to be observed in protein molecules.

4. CONCLUSION

As mentioned above, β-β-hairpins and β-β-arcs can differ in length and conformation of the loops. Moreover, the greater the length of a loop, the more possible conformations it can have. This is why short loops play a particular role in proteins, since they have a limited number of stereochemically allowed conformations, and in the extreme case of the shortest loop there is only one conformation. This means that the shortest loops can connect α-helices and (or) β-strands to form a unique structure which can be a nucleus of protein folding.

Longer loops can be represented as combinations of small standard structures observed in irregular regions of proteins [6]. Most of these small structures have definite 'clue' positions which must be occupied by hydrophilic, flexible or glycine residues and provides us with an opportunity to understand the relation between the structures and amino acid sequences of long loops.

REFERENCES

[1] Efimov, A.V. (1984) FEBS Lett. 166, 33–38.
[2] Efimov, A.V. (1986) Mol. Biol. (USSR) 20, 329–339.
[3] Efimov, A.V. (1986) Mol. Biol. (USSR) 20, 340–345.
[4] Venkatachalam, C.M. (1968) Biopolymers 6, 1425–1436.
[5] Sibanda, B.L. and Thornton, J.M. (1985) Nature 316, 170–174.
[6] Efimov, A.V. (1986) Mol. Biol. (USSR) 20, 250–260.
[7] Lim, V.I. (1974) J. Mol. Biol. 88, 857–872.
[8] Richardson, J.S., Getzoff, E.D. and Richardson, D.C. (1978) Proc. Natl. Acad. Sci. USA 75, 2574–2578.
[9] James, M.N.G. and Sielecki, A.R. (1983) J. Mol. Biol. 163, 299–361.
[10] Bode, W., Epp, O., Huber, R., Laskowski, M. jr and Ardelt, W. (1985) Eur. J. Biochem. 147, 387–395.
[11] Bolognesi, M., Gatti, G., Menegatti, E., Guarneri, M., Marquart, M., Papamokos, E. and Huber, R. (1982) J. Mol. Biol. 162, 839–868.
[12] Vijay-Kumar, S., Bugg, C.E. and Cook, W.J. (1987) J. Mol. Biol. 194, 531–544.
[13] Wright, C.S. (1987) J. Mol. Biol. 194, 501–529.
[14] Nakamura, K.T., Iwahashi, K., Yamamoto, Y., Iitaka, Y., Yoshida, N. and Mitsui, Y. (1982) Nature 299, 564–566.
[15] Holmes, M.A. and Matthews, B.W. (1982) J. Mol. Biol. 160, 623–639.
[16] Newcomer, M.E., Jones, T.A., Åqvist, J., Sundelin, J., Eriksson, U., Rask, L. and Peterson, P.A. (1984) EMBO J. 3, 1451–1454.
[17] Blake, C.C.F., Koenig, D.F., Mair, G.A., North, A.C.T., Phillips, D.C. and Sarma, V.R. (1965) Nature 206, 757–761.
[18] Deisenhofer, J., Epp, O., Miki, K., Huber, R. and Michel, H. (1985) Nature 318, 618–624.
[19] McPhalen, C.A. and James, M.N.G. (1987) Biochemistry 26, 261–269.
[20] Remington, S.J., Anderson, W.F., Owen, J., Ten Eyck, L.F., Grainger, C.T. and Matthews, B.W. (1978) J. Mol. Biol. 118, 81–98.
[21] Guss, J.M. and Freeman, H.C. (1983) J. Mol. Biol. 169, 521–563.
[22] Heinemann, U. and Saenger, W. (1982) Nature 299, 27–31.

[23] Blake, C.C.F., Geisow, M.J., Oatley, S.J., Rérat, B. and Rérat, C. (1978) J. Mol. Biol. 121, 339–356.
[24] Cohen, G.H., Silverton, E.W. and Davies, D.R. (1981) J. Mol. Biol. 148, 449–479.
[25] Wyckoff, H.W., Tsernoglou, D., Hanson, A.W., Knox, J.R., Lee, B. and Richards, F.M. (1970) J. Biol. Chem. 245, 305–328.
[26] Watenpaugh, K.D., Sieker, L.C. and Jensen, L.H. (1979) J. Mol. Biol. 131, 509–522.
[27] Deisenhofer, J. and Steigemann, W. (1975) Acta Crystallogr. B31, 238–250.
[28] Almassy, R.J., Fontecilla-Camps, J.C., Suddath, F.L. and Bugg, C.E. (1983) J. Mol. Biol. 170, 497–527.
[29] Fukuyama, K., Hase, T., Matsumoto, S., Tsukihara, T., Katsube, Y., Tanaka, N., Kakudo, M., Wada, K. and Matsubara, H. (1980) Nature 186, 522–524.
[30] Epp, O., Lattman, E.E., Schiffer, M., Huber, R. and Palm, W. (1975) Biochemistry 14, 4943–4952.
[31] Pflugrath, J.W., Wiegand, G., Huber, R. and Vértesy, L. (1986) J. Mol. Biol. 189, 383–386.
[32] Walkinshaw, M.D., Saenger, W. and Maelicke, A. (1980) Proc. Natl. Acad. Sci. USA 77, 2400–2404.
[33] Bourne, P.E., Sato, A., Corfield, P.W.R., Rosen, L.S., Birken, S. and Low, B.W. (1985) Eur. J. Biochem. 153, 521–527.
[34] McPhalen, C.A., Schnebli, H.P. and James, M.N.G. (1985) FEBS Lett. 188, 55–58.