PLUNC: A novel family of candidate host defence proteins expressed in the upper airways and nasopharynx

Colin D. Bingle1,* and C. Jeremy Craven2

1Respiratory Medicine Unit, Division of Genomic Medicine, University of Sheffield Medical School, Sheffield, S10 2RX, UK and 2Krebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, S10 2TN, UK

Received January 3, 2002; Revised and Accepted February 8, 2002

The upper respiratory tract, including the nasal and oral cavities, is the major route of entry of pathogens into the body, and early recognition of bacterial products in this region is critical for host defence. A well-established family of four proteins involved in this process are bactericidal/permeability-increasing protein (BPI) and lipopolysaccharide-binding protein (LBP), which are central to the host defence against bacteria, and cholesteryl ester transfer protein (CETP) and phospholipid transfer protein (PLTP), which have also been implicated in this response. In this paper, we demonstrate the existence of a related family of seven human proteins, which we designate PLUNC proteins. The PLUNC proteins are encoded by adjacent genes found within a 300kb region of chromosome 20, suggesting that they may be under transcriptional control of shared genomic elements, and expression data shows that these proteins are found in overlapping regions of the pulmonary, nasopharyngeal and oral epithelium, sites where the previously described BPI family members are not expressed. Whereas the BPI family are predicted to share very closely similar three-dimensional structures, the PLUNC family is predicted to have much greater variability in the N-terminal domain, corresponding to the active domain of BPI, thus creating the notion of a BPI/PLUNC structural superfamily. We suggest that members of the PLUNC family may function in the innate immune response in regions of the mouth, nose and lungs, which are sites of significant bacterial exposure.

INTRODUCTION

The early recognition of bacterial products is critical for our survival. Lipopolysaccharide (LPS) constituents of bacterial cell walls, such as LPS of Gram-negative bacteria, lipoteichoic acids of Gram-positive bacteria and lipoarabinomannans of mycobacteria, have profound clinical effects. In sepsis, many cell types, including neutrophils and macrophages, respond to these constituents by releasing potent inflammatory mediators or becoming activated to destroy invading bacteria. When bacterial products are present in relatively large amounts, responses become sufficiently devastating to trigger septic shock, with its concomitant high mortality. The best studied of these bacterial constituents is LPS. For LPS to be recognized by the LPS receptor complexes on the surfaces of cells such as macrophages, it must be bound to a secreted protein that acts as a presentation molecule for the receptor (1,2). In humans, two proteins critical to the mediation of signals from LPS are bactericidal/permeability-increasing protein (BPI) and LPS-binding protein (LBP). These proteins are structurally related and can be considered to have antagonistic functions (1,3). LBP binds LPS and enhances cellular responses to LPS, whilst BPI binds LPS but reduces cellular responses (4). Thus, a critical balance is maintained in our response to bacteria. In addition to this role of neutralizing LPS, BPI also has a more directly bactericidal action (3). Two other related proteins, cholesteryl ester transfer protein (CETP) and phospholipid transfer protein (PLTP), have also been implicated in mediating LPS responses (5). Experiments using mice with a disruption of the LBP gene have yielded conflicting results (6,7), raising doubts about the role of LBP as the only mediator of LPS-induced cell activation. One interpretation of these results is that other LPS-interacting proteins may be present in certain organ systems, which may mediate signalling and the inflammatory response. For example, the existence of such proteins within the respiratory tract would be advantageous, since LBP is a circulating factor whereas a more local source of LPS-sensing molecules within the respiratory tract would...
allow a more immediate response to an inhaled pathogenic signal.

As part of a strategy to identify genes expressed in the upper airways, we recently described the characterization of the novel human gene product PLUNC (8). Human PLUNC is the orthologue of a murine protein expressed in the embryonic palate, nasal epithelium and trachea (9), and appears to be a good candidate marker gene for tracheal epithelium (10).

We now demonstrate that PLUNC is a member of a novel family of seven human proteins, which we designate the PLUNC family. From patterns of gene expression, conservation of genomic organization and predictions of protein structure, we show that PLUNC proteins are strongly related to, yet form a distinct family from, the BPI family. We suggest that members of the PLUNC family may play significant roles in host defence in the mouth, nose and upper airways.

RESULTS AND DISCUSSION

The identification of the family of PLUNC-related proteins arose from an iterative combination of analysis of the published sequence of chromosome 20 coupled with BLAST searches against the GenBank databases, using as the starting point the sequence of PLUNC (8). This procedure identified that PLUNC (which we now designate SPLUNC1) is a member of a family of seven proteins that, strikingly, are encoded by adjacent genes in an approximately 300 kb region of chromosome 20q11. Members of the PLUNC family fall into two groups based on their size. One group, which we designate as ‘short’ proteins, comprises SPLUNC1 (256 amino acids), SPLUNC2 (249 amino acids) and SPLUNC3 (253 amino acids). The other group, designated as ‘long’ proteins, comprises LPLUNC1 (484 amino acids), LPLUNC2 (458 amino acids), LPLUNC3 (463 amino acids) and LPLUNC4 (> 469 amino acids). All of these proteins, with the exception of LPLUNC4, which is not complete, contain putative signal peptides at the N terminus.

Within the PLUNC family, the sequence identity is rather low, ranging typically from 16% to 28%, with LPLUNC3 and LPLUNC4 sharing a somewhat higher pairwise identity of 37%. The BLAST E-values range from $10^{-62}$ (LPLUNC3 and LPLUNC4) to negligible ($>1$). Only for 10 of the possible 21 pairwise combinations within the family is the E-value less than 1. Rat orthologues of LPLUNC3 and LPLUNC4 have previously been reported (11). A weak sequence similarity was noted between these two orthologues, RYA3 and RY2G5, and BPI and LBP. Other non-human proteins related to the PLUNCs have previously been identified in a variety of species. The best studied of these is parotid secretory protein (PSP), which is similar in size to the SPLUNCs. PSP is one of the major secreted proteins of rodent saliva (12–14) although its exact function remains unknown. Other related proteins include SMGB from the rat (14) and BSP30 from the cow (15). The SPLUNC proteins described here are not close homologues of any of the above proteins, having at best a 30% sequence identity.

The pairwise amino acid identities between members of the PLUNC and BPI families range from 13% to 22%. The BLAST E-values between members of the two families range from $10^{-13}$ to negligible ($>1$), and only for 7 of the possible 28 pairwise combinations between the two families is the E-value less than 1. (We note that within the BPI family, all pairwise combinations yield E-values more significant than $10^{-14}$.) This level of pairwise identity is sometimes described as the ‘twilight zone’, where it can be difficult to establish definitive relationships. The PLUNC family was initially delineated using BLAST searches; however, this was an iterative process and no single sequence yielded the whole family in a single search. The more sensitive PSI-BLAST, which uses results of a simple BLAST search to build profiles characteristic of a family and then iterates this process, yields differing coverage of the family depending on the initial sequence used. Only in one case, using LPLUNC1, were all sequences identified. The most direct indication of the familial relationship and, at the same time, fold similarity to BPI was found using the 3DPSSM fold recognition service (16), which predicted a fold similar to that of BPI at the 95% confidence level for all seven members of the PLUNC family (Fig. 1). 3DPSSM is a threading method that uses not only primary sequence comparison but also predicted secondary structure and polarity/hydrophobicity to detect similarities to proteins of known 3D structure, thus improving its sensitivity over PSI-BLAST. The prediction of structural similarity to BPI was much more significant than matches to any other of over 6000 structures scanned by 3DPSSM. For the LPLUNC proteins the distinction was particularly clear, the E-values for BPI being approximately $10^{-10}$, and the next nearest candidates having E-values of greater than 6. For the SPLUNC proteins, the E-values for BPI were larger, ranging from 0.01 to 0.003, with the next nearest candidates having E-values ranging from 0.13 to 0.6. The poorer E-values for the SPLUNC–BPI match presumably reflects the fact that these proteins only contain the more variable N domain (see below). PSI-BLAST searches based on members of the BPI family did not yield all members of the PLUNC family, nor did PSI-BLAST searches based on members of the PLUNC family all yield a member of the BPI family.

BPI is the only member of the BPI family for which a structure has been experimentally determined, and comprises two domains that share the same fold and dock onto each other via a central $\beta$-sheet (17). The backbones of the two domains overlay to within about 3Å, which contrasts with the low pairwise sequence identity of only 15% between the two domains. The BPI structure is a well-studied example of how dissimilar protein sequences can adopt structurally similar protein folds (18). The two-domain structure of the BPI fold accounts for the division of the PLUNC family into short and long proteins, these groups containing either one or both of the BPI domains respectively. For the short sequences, the most confident match made by 3DPSSM was to the N-terminal domain of BPI. All three short proteins contain cysteine pairs, consistent with the location of the disulfide bond present in the N-terminal domain of BPI. In SPLUNC1 and SPLUNC2, the position of the residue putatively equivalent to Cys 175 of BPI aligns with a position eight residues later. This corresponds to being positioned two turns later in the helix, but probably simply reflects the uncertainty in the register of the alignment of this long helical segment and small changes in the relative positioning of the helical and sheet segments.
Figure 1. Conservation and diversity in the PLUNC and BPI families. (A–E) Sequence alignments created by 3DPSSM and coloured according to secondary structure. Helices are shown in yellow, strands in cyan and irregular structure in grey. The groups of proteins are (A) N-terminal domains of BPI family proteins, (B) C-terminal domains of BPI family proteins, (C) C-terminal domains of LPLUNC proteins, (D) N-terminal domains of LPLUNC proteins and (E) SPLUNC proteins. The sequence in the top line of each group, BPI (X-ray), represents the most confident match of the 3DPSSM program in all cases. The other sequences in each group are the query sequences submitted individually to 3DPSSM. These sequences are shown aligned to BPI, with gaps introduced into both sequences of 3DPSSM pairwise alignments as necessary to allow all sequences in a group to be displayed aligned to one representation of the BPI sequence. Where the query sequences align to an insert in the BPI (X-ray) sequence, the relative alignment of the query sequences is simply left justified within the available space. The sequence in the second line of groups (A) and (B) is BPI, which was also submitted to 3DPSSM for completeness and as a control. The BPI (X-ray) sequence is coloured according to the experimentally determined structure of BPI as evaluated by 3DPSSM, whereas the other sequences are coloured according to the 3DPSSM predicted secondary structure. The start points of the N-terminal domain alignments are offset to facilitate comparison. The full-length BPI sequence is numbered from \(-26\), consistent with the numbering used previously (17). (F,G) The degree of secondary structure conservation predicted by 3DPSSM in the PLUNC family, mapped onto the structure of BPI. Regions in red are poorly conserved at the secondary structure level (see Methods). (F) Data for the SPLUNC proteins mapped onto the N domain of BPI. (G) Data for the LPLUNC proteins mapped onto the full length of BPI. (F) and (G) were created using Molscript (31), using the coordinates 1ewf.pdb (18).
Whereas BPI is a product of inflammatory cells (19) and LBP is primarily produced within the liver (2), we have previously shown that SPLUNC1 is most prominently expressed in the nose, salivary glands and upper respiratory tract (8). The sites of SPLUNC1 expression correspond to locations where significant pathogenic loads are encountered and where bacterial sensing and/or neutralizing proteins might be required. SPLUNC2 and LPLUNC1–4 exhibit a similar but not identical expression pattern in tissues from these regions (Fig. 2A). Expression of SPLUNC3 was not detected in any tissue tested. In addition we were unable to identify SPLUNC3 transcripts by the use of RT-PCR. We extended the studies of LPLUNC1 by using a multiple-tissue dot blot. Expression above background was found in only three of 50 human tissues. Strongest expression was found in trachea, with lesser levels in lung and low expression in stomach (Fig. 2B).

The fact that the PLUNC genes are not only expressed in overlapping tissues but also found in a continuous locus on chromosome 20q11.2 (Fig. 3A) suggests that common regulatory regions may be required for targeting appropriate expression of these genes in vivo. In support of this theory, immediately 3' of the PLUNC gene cluster we identified the human orthologue of ryf3. ryf3 is a protein expressed in the lateral nasal glands and olfactory epithelium of the rat (11,20). Furthermore, we have mapped three of the mouse PLUNC orthologues, including splunc1 and lplunc1, to a single PAC clone from mouse chromosome 2, a region syntenic with human chromosome 20q11. Mouse psp is also found in the same genomic region (21), whereas rat psp and smgb are co-localized on chromosome 3q41–3q42 (22). The further study of these loci may yield regulatory regions with utility in gene therapy applications.

Whilst 3DPSSM indicates a clear relationship between the BPI and PLUNC families, it also indicates a fundamental structural difference between them. The BPI family are predicted to share a very closely similar secondary structure throughout both domains (Fig. 1A and B), a finding consistent with previous studies (17,23,24). In the LPLUNC proteins, the C-terminal domain similarly appears to share the same secondary structure as BPI (Fig. 1C). In contrast, in the N-domain of the LPLUNC proteins and in the SPLUNC proteins, there is a much greater variability (Fig. 1D and E). Much of the variability is in the region that forms one of the tips of the rather elongated molecule (Fig. 1F and G). This region, which includes the β-hairpins containing residues 45 and 96, is particularly important for the bactericidal activity of BPI (3) and is also the region of greatest structural difference between the two domains of BPI itself. It is the N-terminal domain of BPI that appears to mediate most of the LPS-binding and antibiotic effects of the protein. In fact it has been shown that the N-terminal domain of BPI alone exerts identical biological effects to the full-length protein; the single exception to this is the opsonic role of BPI, which is mediated by the C-terminal domain (3). The variability of the predicted structures in the N-terminal domain of the LPLUNC proteins may have implications for specificity of ligand binding. In the short proteins, the strands forming the interdomain interface in BPI are not well predicted. This argues against the hypothesis that the short proteins might simply dimerize via a similar interface.

Curiously, in the long proteins, the first β-strand, which forms an integral part of the interdomain β-sheet in BPI, is also not well predicted, nor is much of helix A, which follows it in the sequence. This helix forms a key part of the putative lipid-binding pocket formed by the two major helices and backed by the major β-sheet (17). However, in the N-terminal region of the SPLUNC proteins and of the N domains of the LPLUNC
proteins, where the structural similarity becomes so weak, one should be very cautious to not over-interpret the alignment. The key result is the qualitative difference between the alignments of the LPLUNC C-terminal domains or both of the BPI-family domains, which are compact and conserved at the secondary structure level, and the diffuse and ill-conserved secondary structure alignment of the N-terminal LPLUNC or SPLUNC domains. Simple pairwise sequence identity alone does not indicate this difference. The mean pairwise sequence identity between the LPLUNC C-terminal domains (23% ± 6%) is indistinguishable from that of the N-terminal domains (24% ± 8%), and the pairwise sequence identity to BPI is only very slightly stronger for the C-terminal domains (21% ± 1%) than for the N-terminal domains (17% ± 2%).

Three of the PLUNC proteins contain glycine-leucine-rich sequences, which do not appear to conform to any well-documented motif. In LPLUNC3, the sequence comprises residues 62–80, (GLLGGGGLGGGGLLGHGG; hydrophobic residues are in bold), in LPLUNC4, the clearest sequence comprises residues 43–71 (GGGLGGGGGLGDSGGGLGGGVGLGVGEGG), whereas in SPLUNC1, there are two shorter sequences comprising residues 58–66 (GLLGGGGLL) and 80–93 (GGTSGGGLGGLLG). The LPLUNC3 and LPLUNC4 sequences are highly conserved in rya3 and ry2g5 (11). It is difficult to predict the role of these sequences, but the glycine linkers could provide them with a large degree of conformational flexibility, which might enable the hydrophobic residues to form a highly adaptable binding surface.

The PLUNC genes are clearly evolutionarily related to each other and to the BPI, LBP, PLTP and CETP genes, since they exhibit a high degree of overall similarity of exon structure (Table 1) (25–27). The longer PLUNC genes are predicted to contain 16 exons (although the complete structures of SPLUNC3, LPLUNC3 and LPLUNC4 have not been determined), and the smaller proteins are encoded by genes containing 7 (SPLUNC3) or 9 exons (SPLUNC1 and SPLUNC2). In all of the complete PLUNC genes, the predicted translation start site is found within exon 2, an observation that also holds true for the PLTP gene (27) and the rodent Psp and Smgb genes (21,22). Furthermore, it is also clear that there is a high degree of conservation of exon sizes amongst the different genes (Table 1), despite rather limited conservation of the primary sequence. We have also previously described how the mouse Splunc1 gene is highly conserved compared with the human SPLUNC1 gene. All of the three rodent salivary gland proteins, rat and mouse Psp and rat Smgb, also have highly conserved genomic structures (22). Similar observations have been reported for the lipocalin family, where the levels of genomic and 3D structural similarity are far greater than would be predicted from primary sequence analysis (28,29).

In conclusion, the PLUNC proteins appear to form a self-contained family that is distinct from the BPI family. The PLUNC family is characterized by its co-localization within the genome and expression in a limited number of overlapping locations. In contrast to the structural conservation in the BPI family, there is predicted to be a much greater variability within the PLUNC family. The hypothesis that the PLUNC family of proteins may be involved in host defence against bacteria is supported by the very recent paper by Sung et al. (30). This role may be mediated by direct antibacterial actions or by an indirect LPS sensing/neutralizing activity. If the PLUNC family of proteins do indeed function in host defense then the variability of the N-terminal domain may confer a high level of diversity of response to microorganisms.
Table 1. Conservation of exon sizes between PLUNC genes and selected related proteins

<table>
<thead>
<tr>
<th>SPLUNC1</th>
<th>SPLUNC2</th>
<th>SPLUNC3</th>
<th>LPLUNC1</th>
<th>LPLUNC2</th>
<th>LPLUNC3</th>
<th>LPLUNC4</th>
<th>BPI</th>
<th>LBP</th>
<th>PLTP</th>
<th>CETP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ex 1</td>
<td>&gt;56</td>
<td>&gt;55</td>
<td>nd</td>
<td>&gt;45</td>
<td>&gt;160</td>
<td>nd</td>
<td>—</td>
<td>—</td>
<td>&gt;76</td>
<td>—</td>
</tr>
<tr>
<td>Ex 2</td>
<td>175</td>
<td>168</td>
<td>310</td>
<td>156</td>
<td>143</td>
<td>112</td>
<td>nd</td>
<td>&gt;171</td>
<td>&gt;230</td>
<td>111</td>
</tr>
<tr>
<td>Ex 3</td>
<td>160</td>
<td>145</td>
<td>151</td>
<td>142</td>
<td>94</td>
<td>57</td>
<td>105</td>
<td>115</td>
<td>100</td>
<td>115</td>
</tr>
<tr>
<td>Ex 4</td>
<td>108</td>
<td>108</td>
<td>108</td>
<td>108</td>
<td>105</td>
<td>105</td>
<td>105</td>
<td>115</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Ex 5</td>
<td>153</td>
<td>153</td>
<td>150</td>
<td>150</td>
<td>147</td>
<td>141</td>
<td>144</td>
<td>144</td>
<td>144</td>
<td>135</td>
</tr>
<tr>
<td>Ex 6</td>
<td>85</td>
<td>82</td>
<td>85</td>
<td>82</td>
<td>61</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>70</td>
</tr>
<tr>
<td>Ex 7</td>
<td>64</td>
<td>64</td>
<td>102/64</td>
<td>64</td>
<td>61</td>
<td>61</td>
<td>61</td>
<td>64</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>Ex 8</td>
<td>75</td>
<td>78</td>
<td>&gt;310</td>
<td>110/86</td>
<td>92</td>
<td>92</td>
<td>92</td>
<td>92</td>
<td>92</td>
<td>92</td>
</tr>
<tr>
<td>Ex 9</td>
<td>&gt;163</td>
<td>&gt;188</td>
<td>180</td>
<td>186</td>
<td>180</td>
<td>201</td>
<td>177</td>
<td>177</td>
<td>177</td>
<td>180</td>
</tr>
<tr>
<td>Ex 10</td>
<td>54</td>
<td>54</td>
<td>54</td>
<td>54</td>
<td>nd</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>51</td>
</tr>
<tr>
<td>Ex 11</td>
<td>159</td>
<td>171</td>
<td>171</td>
<td>171</td>
<td>171</td>
<td>168</td>
<td>168</td>
<td>168</td>
<td>160</td>
<td>165</td>
</tr>
<tr>
<td>Ex 12</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td>Ex 13</td>
<td>46</td>
<td>46</td>
<td>43</td>
<td>43</td>
<td>43</td>
<td>43</td>
<td>43</td>
<td>43</td>
<td>43</td>
<td>34</td>
</tr>
<tr>
<td>Ex 14</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>73</td>
</tr>
<tr>
<td>Ex 15</td>
<td>77</td>
<td>77</td>
<td>77</td>
<td>77</td>
<td>77</td>
<td>77</td>
<td>77</td>
<td>77</td>
<td>77</td>
<td>86</td>
</tr>
<tr>
<td>Ex 16</td>
<td>&gt;170</td>
<td>&gt;427</td>
<td>nd</td>
<td>&gt;28</td>
<td>&gt;358</td>
<td>&gt;380</td>
<td>&gt;300</td>
<td>&gt;250</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The sizes of individual exons (in bp) for the human PLUNC genes are compared with those of human BPI, LBP, PLTP and CETP. nd represents unidentified exons, — represents where an exon is missing. * in the CETP gene represents where there are two exons, and two figures separated by / represents the sizes of alternatively spliced exons.

MATERIALS AND METHODS

Characterization of PLUNC genes

PLUNC genes were identified by a combination of nucleotide and protein BLAST searches (blastx and blastn) of the public databases using the human PLUNC sequence as a start point. Individual gene products thus identified were themselves used for further searches. ESTs identified by such searching were assembled into contigs. Where required for completion of sequences, individual Image clones were obtained from the MRC HGMp, Cambridge, UK and sequenced using an ABI 377 sequencer with appropriate flanking or specific oligonucleotides as required. The genomic structure of individual genes was determined from genomic sequence from the chromosome 20 sequencing project. Subsequent to our studies, the PLUNC gene structures and predicted protein products have appeared in the chromosome 20 sequence. With the exception of the genomic structure of SPLUNC3, our determination is in agreement with the annotation assigned in GenBank.

RNA isolation and expression studies

Total RNA was isolated from human nasal septal epithelium, lung and submandibular gland using the RNAeasy system (Qiagen). Samples were resolved on replicate denaturing agarose gels, northern-blotted and hybridized with random primed cDNA probes as previously described (8). The PLUNC probes corresponded to either ESTs or to RT–PCR fragments. RT–PCR analysis was performed using primer pairs to all of the PLUNCs with 1μg of both nasal septal epithelium and lung RNA as the starting templates for oligo dT priming. Appropriately sized products were directly cloned in TOPO pCRII and sequenced for verification. LPLUNC1 expression was also investigated in a wider range of tissues by hybridizing a multiple-tissue RNA dot blot containing RNA from 50 different human tissues (Clontech No. 7770-1).

Sequence alignment and structural analysis

BLAST and PSI-BLAST were performed via the National Centre for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/BLAST) and used the non-redundant sequence database. PSI-BLAST searches used the default threshold of E < 0.005 and were iterated to convergence. Sequences that fell outside this tolerance were analysed up to an E-value of 10 to check that missing sequences were not marginally detected. Sequences were individually submitted to 3DPSMM (http://www.bmm.icnet.uk/~3dpssm/) (16). An in-house python script was used to bring the matched BPI sequences in a group of pairwise alignments into register by inserting gaps in both sequences of individual pairwise alignments as necessary. The classification of residue positions as conserved or variable was performed as follows. Each position in the alignment of a group of proteins (as in Fig. 1A–E) was taken in turn. If this position corresponded to a gap in the BPI (X-ray) sequence then the position was not classified. Otherwise the predicted secondary structure for the query sequences were compared with the experimental secondary structure of BPI. If there was agreement for three or more of the long proteins (two or more for the short proteins) then the position was denoted conserved. Otherwise the position was denoted variable. Gaps in the query sequence were counted as a secondary structure mismatch. According to the EVA project (http://cubic.bioc.columbia.edu/eva/sec/common.html), the PSIPRED method currently has a 3-state accuracy (Q3) of 76%. This is very close to the value observed here for BPI (82%). PSIPRED uses a PSI-BLAST procedure to produce a sequence profile upon which the prediction is made, and thus the predictions for a homologous family are correlated to an extent. However, the PSI-BLAST search is restricted to avoid sequence drift, and thus the profile obtained is quite
strongly biased towards the input sequence (David Jones, personal communication). Pairwise sequence identities were evaluated using simple sequence directed alignments using clustalw v1.7 in a pairwise fashion.

Accession numbers for PLUNC proteins

The individual PLUNC gene products are represented in the database with the following GI numbers: SPLUNC1 7958616, SPLUNC2 9801234, LPLUNC1 14772576, LPLUNC2 11877274, LPLUNC3 11877275, LPLUNC4 11877276. The deduced sequence of SPLUNC3 was determined from multiple ESTs represented by BG71970 or AW340528.

ACKNOWLEDGEMENTS

We are grateful to Jonathan Gitlin, Beth LeClair, David Jones, Paul Thaw, Jon Waltho, Mike Williamson and Moira Whyte for helpful discussions. We thank Deirdre Devine, Alec High and Rob Read for tissue samples. Image Clones were provided by the UK-HGMP Centre, Cambridge, UK. The Krebs Institute is a designated BBSRC Centre and a member of NESBIC. This work was supported in part by project and equipment grants from The Birth Defects Foundation.

REFERENCES
