**Structure of human BPI (bactericidal/permeability-increasing protein) and implications for related proteins**

L.J. Beamer

Department of Biochemistry, 117 Schweitzer Hall, University of Missouri, Columbia, MO 65211, U.S.A.

**Abstract**

Human bactericidal/permeability-increasing protein (BPI) belongs to a family of mammalian lipopolysaccharide-binding and lipid transport proteins. Recent sequence database searches indicate that several other protein families, including the palate, lung and nasal epithelial clone (PLUNC), parotid secretory protein (PSP) and BPI-like proteins, are likely to share the BPI fold, which was determined through X-ray crystallographic studies. As the single representative of its fold family of known structure, the three-dimensional model of BPI suggests structural features that are likely to be conserved across this large and varied group of proteins.

**Lipid transfer protein (LTP)/lipopolysaccharide (LPS)-binding family**

In humans and other mammals, a family of lipid-interactive proteins has been defined that includes four well-characterized members: bactericidal/permeability-increasing protein (BPI), LPS-binding protein (LBP), cholesteryl ester transfer protein (CETP) and phospholipid transfer protein (PLTP). Two of these proteins, BPI and LBP, function in the innate immune response to bacterial infection and modulate the inflammatory cascade triggered by the bacterial glycolipid, LPS [1,2]. The other two proteins, CETP and PLTP, function in the transport of lipids between lipoprotein particles in the bloodstream, and may play a role in the pathogenesis of arteriosclerosis [3]. LBP, CETP and PLTP are produced by the liver and circulate in the blood, while BPI is abundant in the primary granules of polymorphonuclear neutrophils. Although the overall amino acid identities between sequence pairs in this family are somewhat limited (less than 20% for some pairs), their alignments extend over more than 400 residues with few gaps or deletions, and include two highly conserved cysteine residues. These sequence alignments clearly define a related protein family that has been termed the LTP/LPS-binding family. Although the biological roles of the family members are distinct and varied, they appear to share a common ability to interact with membranes, and to bind lipids and other apolar ligands. They also lack any apparent enzymic activity.

Concurrent with the rapid increase in available genomic sequence data in recent years, it has become apparent that the LTP/LPS-binding proteins are part of a much larger family that extends across multiple eukaryotic lineages, and also includes a number of distantly related mammalian homologues. To date, related sequences have been identified (via PSI-BLAST [4] searches or similar methods) in non-mammalian vertebrates including birds and fish, in invertebrates such as *Caenorhabditis elegans*, and in at least one plant species, *Arabidopsis thaliana* [5,6]. No related prokaryotic sequences have yet been noted. In addition, a growing number of mammalian homologues have been identified that are expressed in cell and tissue types distinct from those expressing the four members of the LTP/LPS-binding family. They include (but are not limited to) members of the overlapping families known as palate, lung and nasal epithelial clone (PLUNC) [7], parotid secretory protein (PSP) [8], and the ‘BPI-like’ proteins [9]. For simplicity, we will refer to this larger, extended family as the BPI/PLUNC/PSP superfamily. Many of these sequences are similar in size to the members of the LTP/LPS-binding family ($\approx$450 amino acids); however, some (including the short or SPLUNC proteins) are approximately one-half of that size. Despite considerable diversity, nearly all of these sequences share the two conserved cysteines found in the LTP/LPS-binding family. No clear functional roles have yet been established for the distantly related mammalian homologues of the LTP/LPS-binding family.

Three-dimensional structural information on proteins is an invaluable tool for the interpretation of functional data, and can greatly facilitate the design of future experiments. Despite widespread interest in the biological roles of proteins in the BPI/PLUNC/PSP superfamily, structural information is currently limited to a single example: that of human BPI (Figure 1) [10]. An overview of the BPI structure, with emphasis on features that are likely to be conserved throughout the protein superfamily, is given in the following sections. A discussion of the difficulties faced in homology modelling between distantly related structural templates and
Three-dimensional structure of BPI

The structure of full-length (456 residues) human BPI was determined by X-ray crystallography [10] and refined to 1.7 Å resolution [11]. The BPI molecule has a highly elongated, ‘boomerang’ shape, consisting of two structurally similar domains of nearly equal size (Figure 1A). These two domains exhibit a novel protein fold, with similar secondary structural elements and topology, giving the molecule pseudo 2-fold symmetry. A barrel-like structural unit is found at each end of the protein, composed of a highly twisted, antiparallel \( \beta \)-sheet and two \( \alpha \)-helices (Figure 1B). The two barrels are bridged by a \( \beta \)-sheet in the centre of the protein that contains residues from both the N- and C-terminal domains. In the N-terminal barrel, the two cysteine residues conserved in the LTP/LPS-binding family (residues 135 and 175 in BPI) form a disulphide bond that anchors the edge of the \( \beta \)-sheet to one of the \( \alpha \)-helices (Figure 1C).

The BPI ‘barrel’ appears to be an independent folding module. This has been best demonstrated in the case of the N-terminal barrel of BPI, which the crystal structure shows extends from residues 10 to 193 [10]. Several N-terminal constructs of BPI that include these residues have been well characterized, and shown to retain their biological activity (for a review, see [2]). In addition, deletion analyses have demonstrated that residues 1–12 are unnecessary for activity, while deletion of residues extending further into the barrel units (from either the N- or the C-terminus) results in reduced expression and/or inactive protein [12]. The conserved disulphide bond is also critical for expression of active N-terminal BPI fragments [13], explaining its conservation in the family. Corresponding studies of the activity and stability of the C-terminal barrel of BPI (residues 260–421), which lacks a disulphide bond, have not yet been performed. However, functional N-terminal fragments of LBP have also been characterized [14,15].

Other experiments have provided insight into the modular nature of the barrel unit. For example, chimaeras of BPI and LBP have been created, in which the first \( \approx 200 \) residues of BPI were combined the last \( \approx 250 \) residues of LBP (or vice versa), which essentially corresponds to exchanging the N-terminal barrel of each protein into the remaining structural framework of the partner. The successful expression
and characterization of these chimaeric proteins has permitted important functional comparisons of the N- and C-terminal domains of BPI and LBP [16,17]. The construction of analogous chimaeras may prove useful for elucidating structure–function relationships for other members of the BPI/PLUNC/PSP family. However, such experiments are most likely to succeed when domains are exchanged between closely related sequences: BPI and LBP are ≈45% identical. Biochemical confirmation of the structural borders of the barrel regions, such as through limited proteolysis, may help to increase the likelihood of success of such experiments.

Some members of the extended BPI family appear to exist as a stably folded, single barrel. Database searches reveal a number of shorter, related sequences of approx. 200 residues that have identity with either the N- or the C-terminal half of BPI [5,7]. These smaller, ‘single-barreled’ proteins may be similar to an ancestral protein of the family, whose gene duplication presumably brought about the larger, pseudo-symmetrical molecules such as BPI. For structural studies, these single-barrel sequences are potentially of great interest, since they may be easier to express in bacterial systems than the more complicated fold seen in the two-domain proteins. In addition, their smaller size should make them accessible to analysis by NMR spectroscopy.

A wide variety of amino acid sequences can apparently be utilized to form the BPI barrel unit. This is clearly illustrated by a comparison of the N- and C-terminal domains of BPI [11], which are structurally quite similar, but lack detectable sequence similarity (<15% identity based on the structural superimposition, and not identifiable by PSI-BLAST searches). Detectable sequence relationships between the putative N- and C-terminal barrel regions are not found in other members of the BPI/PLUNC/PSP family either, further demonstrating the wide range of sequences that can fold successfully into the barrel unit. Thus, despite the large sequence variation between members of the superfamily, precedence for overall structural similarity between BPI and the distantly related protein sequences is clear. Comparison of BPI’s N- and C-terminal barrels also reveals the location of greatest structural variation: loops at the tips of the barrels, furthest from the interface with the central β-sheet, which is presumably conserved for structural reasons [10]. Therefore it is likely that the structures of the proteins in the extended family will also differ from each other most in the tip regions of the barrels.

Two lipid-binding pockets
The crystal structure of BPI reveals two phospholipid molecules bound in deep, apolar pockets of the protein [10]. Each domain contains one binding pocket, maintaining the pseudo 2-fold symmetry seen in the protein structure. The openings of the two pockets are found on the same side of the protein, and extend deep into the interior of the protein. The phospholipids bind with their acyl carbon chains buried in the pockets, while the charged head groups are exposed to solvent near the surface of the protein. The biological significance of the bound phospholipids for BPI remains in question, although sequence conservation of residues involved in forming the pockets suggests that they may have a functional role [11,18].

The functional implications of the lipid-binding pockets discovered in the BPI structure are clearer for other members of the LTP/LPS-binding family. Both CETP and PLTP have well characterized phospholipid transfer activity, which is critical for their lipid transfer activity in the blood. The BPI-phospholipid complex, with its two largely buried ligands, provides a model for the sequestration and transfer of apolar ligands through an aqueous environment. Because most of the contact surface between the protein and bound ligands is between apolar atoms, and therefore geometrically non-specific, it appears that the pockets could be adapted to bind a variety of apolar ligands through variation in the identity of the side chains surrounding the ligand. Thus it seems reasonable to expect that other members of the extended BPI/PLUNC/PSP superfamily may also function as ligand-binding proteins.

Modelling of related proteins
As the only representative of its fold family of known structure, BPI is of considerable interest as a template for homology modelling. A number of programs and automated web servers for modelling have been developed, making this technique widely accessible to a general scientific audience. The resulting models often appear to be chemically reasonable, exhibiting good protein geometry and energetic profiles. However, as demonstrated in the recent Fourth Meeting on the Critical Assessment of Methods of Protein Structure Prediction (CASP4), even state-of-the-art modelling procedures are prone to frequent errors [19]. This is particularly true when the amino acid identity between the template structure and the sequence to be modelled is less than 30%. However, in lieu of other structural information, homology models can be helpful for the design of future experiments.

Since all the members of the BPI/PLUNC/PSP family have less than 30% sequence identity with BPI (with the exception of LBP), genuine differences between the structure of the template and the structure of the target protein, such as the differences seen in the tips of the barrel regions of BPI, are likely to exist. This makes it difficult to predict the precise role of a given residue. Nonetheless, homology models may be useful in a broader sense, for example in identifying the general location of a residue, and the structural element to which it belongs (e.g. N- versus C-terminal barrel, or central β-sheet). In the case of human PLTP, which shares 26% sequence identity with BPI, homology modelling was used successfully as a tool to identify residues in the two lipid-binding pockets [20]. Importantly, the roles of these residues were then confirmed by site-directed mutagenesis. Similar experiments to test homology models should be possible with other members of the BPI/PLUNC/PSP family. These models may be useful as tools to be used in concert with...
independent biochemical and structural characterization of proteins of the BPI/PLUNC/PSP superfamily.

**Future directions**

Obtaining structural information on more members of the BPI/PLUNC/PSP superfamily should be a top priority. The increasing number of related sequences, often available from several different organisms, should facilitate the production of proteins amenable to crystallization. The expression of soluble, active protein for the larger family members (>400 residues) has not been successful in prokaryotic systems; this may be more tractable for the smaller, single-domain relatives. Traditional biochemical techniques, such as limited proteolysis, should be helpful in defining fragments that correspond to the barrel folding module, and which are likely to be compact domains suitable for structural analysis. In the meantime, site-directed mutagenesis experiments, perhaps in tandem with modelling efforts, may be able to provide important clues for functional characterization of the family members. Together, these studies will not only illuminate the protein of choice, but also increase structural understanding of all members of the BP/PLUNC/PSP superfamily. The extensive sequence divergence in this widespread and intriguing protein family undoubtedly indicates that many structural surprises await future investigation.

This material is based on work supported by the National Science Foundation under grant no. MCB 9974912.

**References**


Received 21 January 2003