The EMBO Journal vol.12 no.3 pp.861-867, 1993

OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences

Alexey G.Murzin

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK, and Institute of Mathematical Problems of Biology, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia

Communicated by C.Chothia

A novel folding motif has been observed in four different proteins which bind oligonucleotides or oligosaccharides: staphylococcal nuclease, anticodon binding domain of asp-tRNA synthetase and B-subunits of heat-labile enterotoxin and verotoxin-1. The common fold of the four proteins, which we call the OB-fold, has a five-stranded β-sheet coiled to form a closed β-barrel. This barrel is capped by an α-helix located between the third and fourth strands. The barrel-helix frameworks can be superimposed with r.m.s. deviations of 1.4-2.2 Å, but no similarities can be observed in the corresponding alignment of the four sequences. The nucleotide or sugar binding sites, known for three of the four proteins, are located in nearly the same position in each protein: on the side surface of the β-barrel, where three loops come together. Here we describe the determinants of the OB-fold, based on an analysis of all four structures. These proposed determinants explain how very different sequences adopt the OB-fold. They also suggest a reinterpretation of the controversial structure of gene 5 ssDNA binding protein, which exhibits some topological and functional similarities with the OB-fold proteins.

Key words: asp-tRNA synthetase/bacterial cytotoxins/gene 5 DNA binding protein/staphylococcal nuclease/structure-function relationships

Introduction

The architectures of many proteins are based on closed β-sheet structures described as β-barrels (McLachlan, 1979) or orthogonally packed β-sheets (Chothia and Janin, 1982). A recent theoretical analysis of these proteins has produced a general classification of the observed structures (A.G.Murzin, A.M.Lesk and C.Chothia, in preparation). One of the theoretically derived classes brings together four different proteins that share a folding motif: the five-stranded Greek-key β-barrel capped by an α-helix located between the third and fourth strands (Figure 1).

This motif has been found in staphylococcal nuclease, in the anticodon binding domain of yeast asp-tRNA synthetase and in the B-subunits of cholera-like and Shiga-like cytotoxins from *Escherichia coli*. (After the first version of this paper had been submitted, a similar motif was observed in the N-terminal domain of staphylococcal enterotoxin B; see note in the conclusion.) The amino acid sequences of these four proteins show no significant similarity. Their binding sites for oligonucleotides or oligosaccharides, known in three of the four proteins, are found on the same part of the barrel surface. Because of its capacity to bind oligosaccharides and oligonucleotides this motif is called here the OB(oligomer binding)-fold.

Staphylococcal nuclease (SNS) is a small (149 amino acids) enzyme catalysing the hydrolysis of the phosphate backbone of DNA and RNA. The structures of SNS (Hynes and Fox, 1991) and its complex with inhibitor pdTp (Arnone *et al.*, 1971; Loll and Lattman, 1989) have been solved and refined at high resolution, ~1.7 Å .

Yeast aspartyl-tRNA synthetase (DRS) is a homodimeric enzyme that binds $tRNA^{Asp}$, ATP and aspartate. The crystal structure of DRS in complex with two cognate tRNAs has been solved at 3.0 Å and is being refined (Ruff *et al.*, 1991). The DRS subunit (557 residues) consists of two domains: the large C-terminal domain is characteristic of class II aminoacyl-tRNA synthetases (RS) (Moras, 1992), but the N-terminal part of the sequence forms the anticodon binding domain, which has no equivalent in the four other known RSs.

Heat-labile enterotoxin (LT) from *E.coli* is closely related to cholera toxin. These toxins have an AB_5 oligomeric structure with five identical B-subunits (LTB) which bind to oligosaccharides on the cell surface, and a catalytic A-subunit (LTA) that penetrates into the cytoplasm. Crystal structures of AB_5 LT and its complex with lactose have been determined recently at 2.3 Å resolution (Sixma *et al.*, 1991, 1992). Lactose binds each B-subunit in a virtually identical manner.

Verotoxin-1 (VT1) from *E.coli* belongs to the Shiga family

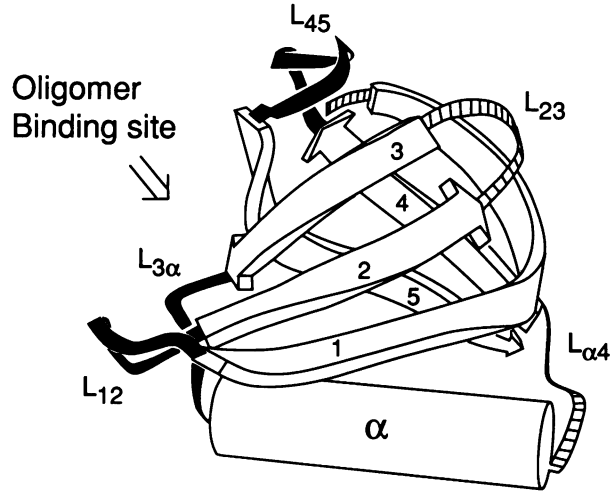


Fig. 1. Description of the OB-fold, based on its smallest representative, the B subunit of verotoxin-1 (VT1B). Five β-strands (numbered arrows) form a closed β-sheet (β-barrel), capped by an α-helix (cylinder). Protein loops are shown as ribbons, connecting the structural segments and numbered accordingly. Three variable loops (black) contribute residues in the oligomer binding site, as determined for the three other structures.

of enterotoxins and also has an AB_5 organization similar to that of the cholera toxin family. [The sequence of the B-subunit of verotoxin-1 (VT1B) is identical to that of Shiga toxin.] The crystal structure of the cell binding B-pentamer has been resolved at 2.2 Å resolution and resembles that of LTB (Stein *et al.*, 1992). The exact location of oligosaccharides bound to VT1B is not known from crystallography yet, but some information about the residues involved in sugar binding is available from site-directed mutagenesis studies (Jackson *et al.*, 1990; Perera *et al.*, 1991; Tyrrell *et al.*, 1992).

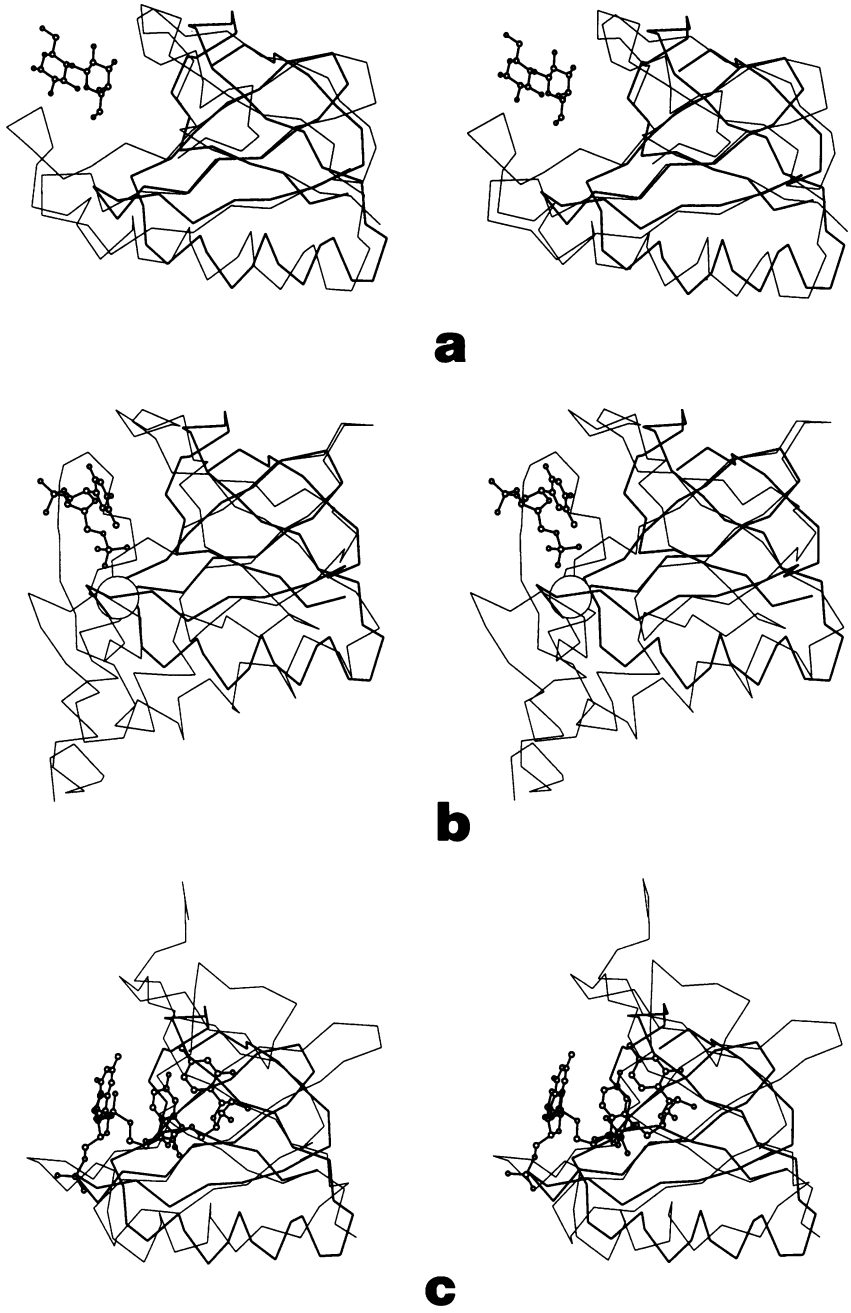
This paper presents comparative analysis of the four protein structures and reveals the structural determinants of their common fold. It discusses physical reasons for the observed architectural and functional similarities, and separates them from evolutionary ones.

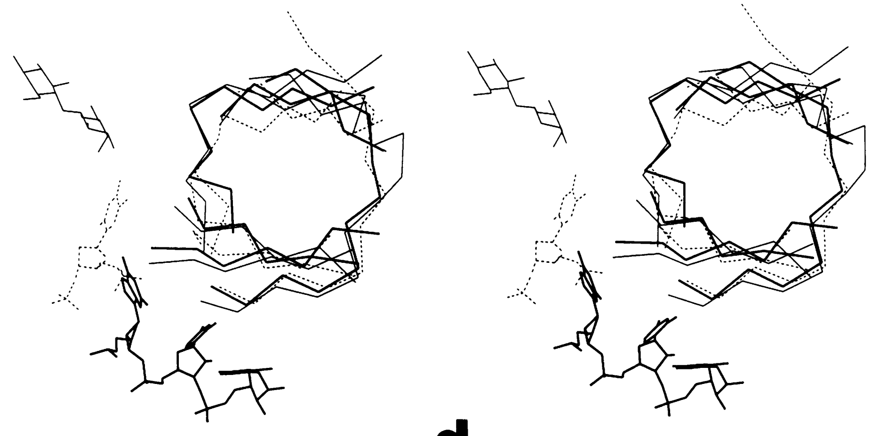
The structure of single strand DNA binding protein (GN5) coded by gene 5 of filamentous phage is found here to exhibit some topological and functional features of the OB-fold. A reinterpretation of current GN5 crystal structure is proposed. This reinterpretation brings together the new definition of GN5 secondary structure provided by NMR (Folkers *et al.*, 1991), the chain topology derived from the crystal structure (Brayer and McPherson, 1983) and the determinants of this fold presented in the first part of the paper.

Results and discussion

Structural comparisons and sequence alignment

The structures were superimposed initially to match their β-sheets, after which structurally homologous positions in helical and loop regions were added. The resulting superpositions are presented in Figures 2 and 3. Major differences in the β-sheet structures occur around non-homologous β-bulges. The helices in VT1B, LTB and DRS have very similar orientations and can be superimposed from





d

Fig. 2. (a, b and c) Stereodrawings of pairwise superpositions of VT1B (thick lines) on the other OB-fold proteins (thin lines, ball-and-stick drawing for the protein ligands): (**a**) LTB with lactose; (**b**) SNS with calcium ion (open circle) and inhibitor pdTp; (**c**) DRS anticodon binding domain with the $tRNA^{Asp}$ anticodon. (**d**) Superposition of the barrels of LTB (thin lines), SNS (dashed lines) and DRS (thick lines) showing similar location of their ligands (lactose, inhibitor pdTp and the $tRNA^{Asp}$ anticodon).

```
                                  1         L12    2    L23   3      L3α
                             t   m b               b m t    t m b
VT1B    1                tPDcVT--GKVEYTKYN-d--DDTFTVKVG----DKELFTn------
                          --^-------^-----      ------      -----
LTB     1  apqtitelcseyrntQI-YTinDKILSYTESmagkREMVIITFks---GETFQVev-----
               =====     ------^^---^-----      ------      -----
SNS     1          atstkklhk-EP--ATLIKAID-------gDTVKLMykgQPMTFRLllvdtpe
                         -----------^--           ------  -------
DRS   101        -eakdsdkeVL-FR--ARVHNTRQ-qgat--LAFLTLRqqaSLIQGLVkank---
                        ------------^----       --------  --------


          L3α                  α   Lα4    4         L45           5
                             *         b  m t                  t m  b
VT1B   33 ------------RWNLQSLLLSAQITGMTV-TIKTna----chn--------GGGFSEVIFR   69
                       =============  -------                  ---^-----
LTB    53 pgsqhidsqkkaIERMKDTLRITYLTETKIdKLCVwn----nkt--------PNSIAAISMKn 103
               =====================  --^----                  ---^------
SNS    44 tkhpkkgvekygpeasaftkkMVE-nakkI-EVEFd--kgqrtdkyg-----RGLA-YIYAdg- 96
                     =============    --------                 --------
DRS   146 --------egtiSKNMVKWAGSLnl-ESIV-LVRGivkkvdepiksatvqnleIHITKIYTIs-198
                       =========     ------------            -----^----
```

Fig. 3. Structurally based alignment of the protein sequences (in single letter code): VT1B, verotoxin-1 B subunit; LTB, heat-labile toxin B subunit; SNS, staphylococcal nuclease; DRS, asp-tRNA synthetase; N- and C-terminal residues are numbered. Structurally homologous residues are denoted by large case letters. Secondary structure regions are underlined: ===, α-helices; ---, β-strands; -ˆ -, insertions (β-bulges). t, m and b, positions in the barrel top, middle and bottom layers, respectively (due to the insertion in LTB the side chain of D22 points outward the barrel and I20 takes its place in the middle layer). (*) Helix anchor residue positions in VT1B, LTB and DRS. Highlighted residues are those in active sites as determined by crystallography for LTB (Sixma *et al.*, 1992), SNS (Loll and Lattman, 1989) and the anticodon binding domain of DRS (Cavarelli *et al.*, 1993), or suggested for VT1B from site-directed mutagenesis studies (Jackson *et al.*, 1990; Perera *et al.*, 1991).

the beginning to the end, but in SNS the helix is tilted away and only its C-terminal residues superimpose well with corresponding residues in the other structures. Despite rather small r.m.s. deviations (1.4–2.2 Å, see Table I) between the barrel–helix frameworks of the four structures, no significant homology of their sequences can be observed in their structural alignment (Figure 3).

The superposition of the four proteins allows the analysis of the structural determinants of the OB-fold, and the identification of positions in the sequence crucial for the formation of the fold. The amino acid residues that occupy these positions in the sequences are discussed.

Table I. R.m.s. deviations between the structures (Å) and numbers of structurally homologous residues (in brackets)

	VT1B	LTB	SNS	DRS
VT1B		1.5 (59)	1.8 (38)	1.9 (53)
LTB	1.4		2.2 (37)	1.8 (52)
SNS	1.8	2.3		1.4 (37)
DRS	1.3	1.6	1.2	

Shown below the diagonal are the deviations between only the β-barrel residues.

Structural determinants of the closed β-sheet

Hydrogen-bond network, geometrical parameters, β-bulges. The general hydrogen-bond network for the β-sheet structure in the proteins considered here is presented in Figure 4. The strands are linked together through main chain H-bonds to form a closed structure. [This closed structure was previously described as two separate β-sheets (Sixma *et al.*, 1991; Stein *et al.*, 1992): the first being the beginning of strand 1 and

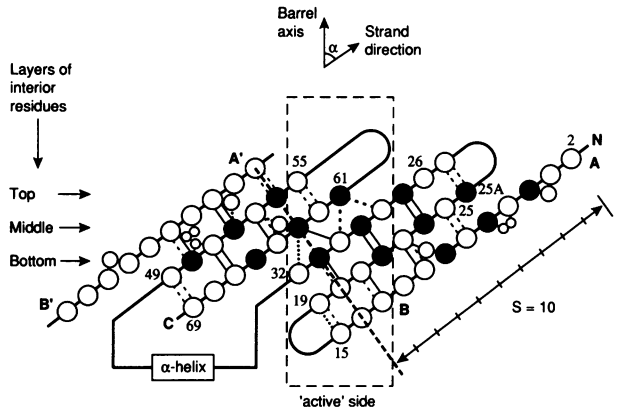


Fig. 4. Hydrogen-bond network and interior residue packing in the OB-barrels. The closed β-sheet is unrolled with the first strand repeated twice on both sides to show its shear number 10. The residues forming the three-layer structure in the interior are shown by black circles (residue numbering according VT1B with insertion between 25 and 26 residues). Conserved and non-conserved (see text) β-bulges in VT1B are shown with middle-size circles; the additional bulge and two-residue insertion in LTB are shown with the smallest circles. The hydrogen bonds common to all four structures are shown by thin solid lines; those observed in two or three structures by dashed lines, and those observed in only one structure by dotted lines. The 'active' side of the barrel is framed.

strands 4 and 5, and the second being the rest of strand 1 and strands 2 and 3.]

Closed β-sheet structure may be described by two integral parameters: the number of strands, n, and the shear number, S (McLachlan, 1979). In all four structures n is 5 and S is 10 (see Figure 4 for definition of S). These parameters allow the calculation of some geometrical features of the closed β-sheet (the mean radius, R, and the average angle between the stand directions and the barrel axis, α). They also link together the twist and the coiling (curvature) of the β-strands (A.G.Murzin, A.M.Lesk and C.Chothia, in preparation); for example, most of the strands in a β-sheet with this strand and shear number must be strongly coiled. As can be seen from Table II the observed geometries and those calculated from McLachlan's equations are close. The observed radii are all slightly larger than the theoretical values. This is because some β-strands are interrupted by β-bulges or other insertions that are ignored in our definition of shear number S.

β-Bulges can provide the small increase of the barrel radius and the required coiling of the β-strands. All four structures have one common β-bulge in the first strand, three of the four (except SNS) have a common bulge position in the fifth strand, and VT1B has an additional bulge in the beginning of the first strand. LTB with its wider barrel has a two-residue insertion in the middle of the first strand, H-bonded to the additional non-homologous bulge in the fourth strand (Figure 4).

Residue packing inside the closed β-sheet. All four structures have a regular residue packing in the interior of the closed β-sheet comprising the three stacking layers of residues. The three-layer structure previously observed in the α/β-barrel fold (Lesk *et al.*, 1989) and in the β-trefoil fold (Murzin *et al.*, 1992) seems to be a general principle for packing barrel interiors. Each β-strand in the OB-fold barrel contributes one residue in each layer (see Figures 3 and 4) like those in the β-trefoil fold. [This is due to the same

Table II. Geometrical parameters and mean volume of interior residues of the β-barrels in OB-fold proteins

Protein	Radius R (Å)	Angle α (deg)	Mean residue volume v (Å^3)
Theory	6.7	56	
VT1B	6.9	51	141
LTB	7.5	60	172
SNS	7.1	57	151
DRS	6.9	55	140

Theoretical mean radius of the barrel, R, and average angle between strand directions and the barrel axis, α, have been calculated according to McLachlan (1979): $\tan\alpha = Sa/nb$; $R = b/2\sin(\pi/n)\cos\alpha$; with average β-structure parameters a =3.3 Å and b = 4.4 Å and the observed number of strands, n = 5 and shear number, S = 10.

relationship, $S = 2n$ between shear number, S and the number of strands, n in both folds (in the β-trefoil fold $n = 6$ and $S = 12$). By contrast, the α/β-barrel fold has $n = 8$ and $S = 8$ (that is $S = n$), and its layers consist of residues from alternate strands (Lesk *et al.*, 1989)].

As in the β-trefoil fold the top layer residues are partly exposed and are heterogeneous in nature. Indeed the top layer is incomplete in LTB and VT1B due to one short strand. The two internal layers (the bottom layer is covered by α-helix) are composed of hydrophobic and neutral residues.

The interior residues (comprising the middle and the bottom layers) in LTB have mean volume ~170 Å^3 to fill its wider barrel, but the three smaller barrels are filled by smaller residues: their mean volume is ~140 Å^3 (VT1B and DRS) or 150 Å^3 (SNS). It should be noted that the middle layer residues vary in size more than those from the bottom layer.

β-Barrel deformation. All four β-barrels are similarly flattened, that is they all have an elliptical cross-section with the long axis going between the third and the fifth strands toward the first strand. The flattening is a result of increasing the interstrand angles and/or the strong coiling of the β-strands at the ends of the long axis. These deformations are due to: (i) the parallel β-structure, which occurs only between the third and the fifth strands, being able to adopt a larger interstrand angle than an antiparallel pair without breaking hydrogen bonds; and (ii) strong coiling or bending of the first strand, which lies opposite to these two. The first strand has the common β-bulge and it is also bent in the middle in all four structures. In LTB this bend results from the insertion, but in the three other structures it is due to small residues (Gly or Ala) contributed by the first strand in the middle layer. To maintain density of this layer some of the other strands contribute large (aromatic) residues to it.

Barrel–helix interface

α-Helices packed against the ends of β-barrels usually display the following features.

(i) The helix packs against the extreme (top or bottom) layer of residues roughly parallel to its plane (i.e. perpendicular to the barrel axis).

(ii) The rest of the β-sheet structure protrudes beyond this layer and packs around the sides of the helix. A cavity on the barrel axis is filled by a large hydrophobic residue coming from the helix.

(iii) The direction of helix axis is determined by the systematic deformation of the barrel (that is, the helix lies along the long axis of the barrel cross-section).

SNS shows reasonable agreement with the pattern described but it is complicated by the presence of an additional helix in the barrel–helix interface. The three other structures follow the pattern completely having very similar interfaces.

The bottom layers in these proteins are composed mainly of Cβ-branched residues: those in VT1B are four Val and one Thr, but in the expanded LTB barrel each of the Val residues is substituted with Ile, and the Thr residue with Val. These residues do not reach the barrel axis or penetrate deeply into the central cavity; that is filled by a residue from the helix: Leu40 in VT1B and Leu72 in LTB. The bottom layer in DRS is composed of three Val, one Ile and one Leu, and the smaller central cavity is filled by Ala157 (that is by the residue homologous to L40 in VT1B and L72 in LTB) with assistance from Leu160. As a result the observed distances between the helix axes and the bottom layer planes are very similar (~7.5 Å), but the DRS helix is shifted by half a turn along its axis in comparison with LTB and VT1B helices.

Relationship between the OB-fold and its binding site(s)

Structural comparisons reveal striking similarities in the locations of the binding sites that are known for three of the four proteins. The details of the binding sites for the inhibitor pdTp and calcium in SNS were described by Loll and Lattman (1989), and those for lactose in LTB by Sixma *et al.* (1992). The detailed analysis of interactions of DRS with its cognate $tRNA^{Asp}$, including the anticodon recognition, will be published elsewhere (Cavarelli *et al.*, 1993).

When structures of complexes of SNS with its inhibitor pdTp, LTB with lactose and DRS with $tRNA^{Asp}$ are superimposed, the inhibitor, lactose and the tRNA anticodon (^{34}GUC^{36}) appear nearly in the same place—the OB site (Figure 2d). The lactose molecule, of which only the galactose moiety binds LTB, and C36 of the anticodon loop are on the opposite ends of this elongated site. The SNS inhibitor pdTp binds in the middle of the site and partly overlaps with galactose and G34. It occupies the central position of three-nucleotide binding site proposed for SNS from NMR experiments (Weber *et al.*, 1992). The other two proposed positions approximately correspond to those occupied by U35 and galactose. If that is so, the three-nucleotide binding sites of SNS and DRS overlap with two positions but accommodate oligonucleotides in the opposite orientations.

VT1B binds longer oligosaccharides than LTB does. The sugar binding site(s) of VT1 is unknown and hardly overlaps with that of LTB, because the lactose molecule bound to LTB is far from the VT1B surface, when the two structures were superimposed. Stein *et al.* (1992) suggested that the VT1B sugar binding site is in a groove between neighbouring B-subunits of the pentamer. However, with respect to individual subunits the proposed site is still a part of the common OB site. If the DRS structure is superimposed on to one subunit of the VT1B pentamer, the anticodon loop comes into the groove.

Although the sites are in similar regions of the structures, few residues involved in binding of these ligands occur in exactly homologous positions, in the β-sheet. Most binding site residues (see Figure 3) come from the loops protruding outward from the barrel axis. These loops differ in sequence, length and conformation, and this makes the detailed structures of binding sites quite different.

The similar location of binding sites results from the chain topology and from the barrel geometry. Due to the large slope of the β-strands to the barrel axis (56°) the L_{12} and L_{45} loops (see Figures 1 and 3 for loop numbering) appear on the same side of the barrel separated by the barrel height (~7 Å between the top and the bottom layers). The shallow grove formed by these loops and the barrel side surface is a good potential site for binding of elongated ligands. The N-terminus and $L_{3\alpha}$ loop are close to this site and may extend it (as they do in LTB). It should be noted that the opposite side of the barrel, where the other two short loops (L_{23} and L_{a4}) appear, shows no similar features.

OB-fold in gene 5 protein

The crystal structure of GN5 displays some topological and functional characteristics of OB-fold. It can be superimposed on VT1B structure (Figure 5) giving r.m.s. deviation 1.9 Å for 28 'barrel' Cα atoms. This superposition puts the single strand DNA binding site proposed for GN5 in the common OB site.

However, the structure deposited in the Protein Data Bank shows neither regular H-bonding in its β-sheet structure nor layer-like packing in its interior. As has been demonstrated by modern tests of quality of protein coordinates, the GN5 crystal structure has poor stereochemistry (Morris *et al.*, 1992) and rather low 3D profile score against its own amino acid sequence (18.59, or 46% of expected value of 40; Lüthy *et al.*, 1992). Moreover, NMR studies of GN5 reveal significant discrepancies with the crystal structure (Folkers *et al.*, 1991). The NMR gives more prominent β-sheet secondary structure that is consistent with topology provided by the crystal structure, but it is more regular and differs in residue assignment. Therefore it is very probable that GN5 has a structure that is much closer to the OB-fold.

The structural principles for OB-fold revealed in this study are applicable to GN5 sequence. The GN5 secondary structure provided by NMR (Folkers *et al.*, 1991) is consistent with an H-bond network generally similar to the OB-fold (Figure 6). Only few additional H-bonds are needed to form closed β-sheet with $S = 10$. The interior of this sheet would have three hydrophobic layers, compositions of which are consistent with the barrel radius. Though the bottom layer in the proposed structure is not homogeneous, here this is not essential because GN5 has an irregular connection between the third and the fourth strands instead the α-helix. Tyr56 in this connection could fill a cavity in the barrel bottom. There is a large insertion in the first strand that facilitates its bending.

Conclusion

This study shows that the OB-fold represents the new common fold. Like other common folds it is suitable for a wide variety of protein sequences. It also provides a clear example of a fold-related binding site, the position of which and kind of ligands depend on the fold architecture and topology rather than on protein sequences. Such relationships between protein fold and active site are not unusual in other

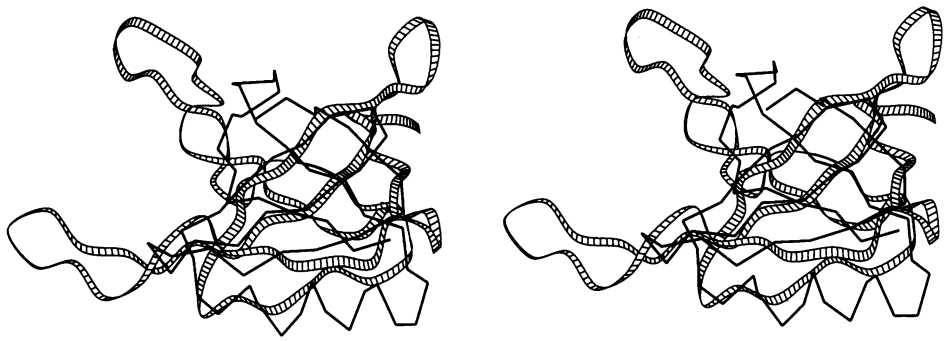


Fig. 5. Stereodrawing of a superposition of VT1B (thick lines) on the crystal structure of GN5 (ribbon), showing similarities in the topology of the protein chain and in the β-structural core.

common folds (the best example is α/β-barrel fold, see Brändén, 1991; Wilmanns *et al.*, 1991).

There are two general views on the origin of common protein folds. One is that proteins sharing the same fold diverged from a common ancestor. The alternative view is that the number of possible stable folds for protein chain is limited by physical reasons, and that protein chains that adopt similar folds arise independently.

On a first look the analysis of OB-fold proteins supports the second view. The similar features in structural and functional organization can be explained by general principles governing the protein structure. Moreover, the Greek-key topology of the OB-fold also is one of the most favourable topologies for β-sheet proteins (Richardson, 1977, 1981). However, the theoretical analysis of protein structures also shows us how varied the stable folding motifs can be, even if they have similar secondary structure. For example, five-strand closed β-sheets can have $4!2^4 = 384$ different topologies, and a few (~ 10) of them have favourable interstrand connectivity like that of the OB-fold. The five-strand β-sheet can be closed not only with shear number $S = 10$, but also with $S = 8$, as observed in several protein structures. It could well be deformed in a manner different from that in the OB-fold, etc. All these alternatives reduce the probability of the independent origin of the four OB-fold proteins.

A possible solution for this dilemma comes from an estimation of a number of different folds for the pool of natural protein sequences, that are based on an analysis of sequences and structures deposited in data banks. A recent estimate made by Chothia (1992) shows this number is of the order of 10^3, i.e. much less than the number of theoretically expected stable folds. If this is true, the OB-fold, and other common folds, may represent the stable folding motifs that appeared at the origin of proteins. Wide appearance of these motifs amongst present protein structures could have been provided by their two intrinsic features: (i) a fold-related potential binding site(s) that could be edited by evolution for many different functions, and (ii) an ability to accommodate a large variety of quite different sequences that allowed their fast and broad divergence.

OB-fold in staphylococcal enterotoxin B?

After the first version of this paper had been submitted, the structure of staphylococcal enterotoxin B (SEB) was published (Swaminathan *et al.*, 1992). It consists of two domains, with the N-terminal domain having secondary structure and topology of the OB-fold. If this domain does contain the OB-fold it will be the fourth example of this fold observed in the proteins secreted by bacterial pathogens. This may have the following implications: (i) these four proteins sharing the OB-fold may be evolutionarily related; (ii) it may be expected that the OB-fold will be found again in other bacterial toxins; (iii) the region in the SEB N-terminal domain, that corresponds to the oligomer binding site of the OB-fold, may be of functional importance.

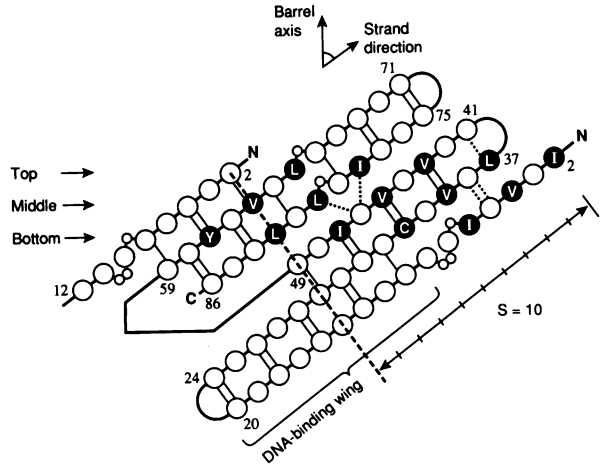


Fig. 6. OB-fold based model of a hydrogen-bond network and an interior residue packing proposed for GN5 (cf. Figure 4). The hydrogen bonds observed by NMR (Folkers *et al.*, 1991) are shown by thin solid lines, the hypothetical hydrogen bonds closing the β-sheet with $S = 10$ are shown by dotted lines.

Coordinates and calculations

Coordinates of staphylococcal nuclease complex with inhibitor (file 1SNC) and those of gene 5 protein (file 2GN5) were taken from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977). Coordinates of heat-labile enterotoxin complex with lactose (file 1LTT) were provided by T.K.Sixma and W.G.J.Hol, those of verotoxin-1 B pentamer (file 1BOV) by P.E.Stein and R.J.Read. Coordinates of the complex of asp-tRNA synthetase with cognate tRNA were provided by B.Rees, J.Cavarelli and D.Moras.

Initial comparisons of the structures and selection of structurally homologous positions were done using FRODO (Jones, 1985) on an Evans & Sutherland ESV Workstation. Detailed comparison and all the other calculations (including graphics of Figures 1, 2 and 5) were carried out with programs of A.M.Lesk (Lesk, 1986; and references contained therein).

Acknowledgements

This paper is dedicated to my teacher, Professor O.B.Ptitsyn, whose initials coincide with the fold name not quite by chance. I thank C.Chothia and A.M.Lesk for help and valuable discussion. I am grateful to T.K.Sixma, W.G.J.Hol, P.E.Stein, R.J.Read, B.Rees, J.Cavarelli and D.Moras for making the coordinates of LT, VT1B and DRS available to me prior to their release and for comments on the manuscript. This work was supported by the EMBO long-term fellowship ALTF 287-1990.

References

Arnone,A., Bier,C.J., Cotton,F.A., Day,V.W., Hase,E.E.,Jr, Richardson,D.C., Richardson,J.S. and Yonath,A. (1971) *J. Biol. Chem.*, **246**, 2302-2316.

Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F.,Jr, Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535-542.

Brändén,C.-I. (1991) *Curr. Opin. Struct. Biol.*, **1**, 978-983.

Brayer,D.D. and McPherson,A. (1983) *J. Mol. Biol.*, **169**, 565-596.

Cavarelli,J., Rees,B., Ruff,M., Thierry,J.C. and Moras,D. (1993) *Nature*, in press.

Chothia,C. (1992) *Nature*, **357**, 543-544.

Chothia,C. and Janin,J. (1982) *Biochemistry*, **21**, 3955-3965.

Folkers,P.J.M., van Duynhoven,J.P.M., Jonker,A.J., Harmsen,B.J.M, Konings,R.N.H. and Hilbers,C.W. (1991) *Eur. J. Biochem.*, **202**, 349-360.

Hynes,T.R. and Fox,R.O. (1991) *Proteins*, **10**, 92-105.

Jackson,M.P., Wadolkowski,E.A., Weinstein,D.L, Holmes,R.K. and O'Brien,A.D. (1990) *J. Bacteriol.*, **172**, 653-658.

Jones,T.A. (1985) *Methods Enzymol.*, **115**, 252-270.

Lesk,A.M. (1986) In Saccone,C. (ed.), *Biosequences: Perspectives and User Services in Europe*. EEC, Bruxelles, pp. 23-28.

Lesk,A.M., Brändén,C.-I. and Chothia,C. (1989) *Proteins*, **5**, 139-148.

Loll,P.J. and Lattman,E.E. (1989) *Proteins*, **5**, 183-201.

Lüthy,R., Bowie,J.U. and Eisenberg,D. (1992) *Nature*, **356**, 83-85.

McLachlan,A.D. (1979) *J. Mol. Biol.*, **128**, 49-79.

Moras,D. (1992) *Trends Biochem. Sci.*, **17**, 159-164.

Morris,A.L., MacArtur,M.W., Hutchinson,E.G. and Thornton,J.M. (1992) *Proteins*, **12**, 345-364.

Murzin,A.G., Lesk,A.M. and Chothia,C. (1992) *J. Mol. Biol.*, **223**, 531-543.

Perera,L.P., Samuel,J.E., Holmes,R.K. and O'Brien,A.D. (1991) *J. Bacteriol.*, **173**, 1151-1160.

Richardson,J.S. (1977) *Nature*, **268**, 495-500.

Richardson,J.S. (1981) *Adv. Protein Chem.*, **34**, 167-339.

Ruff,M., Krishnaswamy,S., Boeglin,M., Poterszman,A., Mitshler,A., Podjarny,A., Rees,B., Thierry,J.C and Moras,D. (1991) *Science*, **252**, 1682-1689.

Sixma,T.K., Pronk,S.E., Kalk,K.H., Wartna,E.S., van Zanten,B.A.M., Witholt,B. and Hol,W.G.J. (1991) *Nature*, **351**, 371-377.

Sixma,T.K., Pronk,S.E., Kalk,K.H., van Zanten,B.A.M., Berghuis,A.M. and Hol,W.G.J. (1992) *Nature*, **355**, 561-564.

Stein,P.E., Boodhoo,A., Tyrrell,G.J., Brunton,J.L. and Read,R.J. (1992) *Nature*, **355**, 748-750.

Swaminathan,S., Furey,W., Pletcher,J. and Sax,M. (1992) *Nature*, **359**, 801-806.

Tyrrell,G.J., Ramotar,K., Toye,B., Boyd,B., Lingwood,C.A. and Brunton,J.L. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 524-528.

Weber,D.J., Gittis, A.G., Mullen,G.P., Abeygunawardana,C., Lattman,E.E. and Mildvan,A.S. (1992) *Proteins*, **13**, 275-287.

Wilmanns,M., Hyde,C.C., Davies,D.R., Kirschner,K. and Jansonius,J.N. (1991) *Biochemistry*, **30**, 9161-9169.

Received on October 5, 1992; revised on November 16, 1992