FOR THE RECORD

Detecting distant relatives of mammalian LPS-binding and lipid transport proteins

LESA J. BEAMER,¹ DANIEL FISCHER,^{2,3} AND DAVID EISENBERG²

¹Biochemistry Department, University of Missouri-Columbia, Columbia, Missouri 65211

²UCLA/DOE Laboratory of Structural Biology, Molecular Biology Institute, UCLA, P.O. Box 951570, Los Angeles, California 90095-1570

(RECEIVED February 23, 1998; ACCEPTED March 15, 1998)

Abstract: In mammals, a family of four lipid binding proteins has been previously defined that includes two lipopolysaccharide binding proteins and two lipid transfer proteins. The first member of this family to have its three-dimensional structure determined is bactericidal/permeability-increasing protein (BPI). Using both the sequence and structure of BPI, along with recently developed sequence-sequence and sequence-structure similarity search methods, we have identified 13 distant members of the family in a diverse set of eukaryotes, including rat, chicken, *Caenorhabditis elegans*, and *Biomphalaria galbrata*. Although the sequence similarity between these 13 new members and any of the 4 original members of the BPI family is well below the “twilight zone,” their high sequence-structure compatibility with BPI indicates they are likely to share its fold. These findings broaden the BPI family to include a member found in retina and brain, and suggest that a primitive member may have contained only one of the two similar domains of BPI.

Keywords: lipid transport proteins; LPS binding; sequence-structure compatibility

A family of four mammalian lipid-binding proteins has been previously characterized. It includes bactericidal/permeability-increasing protein (BPI), lipopolysaccharide-binding protein (LBP), cholesteryl ester transport protein (CETP), and phospholipid lipid transport protein (PLTP). BPI and LBP are involved in host response to Gram-negative bacterial infections and bind with high affinity to a complex bacterial glycolipid called lipopolysaccharide (LPS) (Schumann et al., 1990; Elsbach & Weiss, 1993). CETP and PLTP are serum proteins, which are involved in the transport and exchange of lipids between lipoprotein particles (Tall, 1995). Together, these proteins are called the lipid transport protein (LTP)/LPS-binding family.

Recently, the first three-dimensional structure of a member of the LTP/LPS-binding family, that of human BPI, was determined in our laboratory (Beamer et al., 1997) by X-ray crystallography (Fig. 1A). In addition to its high affinity binding to LPS, BPI is potently bactericidal to Gram-negative bacteria, and has potential therapeutic applications (Elsbach & Weiss, 1995). Full-length (456 amino acids) human BPI was found to have an unusual boomerang shape (approximately $35 \times 35 \times 135$ Å) and two structurally similar domains of about 200 residues (called N- and C-terminal). Although these two domains have no significant sequence identity, they are structurally quite similar (root-mean-square deviation of 3.0 Å for 169 Cα pairs). This two-domain protein, which can be thought of as a covalently linked pseudo-dimer, presumably evolved from a gene duplication event of a single domain, monomeric ancestor. In each domain, an apolar pocket binds a molecule of phosphatidylcholine. This bound lipid may be related to BPI's known LPS-binding activity (Beamer et al., 1997).

In total, 12 amino acid sequences are known for the four members of the LTP/LPS-binding family, all from mammals. The sequence identities between members of this family vary, but are as low as 17%, depending on the protein and organism compared. Although in general it is difficult to reliably identify relationships between proteins when the sequence identity falls below 25%, the relationship between the four family members (BPI, LBP, CETP, PLTP) is clear, and can be detected using sequence similarity search programs such as BLAST (Altschul et al., 1990). Multiple sequence alignments show regions of similarity extending over the full length of the proteins (about 450 amino acids) and all members share a single conserved disulfide bond in the N-terminal half of the protein (between Cys 135 and 175 in human BPI). Although some variations occur, all four members of the family are expected to share the overall BPI fold.

By applying sensitive sequence-sequence and sequence-structure similarity searches that have been recently developed, we have identified a set of proteins with even lower amino acid sequence similarities, which are also likely to share the BPI fold. We began our search by using a new version of BLAST, called PSI-BLAST (Altschul et al., 1997), which iteratively searches sequence databases using a position-specific score matrix. The human BPI sequence (gi|115111) without its leader peptide was submitted as the

Reprint requests to: David Eisenberg, UCLA/DOE Laboratory of Structural Biology, Molecular Biology Institute, UCLA, P.O. Box 951570, Los Angeles, California 90095-1570; e-mail: david@mbi.ucla.edu.

³Present address: Faculty of Natural Sciences, Department of Math and Computer Science, Ben Gurion University, Beer-Sheva 84015, Israel.

Table 1. A summary of the high scoring sequences (HSS) obtained from PSI-BLAST (Altschul et al., 1997) and SDP (Fischer & Eisenberg, 1996)^a

Source	Name/PID	NH ₂ -terminal domain	COOH-terminal domain	PSI-BLAST	SDP	%identity
Human	BPI/gil115111		456	e-174	178.8	89
Human	CETP/gil180269		473	e-124	23.2	21
Rat	gil57734		473	3e-81	25.8	17
Rat	gil57732		470	5e-77	28.3	19
Biomphalaria galbrata	gil1323800		336	e-61	14.6 (19.4)	16
Gallus gallus	tenp/gil2599572		440	4e-43	15.9	14
<i>C. elegans</i>	gil1070040		487	5e-88	17.4	17
	gil2497618		464	3e-88	18.8	15
	gil2462151		486	2e-82	20.4	14
	gil1255822		617	3e-82	20.3	14
	gil1086641		412	2e-50	13.0	14
	gil1255435		284	2e-50	5.0 (19.5)	14
	gil458988		554	3e-46	15.4	13
	gil2253128		824	2e-42	10.5	18
	gil1523957		576	e-40	15.7	15

^aThe source, name (if applicable), and Genbank PID is shown for each sequence at the right. The schematic alignments indicate approximate positions of correspondence to BPI and overall length of the aligned sequences. The gray bars indicate the part(s) of the sequence that were aligned to human BPI by PSI-BLAST. Unaligned regions of the sequence are indicated by a black line. Human BPI and CETP are shown at the top for comparison. On the right are the scores obtained from PSI-BLAST and SDP, followed by the percent sequence identity of each protein with human BPI. For the two shortest sequences, SDP scores using the local alignment algorithm are shown in parentheses. This option compensates for the marked difference in length between BPI and these two proteins and more accurately reflects their sequence-structure compatibility. The sequence identity of BPI with itself is only 89% because the SEG option of BLAST found 48 residues of BPI to have low sequence complexity. These regions were "filtered" and not used in the search or taken into account when calculating the sequence identity. PSI-BLAST scores smaller than 0.00001 are considered to be very significant. SDP scores above 7.0 indicate a strong sequence-structure compatibility.

are from eukaryotic organisms. Six had been previously noted to have weak similarity to either BPI, LBP, or CETP. This includes two proteins from rat olfactory mucosa, and four proteins from *Caenorhabditis elegans*. However, seven other HSS were found, including one from *Gallus gallus* (chicken), and another from *Biomphalaria galbrata*, a blood fluke in the mollusk family. In addition, five more sequences from *C. elegans* were identified, making a total of nine from that organism. It is possible that this reflects the relatively large proportion of the *C. elegans* genome, which has been sequenced compared to other eukaryotes.

The HSS all have amino acid identities of <20% with BPI (Table 1). Except for the two rat proteins, which are 37% identical, the highest pairwise identity between any of the sequences in Table 1 is 22%, indicating their relatively distant relationships, even within the same organism. The HSS range in size from 284 to 834 amino acids. Based on alignments with the BPI sequence and structure, most of these putative proteins contain both the N- and C-terminal domains of BPI (residues 1 to 230 and 250 to 456). Multiple sequence alignments indicate that 5 of the 13 HSS may share the conserved disulfide bond in the N-terminal domain (data

not shown). The *B. galbrata* sequence (gi|1323800; 336 amino acids) aligns primarily with BPI's N-terminal domain, while one *C. elegans* sequence (gi|1255435; 284 amino acids) aligns with BPI's C-terminal domain. It is possible that one or both of these proteins represent a monomeric version of the covalently linked, two-domain BPI "dimer," potentially pre-dating the present-day two domain protein. Three of the *C. elegans* proteins (gi|458988, gi|1523957, and gi|2253128) have one or more insertions of at least 50 amino acids relative to BPI.

Very little functional characterization has been performed on the HSS found in our search. To our knowledge, none of the *C. elegans* proteins has been biochemically characterized. The two rat olfactory proteins have been called probable ligand binding proteins based on the weak sequence similarity to BPI. The protein from *B. galbrata* is known only to be a developmentally regulated albumen gland gene. The tenp protein (Yan & Wang, 1998) from *G. gallus*, however, was isolated as a transiently expressed gene in neural precursor cells in retina and brain, and has been proposed to function in the transition to cell differentiation in neurogenesis. After expression in chicken embryonic fibroblast cells, tenp was immu-

not detected in membrane fractions, and two putative transmembrane helices were predicted from its sequence (Yan & Wang, 1998). Although weak sequence similarity between tenp and the two rat olfactory proteins was noted, it was not extended to include members of the LTP/LPS-binding family. Given the scores of tenp for both the PSI-BLAST and SDP searches, it appears likely that tenp shares the BPI fold and is therefore not an integral membrane protein. However, as has been proposed for BPI, tenp may be membrane associated (Weiss & Olson, 1987). A sequence-structure alignment between tenp and human BPI obtained from SDP is shown in Figure 1B.

Based on our sequence-sequence and sequence-structure similarity analyses, we expect the HSS to share the overall protein fold seen in the BPI crystal structure. It is possible some of the HSS bind to lipids or other hydrophobic ligands. However, it is not uncommon to find proteins with low levels of sequence identity which adopt the same fold and have apparently unrelated functions (Orengo et al., 1993; Fischer et al., 1995). Our results strengthen the previously noted weak sequence similarity between six of the HSS and members of the LTP/LPS-binding family. In addition, we have identified seven new sequences that are likely to share the BPI fold. We, therefore, propose that this protein fold will be found in a wide variety of eukaryotic organisms. Classification of the HSS as relatives of the LTP/LPS-binding family may aid in functional characterization of these proteins, and alignments with the BPI structure should help provide a framework for future structure-function analyses.

Acknowledgments: We thank R.-T. Yan and S.-Z. Wang for sending us a preprint of their manuscript, and the National Institutes of Health and Department of Energy for support.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Bio* 215:403–410.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein data-base search programs. *Nuc Acid Res* 25:3389–3402.
- Beamer LJ, Carroll SF, Eisenberg D. 1997. Crystal structure of human BPI and two bound phospholipids at 2.4 Å resolution. *Science* 276:1861–1864.
- Elsbach P, Weiss J. 1993. Bactericidal/permeability increasing protein and host defense against Gram-negative bacteria and endotoxin. *Curr Opin Immun* 5:103–107.
- Elsbach P, Weiss J. 1995. Prospects for use of recombinant BPI in the treatment of Gram-negative bacterial infections. *Infect Agent Dis* 4:102–109.
- Fischer D, Eisenberg D. 1996. Protein fold recognition using sequence-derived predictions. *Protein Sci* 5:947–955.
- Fischer D, Eisenberg D. 1997. Assigning folds to the proteins encoded by the genome of *Mycoplasma genitalium*. *Proc Natl Acad Sci USA* 94:11929–11934.
- Fischer D, Tsai CJ, Nussinov R. 1995. A 3D sequence-independent representation of the protein data bank. *Protein Eng* 8:981–997.
- Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
- Kraulis P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* 24:946–950.
- Orengo CA, Flores TP, Jones DT, Taylor WR. 1993. Recurring structural motifs in proteins with different functions. *Curr Biol* 6:131–139.
- Rost B, Sander C. 1993. Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol* 232:584–599.
- Schumann RR, Leong SR, Flaggs GW, Gray PW, Wright SD, Mathison JC, Tobias PS, Ulevitch RJ. 1990. Structure and function of lipopolysaccharide binding protein. *Science* 249:1429–1430.
- Tall A. 1995. Plasma lipid transfer proteins. *Annu Rev Biochem* 64:235–257.
- Weiss J, Olson I. 1987. Cellular and subcellular localization of the bactericidal/permeability-increasing protein of neutrophils. *Blood* 69:652–659.
- Yan R-T, Wang S-Z. 1998. Identification and characterization of tenp, a gene transiently expressed before overt cell differentiation during neurogenesis. *J Neurobiol* 34:319–328.