# BOF: a novel family of bacterial OB-fold proteins

Krzysztof Ginalski[a,*], Lisa Kinch[b], Leszek Rychlewski[c], Nick V. Grishin[a,b]

[a]*Department of Biochemistry, University of Texas, Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-9038, USA*
[b]*Howard Hughes Medical Institute, University of Texas, Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-9050, USA*
[c]*BioInfoBank Institute, ul. Limanowskiego 24A, 60-744 Poznań, Poland*

Received 1 April 2004; accepted 19 April 2004

Available online 12 May 2004

Edited by Robert B. Russell

**Abstract Using top-of-the-line fold recognition methods, we assigned an oligonucleotide/oligosaccharide-binding (OB)-fold structure to a family of previously uncharacterized hypothetical proteins from several bacterial genomes. This novel family of bacterial OB-fold (BOF) proteins present in a number of pathogenic strains encompasses sequences of unknown function from DUF388 (in Pfam database) and COG3111. The BOF proteins can be linked evolutionarily to other members of the OB-fold nucleic acid-binding superfamily (anticodon-binding and single strand DNA-binding domains), although they probably lack nucleic acid-binding properties as implied by the analysis of the potential binding site. The presence of conserved N-terminal predicted signal peptide indicates that BOF family members localize in the periplasm where they may function to bind proteins, small molecules, or other typical OB-fold ligands. As hypothesized for the distantly related OB-fold containing bacterial enterotoxins, the loss of nucleotide-binding function and the rapid evolution of the BOF ligand-binding site may be associated with the presence of BOF proteins in mobile genetic elements and their potential role in bacterial pathogenicity.**


*Keywords:* OB-fold; Fold recognition; Structure–functional assignment; Binding site; Bacterial pathogenicity

## 1. Introduction

A number of different protein families possess an oligonucleotide/oligosaccharide-binding (OB)-fold domain, which consists mainly of five antiparallel β-strands forming a closed or partly opened barrel [1,2]. In general, the OB-fold core architecture supports a common binding face for a range of biological molecules including nucleic acids, oligosaccharides, and proteins. The structural classification of proteins (SCOP) [3] currently defines nine OB-fold superfamilies including the nucleic acid-binding protein superfamily and the bacterial enterotoxin superfamily for which the fold was first named and described [4]. Here, we describe a novel family of bacterial OB-fold (BOF) proteins from the nucleic acid-binding superfamily. The BOF family encompasses exclusively uncharacterized hypothetical sequences of unknown function and is partially catalogued in Pfam [5] (DUF388) and COG [6] (COG3111) databases.

* Corresponding author. Fax: +1-214-648-9099.
*E-mail address:* kginal@chop.swmed.edu (K. Ginalski).

## 2. Materials and methods

### 2.1. Identification of BOF family members

To perform this task, an exhaustive, transitive PSI-Blast search procedure was applied. Initially, PSI-Blast [7] searches against the NCBI non-redundant protein database (nr posted Oct 29, 2003, 1 529 764 sequences) until profile convergence with inclusion threshold of 0.01 were carried out using the consensus sequence of DUF388 as an initial query. Consequently, collected sequences were subjected to further PSI-Blast searches until no new sequences were found.

### 2.2. Structural assignment for BOF proteins

Initially, the consensus sequence of DUF388 was subjected to newly developed meta profile [8] alignment method Meta-BASIC [9] available at http://basic.bioinfo.pl. This fold recognition approach utilizes comparison of sequence profiles combined with predicted secondary structure (what we call meta profiles) enabling detection of very distant relationships between proteins even if the tertiary structure for the reference protein is not known. Specifically, the consensus sequence of DUF388 was compared to all 6249 PfamA [5] families and to 7225 proteins (representatives at 90% of sequence identity) extracted from Protein Data Bank (PDB) [10]. The same comparison was also conducted using PSI-Blast and RPS-Blast [7].

In addition, the consensus sequence of DUF388 as well as several members of this family were analyzed with CDD [46] (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and SMART [11] (http://smart.embl-heidelberg.de) search tools to detect conserved protein domains annotated in SMART, Pfam and COG [6] databases. This analysis also included searches for transmembrane segments (with TMHMM2 [12]), signal peptides (SignalP [13]), low compositional complexity (CEG [14]) and coiled coil (Coils2 [15]) regions, as well as regions containing internal repeats (Prospero [16]).

Finally, both consensus sequence of DUF388 and one of the family members, *Escherichia coli* protein ygiW precursor (gi|26249594), were submitted to the Meta Server [17] (http://bioinfo.pl/meta) that assembles various secondary structure prediction and top-of-the-line fold recognition methods. Collected predictions were screened with 3D-Jury [18], the consensus method of fold recognition servers. The default servers used by the 3D-Jury system for consensus building included: ORFeus [8], SamT02 [19], FFAS03 [20], mGenTHREADER [21], INBGU [22], RAPTOR [23], FUGUE-2 [24], and 3D-PSSM [25].

Consequently, based on the final sequence-to-structure alignment (see below), a 3D model of an OB-fold domain of *E. coli* protein ygiW precursor (gi|26249594) was built with the MODELLER program [26] using the *E. coli* AspRS structure (PDB code: 1c0a) [27] as a template. Independently, the structure of this domain was also modeled ab initio using the ROSETTA program [28].

### 2.3. Generation of sequence-to-structure alignment

Multiple sequence alignment for BOF family was prepared using PCMA program [29] followed by final manual adjustments with respect to corresponding sequence–structure mapping. Sequence-to-structure alignment between BOF family and selected OB-fold domains of known structure from superfamily of nucleic acid-binding proteins was built using consensus alignment approach and 3D assessment [30] based on 3D-Jury results for DUF388 consensus sequence and *E. coli* protein ygiW precursor (gi|26249594). Sequences of

0014-5793/$22.00 © 2004 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/j.febslet.2004.04.086

selected nucleic acid-binding OB-fold domains belonging to anticodon-binding (AspRS PDB codes: 1c0a [27], 1b8a [31], 1efw [32], 1asy [33]; LysRS PDB code: 1bbu [34]) and SSB (RPA32 PDB code: 1quq [35]) families were aligned first based on the superposition of their 3D structures. To define more precisely general conservation pattern in template sequences, their close homologs were collected with PSI-Blast searches and subsequently aligned using PCMA. Finally, the sequence of a distantly related cholera toxin (PDB code: 3chb) [36] from the superfamily of bacterial enterotoxins was also included in the structure-based alignment.

### 2.4. Phylogenetic analysis

Phylogenetic analysis was performed with the MOLPHY package [37] using final multiple sequence alignment of BOF proteins encompassing complete OB-fold domain. Distances estimated with the amino acid transition probability matrix of Jones, Taylor, and Thornton [38] were used to generate an initial tree topology with Njdist program [37,39]. This topology was subsequently improved using maximum likelihood local rearrangement search (-R option of the PROTML program [37]). Local bootstrap probabilities for each internal branch were estimated by the RELL method ($10^3$ replications) [40].

## 3. Results and discussion

### 3.1. BOF proteins possess an OB-fold structure

This finding is a result of a large-scale structure–functional annotation performed for all PfamA protein families of unknown function (DUF) with Meta-BASIC [9], a novel sensitive approach for recognition of distant similarity between proteins based on consensus alignments of meta profiles [8]. Specifically, Meta-BASIC mapped the consensus sequence of DUF388, with an above threshold (>12) *Z*-score, to a number of OB-fold structures from the superfamily of nucleic acid-binding proteins. Our benchmarking results show that predictions with *Z*-score above 12 have less than 5% probability to be incorrect (using rigorous structural criteria). Importantly, both PSI-Blast [7] and RPS-Blast were unable to find any reliable matches (with *E*-value below 0.1) to other PfamA families or to known protein structures. This interesting, but unexpected, assignment is also not possible with other standard sequence similarity search tools, such as CDD [46] or SMART [11] used with default settings.

To confirm the correctness of Meta-BASIC assignment, further analysis was carried out using the Meta Server [17] coupled with the 3D-Jury system [18]. This protein structure prediction approach was proven to be one of the best performing methods in the CASP5 experiment [41]. 3D-Jury assigned reliable scores above 50 to OB-fold domain for both the consensus sequence of DUF388 and its family member, *E. coli* protein ygiW precursor (gi|26249594). As shown previously [42], 3D-Jury scores above 50 correspond to essentially correct predictions, meaning that in over 90% of the cases the overall fold of the model is similar to the experimental structure (certain exceptions with α-helical domains exist). In particular, the highest scoring 3D-Jury predictions pointed exclusively to several nucleic acid-binding proteins, including anticodon-binding domains of aspartyl-tRNA (AspRS) [27,31–33] and lysyl-tRNA (LysRS) [34] synthetases from different species. In addition, we also observed replication protein A 32 Kda subunit (RPA32) [35] belonging to the family of single strand DNA-binding domains (SSB) as a frequently selected OB-fold template by a number of different fold recognition servers.

Additional indicators of the correct fold assignment for DUF388 include a general conservation of the hydrophobic character of the barrel interior and an excellent mapping of predicted and observed secondary structures (Fig. 1) that embrace all core elements of the OB-fold. Accordingly, we used the *E. coli* AspRS structure (PDB code: 1c0a) [27] selected both by Meta-BASIC and 3D-Jury to build a reliable 3D model for the OB-fold domain of the *E. coli* protein ygiW precursor (gi|26249594) (Fig. 2A). Finally, we obtained similar OB-fold-like structure for this protein using ab initio approach implemented in ROSETTA program [28] (Fig. 2B).

### 3.2. Phyletic distribution and relation to biological function

Exhaustive PSI-Blast searches with OB-fold domain of DUF388 consensus sequence, as well as all members of the BOF family performed against the non-redundant protein database (*E*-value threshold of 0.01) revealed in total 37 hypothetical proteins, including also sequences belonging to an uncharacterized cluster of orthologs COG3111. Notably, several AspRS and LysRS OB-folds appeared as first hits with below threshold scores (*E*-value $\sim 0.1$) to many query BOF sequences after PSI-Blast iterations to convergence. Despite a lack of significant statistical support, these marginal PSI-Blast hits point to an evolutionary relationship between the BOF family and other OB-fold proteins from the nucleic acid-binding superfamily (anticodon-binding and SSB domains) that is further justified by fold recognition prediction. All members of BOF family come from the proteobacterial phylum, including the causitive agents of typhoid fever (*Salmonella typhimurium*), plague (*Yersinia pestis*), cholera (*Vibrio cholerae*), whooping cough (*Bordetella parapertussis*) and a number of other pathogenic (and some non-pathogenic) strains. The various bacterial genomes generally include one or more BOF family members, with several of these proteins encoded by prophage inserts thought to be associated with bacterial virulence and horizontal gene transfer events [43,44] (sequences shown in italics, Fig. 3).

The association of BOF proteins with prophage inserts or virulence plasmids, that can act as mobile genetic elements, provides a mechanism for rapid evolution of the BOF family. Such evolution could lead to adaptation of the potential OB-fold ligand-binding site (see description below), consistent with a general lack of strict amino acid conservation among potential binding determinants of various BOF family members. Such a distribution closely resembles that found for the distantly related OB-fold containing bacterial superantigens and enterotoxins that function to bind various proteins (superantigens) and polysaccharides (enterotoxins), leading to a number of toxic responses in human hosts.

In addition to providing a mechanism for genetic variation, the presence of BOF proteins in several virulence plasmids (Shiga toxin 2 phage 933W from *E. coli* O157:H7, pKDSC50 virulence plasmid of *Salmonella enterica* Serovar Choleraesuis, and the virulence plasmid of *S. typhimurium* strain LT2) and in a plasmid that confers nickel/cobalt resistance (*Hafnia alvei* 5-5 plasmid pNRS148) suggests a potential role of this family in bacterial pathogenicity. Examples of pathogenic characteristics associated with genes contained within such plasmids often include antibiotic resistance determinants, virulence factors associated with invasion and toxicity, and fitness traits that confer increased survival to the bacterium [45].

### 3.3. Evolutionary origin of BOF proteins

All identified BOF proteins encompass a single OB-fold domain with a conserved N-terminal region predicted as a

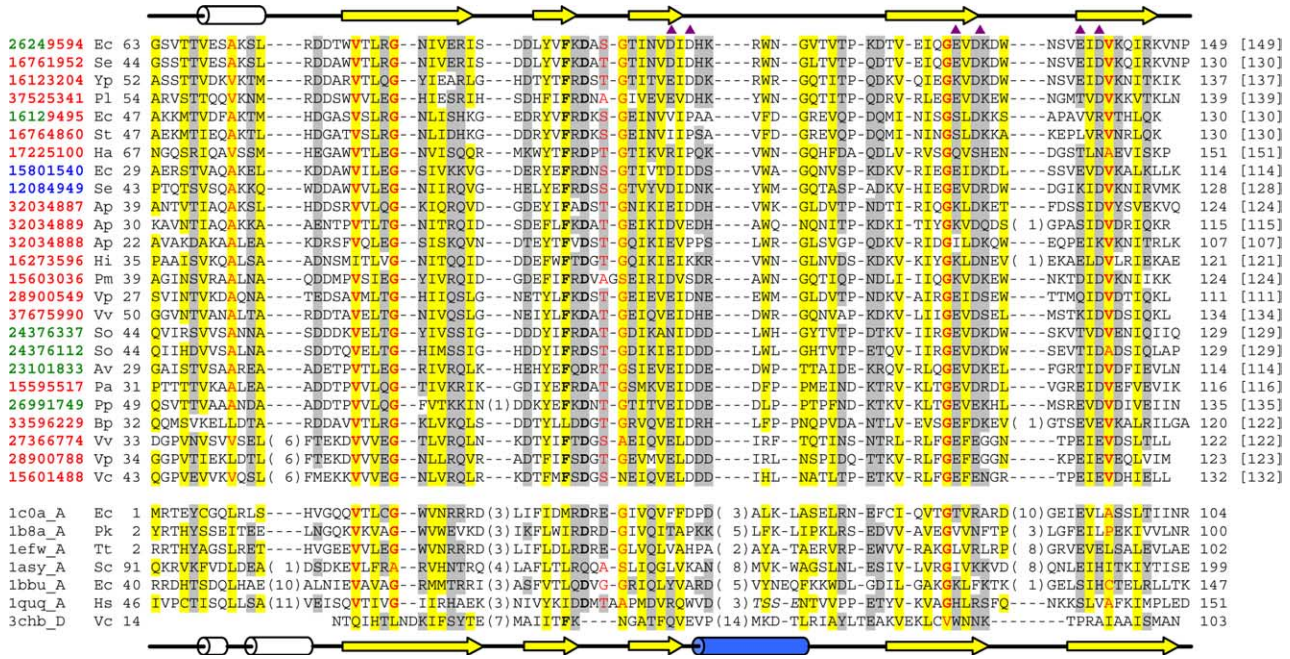

Fig. 1. The sequence-to-structure alignment for BOF family and selected structures from the aniticodon-binding, SSB and bacterial AB5 toxins families. Each sequence is labeled by the NCBI gene identification (gi) number or PDB code followed by an abbreviation of the species name (Ap, *Actinobacillus pleuropneumoniae*; Av, *Azotobacter vinelandii*; Bp, *B. parapertussis*; Ec, *E. coli*; Ha, *H. alvei*; Hi, *Haemophilus influenzae*; Hs, *Homo sapiens*; Pa, *Pseudomonas aeruginosa*; Pk, *Pyrococcus kodakaraensis*; Pl, *Photorhabdus luminescens*; Pm, *Pasteurella multocida*; Pp, *Pseudomonas putida*; Sc, *Saccharomyces cerevisiae*; Se, *S. enterica*; So, *Shewanella oneidensis*; St, *S. typhimurium*; Tt, *Thermus thermophilus*; Vc, *Vibrio cholerae*; Vp, *Vibrio parahaemolyticus*; Vv, *Vibrio vulnificus*; Yp, *Y. pestis*). gi numbers for BOF sequences are colored blue for inclusion in prophage inserts, red for pathogenic bacterial strain, and green for non-pathogenic bacterial strain. Conserved N-terminal signal peptide region in BOF family is not shown. BOF sequences identical in more than 90% to any other sequence and those with possible errors are not presented. The first and last residue numbers are indicated before and after each sequence, with the total BOF sequence length following in square brackets. Numbers in parentheses specify the number of excluded residues. Uncharged residues in mainly hydrophobic positions are highlighted in yellow and conserved small residues are shown in red letters. Highly preserved residues in BOF family are denoted in bold. Locations of the secondary structure elements in *E. coli* protein ygiW precursor (gi|26249594) (consensus of secondary structure predictions) and *E. coli* AspRS (PDB code: 1c0a) are marked above and below the sequences, respectively. N-terminal α-helical region that displays a unique conformation in anticodon-binding and some SSB families but not in other OB-folds is shown in white. Proposed BOF binding site residues from the conserved OB-fold core are marked with violet triangles.

signal peptide, suggesting their possible functional role in periplasm. The multiple sequence alignment presented in Fig. 1 illustrates the BOF family conservation pattern compared to those of other OB-fold domains. Highly conserved Val 80 (numbering from *E. coli* protein ygiW precursor, gi|26249594), Phe 97, and Val 141 contribute to the hydrophobic interior of the barrel, while an invariant glycine typical for OB-fold containing proteins (Gly 84) allows a β-bulge in the first β-strand. A second invariant BOF family glycine (Gly 128) present in template structures from anticodon-binding and SSB families (Fig. 1) seems to be less conserved among OB-fold proteins. The last invariant BOF family residue (Asp 99) is also shared among anticodon-binding and some SSB domains, providing an anchor for the N-terminal part of the structure and stabilizing its unique conformation.

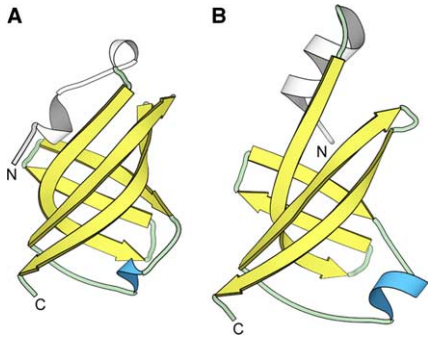


Fig. 2. 3D model of OB-fold domain for *E. coli* protein ygiW precursor (gi|26249594). (A) Fold recognition model based on *E. coli* AspRS structure (PDB code: 1c0a) selected by 3D-Jury method. (B) Similar ab initio structure obtained independently using ROSETTA program. Color shading of secondary structure elements corresponds to that in Fig. 1.

The shared conservation of structurally important residues between the BOF family and other OB-fold domains belonging to the nucleic acid-binding superfamily suggests a common evolutionary origin for these sequences. Although residues contributing to the hydrophobic core and the β-bulge (Gly 84) are generally conserved in majority of OB-fold sequences, two BOF family invariant residues (Gly 128 and especially Asp 99) are more specifically conserved in nucleic acid-binding proteins. A somewhat unique conservation of Asp 99 in anticodon-binding and SSB families correlates with a specific packing of the N-terminal region that extends the OB-fold barrel with a few residues lining against the first β-strand in a parrallel strand-like orientation (Fig. 4A and B). This segment forms a significantly different conformation in cholera toxin [36] (Fig. 4C) as well as in other OB-fold

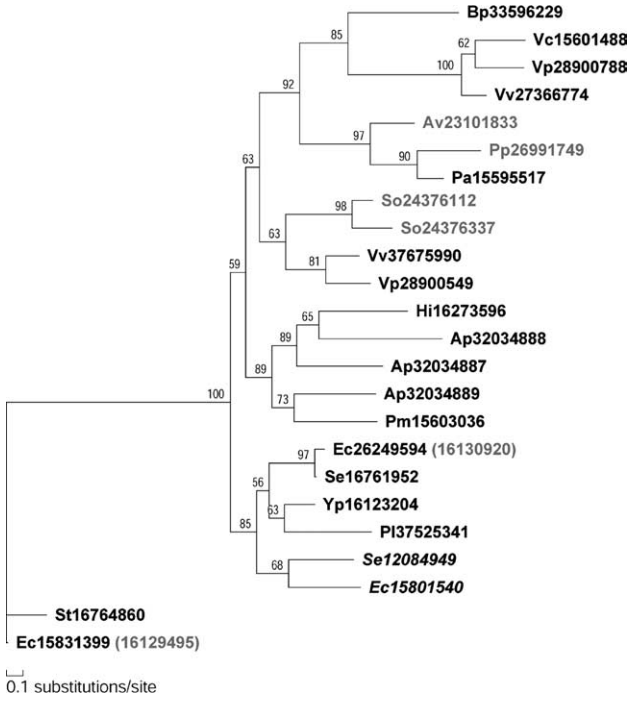


Fig. 3. Phylogenetic tree for BOF sequences containing complete OB-fold domain. Sequences are labeled according to species abbreviation and gi number. Labels are shown in black for pathogenic bacterial strain, in gray for non-pathogenic bacterial strain, and in italics for inclusion in prophage inserts. Labels in parentheses are from identical sequences found in genomes from different strains. Local bootstrap values are indicated (as %) at internal tree nodes.

families. The hypothesis of recent sequence divergence of BOF and anticodon-binding families is also supported by weak PSI-Blast hits over the entire sequence length with the absence of false positives. All these findings demonstrate that BOF family is a distant outlier in the nucleic acid-binding superfamily.

### 3.4. Potential binding site

Despite the convincing conservation of unique structural features between the BOF family and other nucleic acid-binding domains, establishing a precise binding mode for the BOF OB-fold using theoretical methods remains difficult. Residues that contribute to the typical OB-fold ligand-binding face often reside in loop regions that differ considerably between different OB-fold families. The length differences in these loop regions between the BOF family and the closest related nucleic acid-binding families prevent precise modeling of the potential binding site. Fortunately, a significant portion of residues that belong to the reliably predicted core also contributes to the ligand-binding face of OB-fold domains. Fig. 4A and C illustrate two different OB-fold proteins bound to their respective ligands. Side-chains residing within the OB-fold core that could contribute to an analogous binding site in *E. coli* protein ygiW precursor (gi|26249594) include Asp 107, Asp 109, Glu 129, Asp 131, Glu 138 and Asp 140 (Fig. 4B). These residues are relatively conserved within the entire BOF family and are highly conserved among different phylogenetically defined subfamilies, supporting a potential role in ligand binding.

As compared to other nucleic acid-binding OB-folds, the predicted BOF family ligand-binding site contains conservatively replaced polar residues (Fig. 1, violet triangles). However, the nature of these residues may not be consistent with binding nucleotides. The BOF family lacks a highly conserved nucleotide-binding domain phenylalanine residue (Phe 35, 1c0a) that provides stacking interactions and glutamine (Gln 46, 1c0a) that forms hydrogen bonds with the nucleotide base, in addition to lacking a conserved positive charge (Arg 28, 1c0a) that interacts with the sugar backbone. With the exception of a single BOF sub-family, a majority of the predicted ligand-binding residues maintain a negative charge, suggesting a potential interaction with a positively charged ligand. However, the precise nature of this ligand remains unclear. As hypothesized for the distantly related OB-fold containing bacterial enterotoxins, the loss of nucleotide-binding function

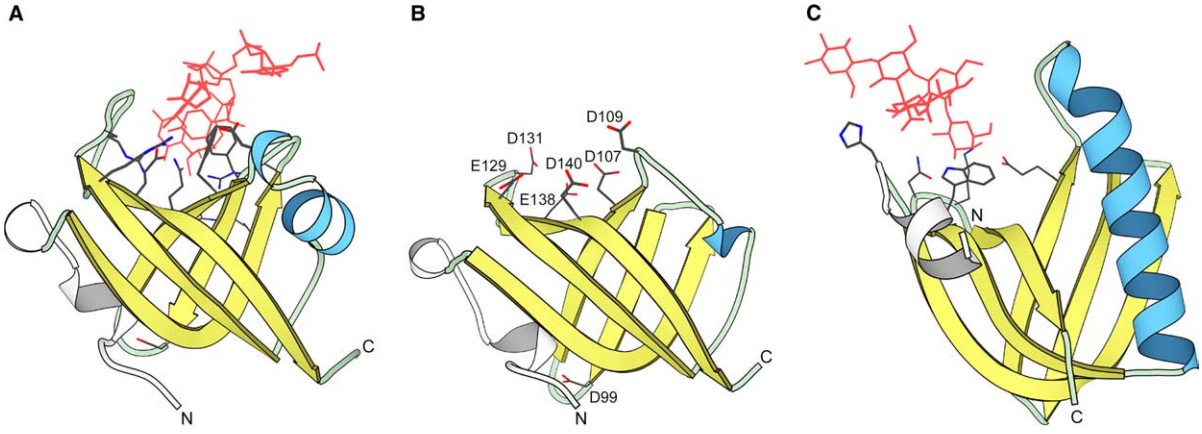


Fig. 4. Comparison of OB-fold binding sites. (A) The N-terminal OB-fold anticodon-binding domain from *E. coli* AspRS complexed with Asp-tRNA (PDB code: 1c0a). Only four nucleotides of the Asp-tRNA molecule (red) that fall within 3.6 Å of the anticodon-binding domain and side chains within 3.6 Å of the ligand (excluding residues within inserted loop regions with respect to BOF family alignment) are displayed. (B) Fold recognition model for OB-fold domain of *E. coli* protein ygiW precursor (gi|26249594) based on *E. coli* AspRS structure (PDB code: 1c0a). Side chains facing proposed binding site are shown. (C) One OB-fold monomer from the bacterial enterotoxin cholera toxin B pentamer (PDB code: 3chb) complexed with pentasaccharide (red). Displayed side chains were selected in the same manner as in panel A. Aspartic acids stabilizing unique conformation of the N-terminal region (white) in A and B are also shown. In all panels displayed side chains are colored according to atom type: nitrogen (blue), oxygen (red), and carbon (gray). Color shading of secondary structure elements corresponds to that in Fig. 1.

and the rapid evolution of the BOF ligand-binding site may be associated with the presence of BOF proteins in mobile genetic elements and their potential role in bacterial pathogenicity. In accordance with this hypothesis, the BOF family sequences share similar phylogenetic distributions (sequences found in mobile genetic elements) and cellular distributions (presence of a signal peptide) with the OB-fold domains of the bacterial enterotoxin superfamily.

While this prediction provides a general structure–functional annotation for previously uncharacterized proteins, their substrate specificity, detailed mechanism of action and exact role in bacterial organisms need to be elucidated through further biochemical experiments.

*Acknowledgements:* This work was supported in part by the NIH grant GM67165 to N.V.G.

## References

[1] Arcus, V. (2002) Curr. Opin. Struct. Biol. 12, 794–801.
[2] Agrawal, V. and Kishan, K.V. (2003) Curr. Protein Pept. Sci. 4, 195–206.
[3] Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) J. Mol. Biol. 247, 536–540.
[4] Murzin, A.G. (1993) EMBO J. 12, 861–867.
[5] Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E.L. (2002) Nucleic Acids Res. 30, 276–280.
[6] Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D. and Koonin, E.V. (2001) Nucleic Acids Res. 29, 22–28.
[7] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[8] Ginalski, K., Pas, J., Wyrwicz, L.S., von Grotthuss, M., Bujnicki, J.M. and Rychlewski, L. (2003) Nucleic Acids Res. 31, 3804–3807.
[9] Ginalski, K., von Grotthuss, M., Grishin, N.V. and Rychlewski, L. (2004) Nucleic Acids Res. 32, in press.
[10] Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D. and Zardecki, C. (2002) Acta. Crystallogr. D: Biol. Crystallogr. 58, 899–907.
[11] Schultz, J., Milpetz, F., Bork, P. and Ponting, C.P. (1998) Proc. Natl. Acad. Sci. USA 95, 5857–5864.
[12] Sonnhammer, E.L., von Heijne, G. and Krogh, A. (1998) Proc. Int. Conf. Intell. Syst. Mol. Biol. 6, 175–182.
[13] Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Protein Eng. 10, 1–6.
[14] Wootton, J.C. (1994) Comput. Chem. 18, 269–285.
[15] Lupas, A., Van Dyke, M. and Stock, J. (1991) Science 252, 1162–1164.
[16] Mott, R. (2000) J. Mol. Biol. 300, 649–659.
[17] Bujnicki, J.M., Elofsson, A., Fischer, D. and Rychlewski, L. (2001) Bioinformatics 17, 750–751.
[18] Ginalski, K., Elofsson, A., Fischer, D. and Rychlewski, L. (2003) Bioinformatics 19, 1015–1018.
[19] Karplus, K., Karchin, R., Barrett, C., Tu, S., Cline, M., Diekhans, M., Grate, L., Casper, J. and Hughey, R. (2001) Proteins Suppl. 5, 86–91.
[20] Rychlewski, L., Jaroszewski, L., Li, W. and Godzik, A. (2000) Protein Sci. 9, 232–241.
[21] Jones, D.T. (1999) J. Mol. Biol. 287, 797–815.
[22] Fischer, D. (2000) Pac. Symp. Biocomput., 119–130.
[23] Xu, J., Li, M., Lin, G., Kim, D. and Xu, Y. (2003) Pac. Symp. Biocomput., 264–275.
[24] Shi, J., Blundell, T.L. and Mizuguchi, K. (2001) J. Mol. Biol. 310, 243–257.
[25] Kelley, L.A., MacCallum, R.M. and Sternberg, M.J. (2000) J. Mol. Biol. 299, 499–520.
[26] Sali, A. and Blundell, T.L. (1993) J. Mol. Biol. 234, 779–815.
[27] Eiler, S., Dock-Bregeon, A., Moulinier, L., Thierry, J.C. and Moras, D. (1999) EMBO J. 18, 6532–6541.
[28] Bonneau, R., Tsai, J., Ruczinski, I., Chivian, D., Rohl, C., Strauss, C.E. and Baker, D. (2001) Proteins Suppl. 5, 119–126.
[29] Pei, J., Sadreyev, R. and Grishin, N.V. (2003) Bioinformatics 19, 427–428.
[30] Ginalski, K. and Rychlewski, L. (2003) Proteins 53 (Suppl 6), 410–417.
[31] Schmitt, E., Moulinier, L., Fujiwara, S., Imanaka, T., Thierry, J.C. and Moras, D. (1998) EMBO J. 17, 5227–5237.
[32] Briand, C., Poterszman, A., Eiler, S., Webster, G., Thierry, J. and Moras, D. (2000) J. Mol. Biol. 299, 1051–1060.
[33] Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C. and Moras, D. (1991) Science 252, 1682–1689.
[34] Onesti, S., Desogus, G., Brevet, A., Chen, J., Plateau, P., Blanquet, S. and Brick, P. (2000) Biochemistry 39, 12853–12861.
[35] Bochkarev, A., Bochkareva, E., Frappier, L. and Edwards, A.M. (1999) EMBO J. 18, 4498–4504.
[36] Merritt, E.A., Kuhn, P., Sarfaty, S., Erbe, J.L., Holmes, R.K. and Hol, W.G. (1998) J. Mol. Biol. 282, 1043–1059.
[37] Adachi, J. and Hasegawa, M. (1992) In: Computer Science Monographs, Institute of Statistical Mathematics, Tokyo.
[38] Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) Comput. Appl. Biosci. 8, 275–282.
[39] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.
[40] Kishino, H., Miyata, T. and Hasegawa, M. (1990) J. Mol. Evol. 30, 151–160.
[41] Kinch, L.N., Wrabl, J.O., Krishna, S.S., Majumdar, I., Sadreyev, R.I., Qi, Y., Pei, J., Cheng, H. and Grishin, N.V. (2003) Proteins 53 (Suppl 6), 395–409.
[42] Ginalski, K. and Rychlewski, L. (2003) Nucleic Acids Res. 31, 3291–3292.
[43] Haneda, T., Okada, N., Nakazawa, N., Kawakami, T. and Danbara, H. (2001) Infect. Immun. 69, 2612–2620.
[44] Plunkett III, G., Rose, D.J., Durfee, T.J. and Blattner, F.R. (1999) J. Bacteriol. 181, 1767–1778.
[45] Hacker, J., Hentschel, U. and Dobrindt, U. (2003) Science 301, 790–793.
[46] Marchler-Bauer, A., Anderson, J.B., DeWeese-Scott, C., Fedorova, N.D., Geer, L.Y., He, S., Hurwitz, D.I., Jackson, J.D., Jacobs, A.R., Lanczycki, C.J., Liebert, C.A., Liu, C., Madej, T., Marchler, G.H., Mazumder, R., Nikolskaya, A.N., Panchenko, A.R., Rao, B.S., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Vasudevan, S., Wang, Y., Yamashita, R.A., Yin, J.J. and Bryant, S.H. (2003) Nucleic Acids Res. 31, 383–387.