Immunology 1993 **78** 159–165

Structural similarity between bovine conglutinin and bovine lung surfactant protein D and demonstration of liver as a site of synthesis of conglutinin

B. L. LIM, J. LU & K. B. M. REID *MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, Oxford, U.K.*

Accepted for publication 10 August 1992

SUMMARY

Conglutinin is a Ca^{2+}-dependent, carbohydrate-binding, serum protein which contains an N-terminal collagen-like region and a C-terminal, C-type lectin domain. To date, conglutinin, which appears to play an important role in defence mechanisms, has been fully described, by protein sequence analysis, only in the bovine system. To allow comparison of lung surfactant protein D (SP-D) with conglutinin, within one species, a full-length cDNA clone for SP-D has been isolated from a bovine lung library. The derived amino acid sequence for bovine SP-D shows a higher (78%) level of identity to the sequence of conglutinin than to the sequence of human or rat SP-D (67 and 65% respectively). However, SP-D and conglutinin are known to have different carbohydrate-binding specificities, therefore some of the 16 residues conserved in the C-type lectin domains of all three species of SP-D, but which are not conserved in conglutinin, appear likely to be involved in determination of specificity. The use of a polymerase chain reaction (PCR)-derived DNA probe for bovine SP-D in Northern blotting studies yielded a signal from bovine liver mRNA as well as the expected signal from bovine lung mRNA. Since SP-D appears to be a lung-specific protein, it seems probable that the liver is the primary site of synthesis of conglutinin.

INTRODUCTION

Pulmonary surfactant is a complicated mixture of phospholipids and proteins which lines the alveolar spaces of lung. Its major function is to modulate surface tension at the alveoli, but it may also play roles in regulating transepithelial fluid permeability and in defending against inhaled micro-organisms. The major constituent of the surfactant is lipid, but four proteins have so far been isolated from the surfactant.[1,2] They are designated as surfactant protein (SP)-A,[3,4] SP-B,[5] SP-C[6] and SP-D.[7–10] SP-B and SP-C are small hydrophobic proteins which appear to facilitate the adsorption of phospholipids to the air–liquid interface,[11] whereas both SP-A and SP-D are collagenous, carbohydrate-binding glycoproteins containing domains similar to the family of C-type mammalian lectins.[4,12] SP-D has been isolated from lung lavage[10] and amniotic fluid,[7] it has a MW of approx. 600,000 on gel filtration in non-dissociating conditions and it yields a 44,000 MW polypeptide chain under reducing conditions in SDS–PAGE.[7] Analysis of cDNA clones of SP-D isolated from rat[13] and human[7] has confirmed that SP-D has a short, cysteine-containing N-terminal non-collagenous domain, a collagenous domain and a C-terminal, calcium-dependent, carbohydrate-recognition domain (CRD) consistent with its ability to bind to α-glucosyl residues in maltose and several other sugars in a calcium-dependent manner.[9] The function of SP-D is not clear, however, the rat and human proteins show a very high homology to bovine conglutinin, a carbohydrate-binding protein specific for *N*-acetylglucosamine and mannose residues.[14,15] Conglutinin is capable of binding to the C1q receptor[16] and the human immunodeficiency virus (HIV)-1 envelope glycoprotein gp160,[17] and it has also been found to have antibacterial activity.[18] The structural similarity between SP-D and conglutinin suggests SP-D may also play a role in defence against pathogens and this view is consistent with recent findings showing that SP-D binds and subsequently aggregates *Escherichia coli*.[19]

In the present study, the use of polymerase chain reaction (PCR) techniques to clone the bovine SP-D (bSP-D) cDNA and also to increase the sensitivity of Northern blotting is described. Analysis of the derived amino acid sequence for bSP-D showed that it has a high degree of sequence identity in comparison with bovine conglutinin (77·8%).[14] Comparison of these two proteins, as well as the analogues of SP-D in rat[13] and human,[7] provides some indication as to the residues which may be important in determining carbohydrate-binding specificity. A

Abbreviations: CRD, carbohydrate recognition domain; IPTG, isopropythio-b-D-galactoside; MBP-A, mannan-binding protein A; MBP-C, mannan-binding protein C; PCR, polymerase chain reaction; PVDF, poly-vinylidene difluoride; SP-A, surfactant protein A; SP-B, surfactant protein B; SP-C, surfactant protein C; SP-D, surfactant protein D; X-gal, 5-bromo-4-chloro-3-indoyl-b-D-galactoside.

Correspondence: Dr K. B. M. Reid, MRC Immunochemistry Unit, Dept. of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

high specific activity, single-stranded DNA probe was generated by the PCR technique and used in Northern blotting. This allowed the detection of a signal in bovine liver mRNA as well as in bovine lung mRNA which indicates that a protein highly homologous to SP-D, presumably bovine conglutinin, is synthesized in bovine liver.

MATERIALS AND METHODS

PCR amplification

SP-D was isolated from bovine lung lavage by affinity chromatography on maltose-Sepharose as described by Lu *et al.*[7] The limited protein sequence information about bSP-D was obtained by N-terminal sequencing analysis of a collagenase-resistant fragment as described by Lu *et al.*[7]

Two degenerate, inosine-containing, oligonucleotide primers were designed from the N-terminal amino acid sequence of the collagenase-resistant fragment of SP-D (Fig. 1), and were synthesized on a model 381A DNA synthesizer (Applied Biosystems, Cheshire, U.K.). The template was obtained from a λgt10 bovine lung cDNA library (Clontech Laboratories Inc, Cambridge BioScience, Cambridge, U.K.) and a PCR reaction was performed with 30 cycles in a Perkin-Elmer/Cetus DNA Thermal Cycler (Buckinghamshire, U.K.) using a cycle of denaturation for 0·5 min at 94°, annealing for 1·0 min at 48° and an extension at 72° for 0·5 min. The reaction mixture (50 μl) contained 25 pmol of each primer (B4 and B5, Fig. 1), 0·25 mM dNTP, 1·25 U Taq polymerase (Promega, Southampton, U.K.) with the manufacturer's buffer system, together with 0·5 μl (about 10^6 phage) of the cDNA library. The PCR product (5 μl) was electrophoresed on a 4% (w/v) NuSieve agarose gel (FMC Bioproducts, Vallensbaek Strand, Denmark). The expected 90 base pair (bp) product was obtained (Fig. 1).

Subcloning and DNA sequencing

The two primers (B4 and B5) were phosphorylated and used to repeat the PCR described above. After electrophoresis, the 90 bp fragment was cut out from the gel and extracted from the gel slice by centrifugation through glass wool.[20] The DNA fragment (50 ng) was blunt-end ligated to the *Hinc*II site of pBluescript SK (0·1 μg) using T4 DNA ligase (1 U, Amersham International, Amersham, U.K.) in a 10-μl reaction volume overnight at room temperature. The *E. coli.* strain XL-Blue 1 was used as the host and transformation of cells with the ligation mixture was performed by a standard method.[21] After transformation, the cells were plated on LAT plates (LB plates with 100 μg/ml ampicillin and 10 μg/ml tetracycline), which had been spread with 5-bromo-4-chloro-3-indoyl-b-D-galactoside (X-gal) and IPTG 2 h before the cells were inoculated. White colonies were screened by the PCR using the same pair of primers and conditions as used previously. Colonies which yielded the 90 bp PCR product were selected and the plasmids were purified by alkaline lysis method[21] from a 10-ml miniprep culture. Double-stranded DNA sequencing was performed on two purified plasmids by the dideoxynucleotide chain termination procedure using T7 polymerase (Pharmacia, Milton Keynes, U.K.) as recommended by the manufacturer.

Screening of the λgt 10 bovine lung cDNA library

A 50-mer oligonucleotide probe was synthesized based on the DNA sequence of the clones (Fig. 1). This oligonucleotide was end-labelled with γ-[^{32}P]ATP using T4 kinase and was used to screen the library (approximately 4×10^4 plaques). Positive plaques were picked up and the phages were eluted in 1 ml SM buffer (50 mM Tris-HCl, 10 mM NaCl, 10 mM $MgSO_4$ and 0·01% gelatin). Screening of the positive plaques was performed by PCR reaction with primers B4 and B5, in 25-μl reaction volume. *Eco*RI inserts from the positive clones were subcloned into pBluescript for sequence analysis. One full-length cDNA clone

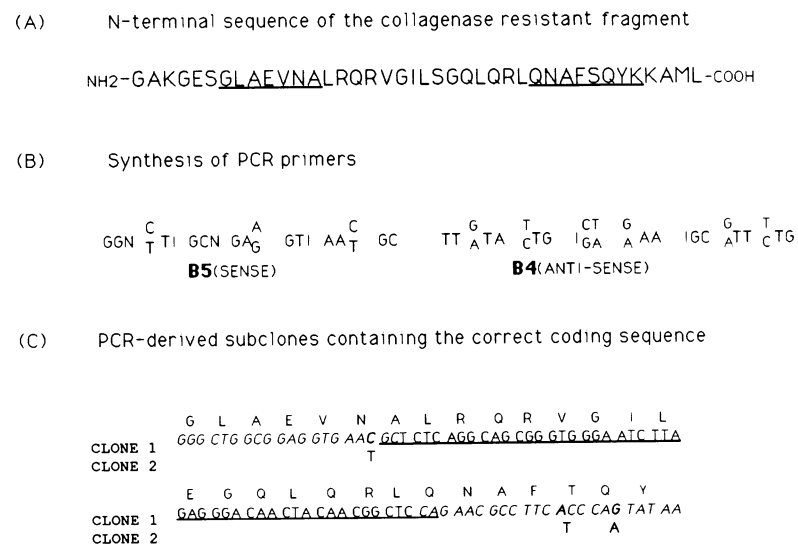

Figure 1. Strategy for PCR cDNA cloning by using degenerate primers. After collagenase digestion of bSP-D, the large, C-terminal, collagenase-resistant fragment was sequenced and the sequence obtained is shown in (A). Two PCR primer mixes containing inosine were synthesized, based on the underlined peptide sequences (B). A 90-mer product was amplified and subcloned into a plasmid. Two clones were found to carry the correct DNA sequence, although there were discrepancies present in the regions originating from both primer mixes (C). A 50-mer oligonucleotide (underlined in C) was synthesized for use in screening a λgt10 cDNA bovine lung library.

```
                                                                                M  L  L  L  P  L  S
aattccgggtgctatagttgcttcctgtaggactgcagactccagtactagtctgtccagagcaacagtgataggaaacaagccagcattgtaagaggacATGCTTCTCCTCCCTCTCT 120
                                                                                                    -20

   V  L  L  L  L  T  Q  P  W  R  S  L  G  A  E  M  K  I  Y  S  Q  K  T  M  A  N  A  C  T  L  V  M  C  S  P  P  E  D  G  L
CCGTGCTGCTCCTGCTCACACAGCCCTGGAGATCCCTGGGAGCAGAAATGAAGATCTATTCCCAGAAAACAATGGCCAACGCCTGTACCCTGGTCATGTGTAGCCCCCCGGAGGATGGTT 240
      -10                                 1                                  10                                  20

   P  G  R  D  G  R  D  G  R  E  G  P  R  G  E  K  G  D  P  G  S  P  G  P  A  G  R  A  G  M  P  G  P  A  G  P  I  G  L  K
TGCCTGGTCGTGATGGACGAGATGGGAGAGAAGGCCCCCGGGGGGAGAAGGGAGATCCAGGTTCACCAGGACCTGCAGGACGAGCAGGAATGCCTGGACCAGCTGGCCCTATTGGGCTGA 360
      30                                  40                                  50                                  60

   G  D  N  G  S  A  G  E  P  G  P  K  G  D  T  G  P  P  G  P  P  G  M  P  G  P  A  G  R  E  G  P  S  G  K  Q  G  S  M  G
AAGGAGACAATGGCTCTGCTGGAGAACCCGGACCAAAGGGAGACACTGGACCACCTGGGCCTCCAGGTATGCCTGGACCAGCTGGAAGAGAGGGCCCCTCAGGGAAGCAGGGGAGCATGG 480
      70                                  80                                  90                                  100

   P  P  G  T  P  G  P  K  G  D  T  G  P  K  G  G  V  G  A  P  G  I  Q  G  S  P  G  P  A  G  L  K  G  E  R  G  A  P  G  D
GACCTCCAGGCACACCAGGCCCCAAAGGAGACACTGGGCCCAAAGGAGGAGTGGGTGCCCCAGGCATTCAGGGCTCCCCAGGCCCTGCAGGTCTCAAAGGAGAGAGAGGTGCCCCTGGTG 600
      110                                 120                                 130                                 140

   P  G  A  P  G  R  A  G  A  P  G  P  R  G  A  I  G  P  Q  G  P  S  G  A  R  G  P  P  G  L  K  G  D  R  G  T  P  G  E  R
ACCCCGGAGCCCCTGGACGTGCTGGGGCACCAGGGCCTCGTGGAGCCATAGGTCCACAGGGGCCTTCAGGTGCCAGGGGCCCCCCAGGACTGAAGGGAGACAGAGGTACTCCTGGAGAAA 720
      150                                 160                                 170                                 180

   G  A  K  G  E  S  G  L  A  E  V  N  A  L  R  Q  R  V  G  I  L  E  G  Q  L  Q  R  L  Q  N  A  F  S  Q  Y  K  K  A  M  L
GAGGAGCAAAGGGGGAGAGTGGGCTTGCAGAGGTCAATGCTCTCAGGCAGCGGGTGGGAATCTTAGAGGGACAACTACAACGGCTCCAGAATGCCTTCTCTCAGTATAAGAAAGCGATGC 840
      190                                 200                                 210                                 220

   F  P  N  G  R  S  V  G  E  K  I  F  K  T  V  G  S  E  K  T  F  Q  D  A  Q  Q  I  C  T  Q  A  G  G  Q  L  P  S  P  R  S
TCTTCCCTAATGGCCGGAGTGTCGGGGAGAAGATCTTTAAGACGGTAGGCTCTGAAAAAACGTTTCAGGATGCCCAGCAGATCTGCACACAGGCTGGAGGACAGTTGCCCTCCCCACGTT 960
      230                                 240                                 250                                 260

   G  A  E  N  E  A  L  T  Q  L  A  T  A  Q  N  K  A  A  F  L  S  M  S  D  T  R  K  E  G  T  F  I  Y  P  T  G  E  P  L  V
CTGGAGCTGAAAACGAGGCCTTGACTCAGCTGGCCACAGCCCAGAACAAGGCTGCTTTCCTGAGCATGAGCGACACCAGGAAGGAGGGTACTTTCATCTACCCCACGGGGGAGCCCCTGG 1080
      270                                 280                                 290                                 300

   Y  S  N  W  A  P  Q  E  P  N  N  D  G  G  S  E  N  C  V  E  I  F  P  N  G  K  W  N  D  K  V  C  G  E  Q  R  L  V  I  C
TCTATTCCAACTGGGCCCCCCAGGAGCCCAACAATGATGGCGGCTCAGAGAACTGTGTGGAGATCTTTCCCAATGGCAAGTGGAATGACAAAGTCTGCGGAGAGCAGCGCCTCGTGATCT 1200
      310                                 320                                 330                                 340

   E  F  *
GCGAGTTCTGAgctcctcctgcacacacacacacacatagtgtgtgttggggcggtgggggtcgggggggggggatgggcagtgcccagagctgcatttttccagtgtttgaataaaat 1320
    349

agtgaccctctactggccagggcttctccacagagccacaggataaggccagaggcagggctcctatggaatacatccctcagaataaagtttgaaactggcttcacacaaaaaaaaaaa 1440

aaaaaccggaattc 1454
```

Figure 2. cDNA sequence and deduced amino acid sequence of bSP-D. The total length of the cDNA is 1454 nucleotides, including both the 5′ and the 3′ untranslated regions. The ATG start codon is located at nucleotide 102 and the TGA stop codon is located at 1209. The whole translated protein has 369 amino acid residues, with the first 20 residues being considered to be signal peptide (in italics). Two potential polyadenylation signals were located, starting at nucleotides 1313 and 1404 respectively. A potential N-linked glycosylation site is double underlined. An anti-sense oligonucleotide complementary to nucleotides 1201 and 1225 was synthesized as primer for generating, by PCR, a single-stranded probe for Northern blotting.

of 1·5 kb was sequenced, in both directions, by the dideoxynucleotide chain termination procedure using T7 polymerase (Pharmacia).

Northern blotting

Total RNA was isolated from tissues of several species (bovine, rat and mouse) by the guanidine thiocyanate method.[22] Total RNA (10 μg) from each sample was electrophoresed on a formaldehyde-containing 1% (w/v) agarose gel. After electrophoresis, the RNA was transferred to a Hybond-N membrane (Amersham) by capillary blotting and fixed onto the membrane by ultraviolet (UV) cross-linking (XL-1500 UV cross-linker, Spectronics Corp. Ams, Burford, U.K.). A high specific activity, single-strand DNA probe was generated by using the PCR method. Whole pBluescript plasmid (0·1 μg) containing the full-length bovine SP-D cDNA clone was used as a template and an antisense primer, complementary to nucleotides 1201–1225 (Fig. 2), was employed. Thirty cycles of denaturation for 0·5 min at 94°, annealing for 0·5 min at 55° and an extension at 72° for 50 seconds was performed. The reaction mixture (50 μl) contained 25 pmol of the primer (Fig. 2), 0·2 mM cold dATP, dTTP and dGTP, 0·625 U Taq polymerase (Promega) with the manufacturer's buffer system, together with 2 μl [^{32}P]α-dCTP (3000 Ci/mmol, 10 mCi/ml). After labelling, the probe was purified on a nick column (Pharmacia) and added directly to the formamide containing prehybridization buffer (50% Formamide, 5 × SSPE, 5 × Denhardt's solution, 0·5% SDS, 20 μg/ml sperm DNA). The membrane was hybridized with the probe overnight at 42°. After hybridization, the filter was washed three times with 2 × SSC, 0·1% SDS at room temperature and then washed once in 0·1 × SSC, 0·1% SDS at 65° for 15 min. The membrane was exposed to X-ray film for 1 day at −70°.

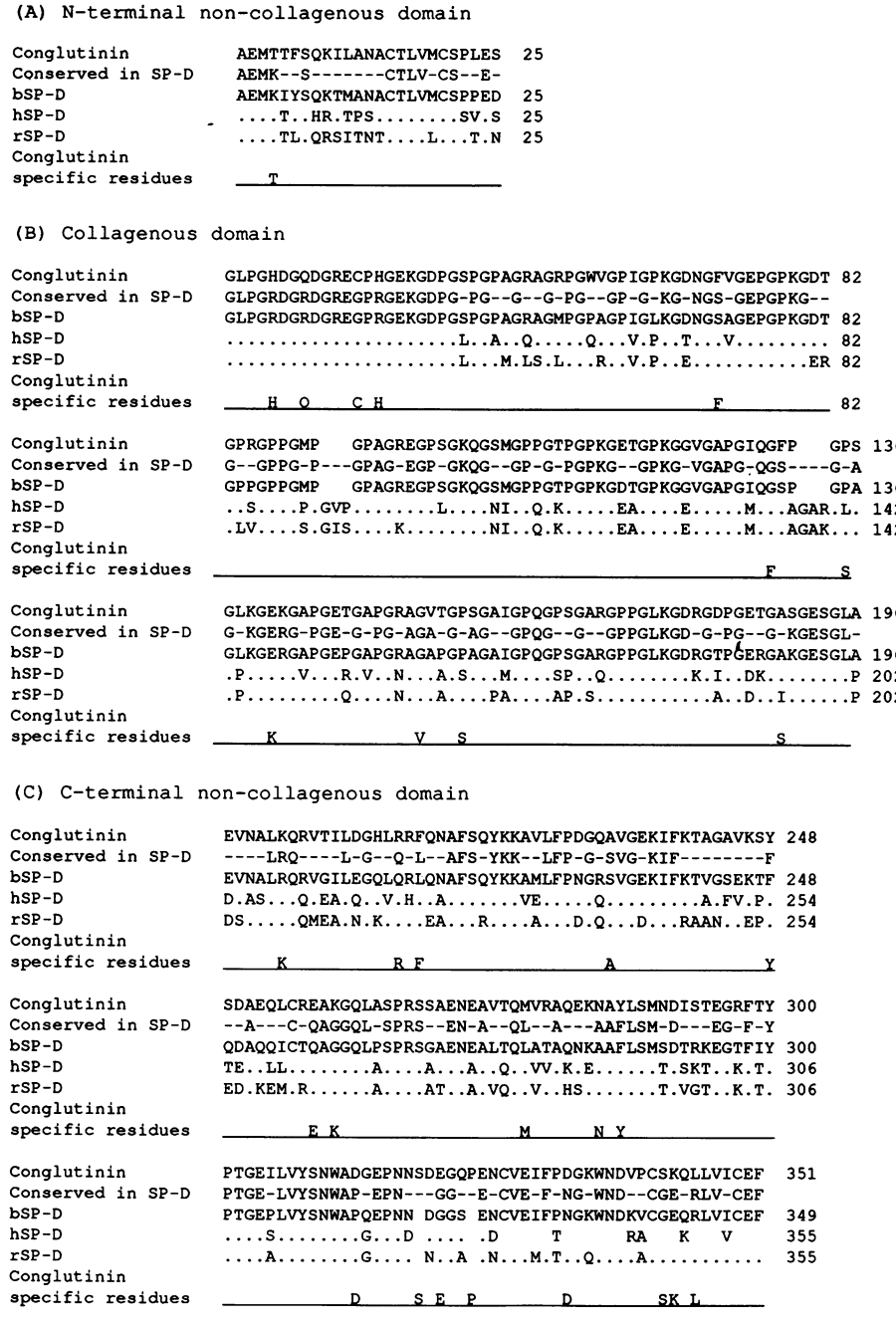

Figure 3. Comparison of the amino acid sequences of SP-D from three species with each other and with the sequence of bovine conglutinin. The three domains of these proteins are compared separately. The first line and the third line are bovine conglutinin and SP-D respectively, whereas the fourth and fifth lines are hSP-D and rSP-D respectively. Residues in hSP-D and rSP-D are replaced by dots if they are the same as bSP-D. The single letter code of the amino acid is shown at positions which differ between bSP-D and hSP-D, or rSP-D. The fifth line displays the conserved residues in all three SP-D sequences. Non-conserved residues are denoted by dashes. The sixth line is the comparison of the SP-D-conserved residues with the corresponding residues in bovine conglutinin, residues in conglutinin which differ from the SP-D concensus sequence (i.e. 'conglutinin-specific' residues) are indicated on the sixth line by the single letter code. The number at the end of each line indicates the residue number of the last residue on that line.

RESULTS

PCR amplification

After collagenase digestion, N-terminal peptide sequencing of the collagenase-resistant fragment yielded a 40 amino acid long stretch of bSP-D sequence (Fig. 1). Two degenerate, inosine-containing oligonucleotides were synthesized based on the amino acids sequence shown in Fig. 1. These two primers had a degeneracy of 8 (B5, sense) and 128 (B4, anti-sense), and contained four and two inosines respectively. Since these two primers were not very long (B5 was 20-mer and B4 was 23-mer),

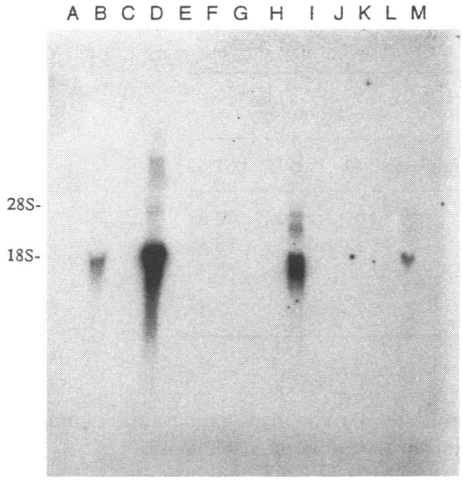

Figure 4. Northern blot analysis. Total cellular RNA (10 μg) from various tissues of bovine, rat and mouse origin was separated by electrophoresis on a formaldehyde-containing 1% (w/v) agarose gel, transfered to Hybond-N by capillary action, and hybridized to a ^{32}P-labelled, single-strand DNA probe, generated by the PCR technique, using bovine cDNA as template. The samples were bovine heart (A), liver (B), spleen (C), lung (D), muscle (E), rat heart (F), liver (G), spleen (H), lung (I), muscle (J), mouse liver (K), heart (L) and lung (M).

and they were degenerate, the annealing temperature was a very critical factor for amplification of the correct product. It was found by performing trial PCR reactions, in the temperature range 42 to 55°, that 48° gave the best yield of the expected 90 nucleotide-long PCR product.

Subcloning of the 90 nucleotide-long fragment

After blunt-end ligation, five potentially positive clones were sequenced and two of them were found to code for bSP-D. Comparing the sequences of two clones (Fig. 1), it was found that the sequences contributed by the two primer mixes were not exactly the same. This was due to the degeneracy of the primers. As a result, only the sequence between the two primers was reliable and a 50-nucleotide-long, single-strand DNA probe was synthesized (Fig. 1) to screen a λgt10 bovine lung cDNA library. It was noticed that the derived peptide sequence had a single amino acid difference compared with the sequence obtained by direct peptide sequencing (Glu for Ser at position 22, Fig. 1A,C). This difference is considered to be due to an error in peptide sequencing.

λgt10 cDNA library screening

About 4×10^5 plaques were screened using the ^{32}P end-labelled 50 nucleotide-long probe and 12 potential positive clones were identified and six of them were found to be positive after rescreening by PCR using B4 and B5 primers. The size of the six inserts in the positive clones varied from 0·9 to 1·5 kb. The 1·5-kb clone was sequenced in both directions and found to contain the full-length cDNA coding for bSP-D (Fig. 2). The overall length of the cDNA insert is 1454 nucleotides. A translation initiation codon (ATG) was found at nucleotide 102–105 and a polypeptide of 369 amino acids in length is predicted to be translated, in which the first 20 amino acids are thought to be a signal peptide thus giving a mature protein of 349 amino acids. The calculated molecular mass of the signal peptide is 2264 and that of the mature protein (without taking into account the possible presence of carbohydrate) is 35,300. A TGA stop codon is present, at positions 1209–1211 and is followed by 243 nucleotide-long, 3′ untranslated region containing two AATAAA polyadenylation signals, starting at nucleotides 1313 and 1404, respectively.

Similar to human SP-D (hSP-D) and rat SP-D (rSP-1), the bSP-D sequence can be divided into three domains, a cysteine-containing non-collagenous N-terminal sequence of 25 amino acids, a collagen-like domain of 171 amino acids (57 Gly-X-Y repeats), and a non-collagenous C-terminal sequence of 153 amino acids. There are 35 proline and nine lysine residues in the collagen-like region. The collagen-like region also contains one potential N-linked glycosylation consensus site of Asn-Gly-Ser, at residues 70–72. There are six cysteine residues present in bovine SP-D, two in the N-terminal non-collagenous region (residues 15 and 20) and four in the C-terminal non-collagenous region (residues 255, 325, 339 and 347) (Fig. 3).

Northern blotting

RNA was extracted from a variety of bovine, rat and mouse tissues. A high specificity ^{32}P-labelled, single-stranded probe was produced by the PCR method as described in the methods. As seen in Fig. 4, a strong signal is obtained from bovine lung RNA at about 1·8 kb, comparable to the size of hSP-D mRNA.[10] Moreover, weaker bands can also be seen in rat and mouse lung RNA tracts, with sizes of approximately 1·7 and 1·8 kb respectively. Unexpectedly, a weak signal of about 1·8 kb is present in bovine liver RNA, but not in rat and mouse liver RNA. Finally, longer exposure of the blot showed that a larger transcript, with a size of approximately 4 kb, was present in bovine lung RNA preparation.

DISCUSSION

Since SP-D contains collagen-like regions it is grouped in the same subset of Ca^{2+}-dependent C-type lectins as mannan-binding protein A (MBP-A), MBP-C, SP-A and conglutinin.[23] It is clear that SP-D shows a greater similarity to conglutinin than to any of the other members of the subset, with, for example, rSP-D[13] having approximately 66, 40, 38 and 40% amino acid sequence identity (after allowing for gaps) to bovine conglutinin,[14] rSP-A,[24] rMBP-A[25] and rMBP-C,[25] respectively. The similarity between SP-D and conglutinin is further illustrated by the finding that bSP-D shows a higher sequence similarity (78% identity) to bovine conglutinin than to hSP-D (73% identity to bSP-D) or rSP-D (72% identity to bSP-D) (Table 1, Fig. 3). The similarity between bSP-D and conglutinin is also emphasized by the presence of a collagen-like region of 171 residues in both molecules which is six residues shorter than the collagen-like regions of rSP-D and hSP-D (Table 1, Fig. 3). Thus the length of the collagen-like region may not be critical to the function of SP-D. It is, however, striking that the first 22 residues of the collagen-like domain are completely conserved in SP-D from all three species (Fig. 3) while four out of the only 10 differences, between SP-D and conglutinin over their collagen-like domains, are located in this region. All three of the SP-D sequences contain strongly hydrophilic residues (D, E, R, K) near the

Table 1. Number of residues in each domain and homology between conglutinin and SP-D from three species

	Length of domain (residues)				Homology		
	hSP-D	rSP-D	bSP-D	Conglutinin	hSP-D	rSP-D	bSP-D
N-terminal domain	25	25	25	25	60·0	60·0	72·0
Collagen domain	177	177	171	171	72·5	70·2	88·3
C-terminal CRD	153	153	153	155	62·1	60·8	66·7
Total/Average	355	355	349	351	66·7	65·0	77·7

beginning of their collagen-like domains, with the two D-G-R sequences (residues 29–34) perhaps having some functional significance. These D-G-R sequences are also present in the collagen-like domains of the other collagenous C-type lectins, MBP and SP-A. The D-G-R sequence is the reverse of the R-G-D recognition sequence found for the integrin superfamily of cell-surface receptors[26] and it is of interest that a cell-surface receptor complex has been reported to bind to an R-G-D sequence within the triple-helical region of type I collagen.[27] In conglutinin, there is only one D-G-R sequence, this could reflect differences in the function of SP-D and conglutinin. In all three species of SP-D there is one potential site of N-linked glycosylation (residues 70 and 72, Fig. 3), while no such site is present in the conglutinin molecule. Since rSP-D is known to contain a single N-linked oligosaccharide chain[28] bSP-D may also be conserved at the conserved potential glycosylation site.

All three SP-D sequences are the same in length over their C-terminal lectin domains (153 residues) and are two residues shorter than bovine conglutinin over this region (Fig. 3), which is perhaps not unexpected in view of the different carbohydrate specificities shown by SP-D and conglutinin. Conglutinin has a strong affinity for non-reducing *N*-acetyl-D-glucosamine,[15] while SP-D is specific for α-glycosyl residues, such as maltose, and has a relatively low affinity for *N*-acetyl-D-glucosamine.[7,9] The alignment of the C-terminal amino acid sequences of bSP-D, hSP-D, rSP-D and bovine conglutinin (Fig. 3) allows an assessment to be made of which residues might be important in conserving the overall structure of the CRD and which residues may be involved in the determination of carbohydrate-binding specificity. For example, the four completely conserved Cys residues found in the CRD would clearly be expected to play a structural role, with an expected disulphide bond pattern Cys225–Cys347 and Cys325–Cys339, as based on the pattern seen in other members of the C-type lectin superfamily. On the other hand, carbohydrate specificity may be determined by some of the 16 residues which are conserved in the CRD of all three species of SP-D but which are not conserved in in conglutinin (Fig. 3). The two 'additional' residues found in the conglutinin CRD could also change the fine structure of the domain and be responsible for differences in the binding specificities. These 'additional' residues (Ser319 and Pro324; Fig. 3) are probably located on the surface of the CRD (i.e. on 'loop 4' as determined from the crystal structure of the MBP CRD reported by Weis *et al.*[29] It is considered that the carbohydrate binding site on the MBP CRD may extend over a substantial portion of the surface.[29] Therefore, comparisons of CRD sequences provide a useful starting point for protein engineering studies involving the generation of mutant CRD for fine analysis of carbohydrate specificity.

In the electron microscope, conglutinin has been observed as an X-shaped, tetrameric molecule with four globular heads protruding outwards, each being connected to a central 'hub' by four elongated collagen 'arms'.[30] Bovine SP-D has been viewed as a mixture of monomers, dimers, trimers and tetramers in the electron microscope (J. Lu, H. Widemann, S. Thiel, R. Timpl and K. Reid, unpublished observations). The tetrameric forms of SP-D have similar overall dimensions to the tetramers seen in the conglutinin preparation and the tetramer seems likely to be the native form of SP-D since it behaves as an approximately 600,000 MW protein on gel filtration in non-dissociating conditions[7] (i.e. with each monomer being composed of three polypeptide chains and the tetramer therefore being composed of 12 polypeptide chains, each of approximately 43,000 MW). On SDS–PAGE, under non-reducing conditions, SP-D has been reported[31] to behave as a single band of 120,000–130,000 MW, which is indicative that there are no disulphide bridges between the monomers. This would be consistent with the finding of a large percentage of monomers when SP-D is viewed by electron microscopy. The value of 160,000 MW obtained by SDS–PAGE for SP-D, under non-reducing conditions, is considered to be an overestimate due to the use of a 5–15% gradient gel system,[7] especially as it is known that the MW of SP-D on SDS–PAGE varies as a function of acrylamide concentration.[31]

It has been suggested from immunohistochemical studies that liver is the site of synthesis of bovine conglutinin[32] and therefore it was of interest that the Northern blotting analysis using the bSP-D, PCR-derived DNA probe, showed the presence of a signal from bovine liver RNA as well as the expected signals from the bovine, rat and mouse lung RNA preparations (Fig. 4). No signal was obtained from rat, or mouse, liver RNA which is consistent with a previous study in which a rSP-D cDNA probe was used in Northern blot analysis.[13] Conglutinin appears to be present in significant amounts in only bovine serum[30] and, although low levels of conglutinin have been detected in human and rat plasma,[33] neither human nor rat conglutinin have yet been purified and characterized. Thus it seems probable that the signal seen in the bovine liver RNA, using the bSP-D probe, may be due to an mRNA transcript for conglutinin especially in view of the high sequence similarity between SP-D and conglutinin, and the fact that bSP-D appears to be lung specific and cannot be detected in serum. It is therefore proposed that bovine conglutinin mRNA cross-reacts with the bSP-D probe and that the major site of synthesis for conglutinin is the liver, which is consistent with the

immunohistochemical data.[32] This view is also supported by the isolation of cDNA clones coding for conglutinin, from a bovine liver cDNA library (J. Lu, unpublished observations).

Recently, rSP-D has been reported to bind to the lipopolysaccharides of Gram-negative bacteria which indicates that SP-D may play an important role in the recognition and clearance of pathogens in the lung (S. F. Kuan, J. Lu, S. Thiel, S. B. Laursen and K. B. M. Reid, personal communication). Bovine conglutinin has been shown to have antibacterial activity[18] and to be able to bind to: iC3b (a breakdown product of the activated third component of complement); the C1q receptor;[16] the gp160 envelope glycoprotein on HIV.[17]

This paper allows comparison of SP-D and conglutinin from the same species and the strong overall structural similarities between the two molecules are suggestive that SP-D may play roles in defence and clearance mechanisms similar to those shown by conglutinin, with SP-D acting in the lung and conglutinin in the serum.

ACKNOWLEDGMENTS

B. L. Lim is supported by the Croucher Foundation and J. Lu was supported by British Council. We are grateful to Mr A. C. Willis for performing the protein sequencing and Mr T. Gascoyne for the synthesis of oligonucleotides.

REFERENCES

1. Possmayer F. (1988) A proposed nomenclature for pulmonary surfactant-associated proteins. *Am. Rev. respir. Dis.* **138,** 990.
2. Hawgood S. & Clements J.A. (1990) Pulmonary surfactant and its apoproteins. *J. clin. Invest.* **86,** 1.
3. Haagsman H.P., White R.T., Schilling J., Lau K., Benson B.J., Golden J., Hawgood S. & Clements J.A. (1989) Studies of the structure of lung surfactant protein SP-A. *Am. J. Physiol.* **257,** L421.
4. Haagsman H.P., Sargeant T., Hauschka P.V., Benson B.J. & Hawgood S. (1990) Binding of Calcium to SP-A, a surfactant-associated protein. *Biochemistry*, **28,** 8894.
5. Glasser S.W., Korfhagen T.R., Weaver T., Pilot-Matias T., Fox J.L. & Whitsett J.A. (1987) cDNA and deduced amino acid sequence of human pulmonary surfactant-associated proteolipid SPL (Phe). *Proc. natl. Acad. Sci. U.S.A.* **84,** 4007.
6. Warr R.G., Hawgood S., Buckley D.I., Crisp T.M. & Schilling J. (1987) Low molecular weight human pulmonary surfactant protein (SP5): isolation, characterization and cDNA and amino acid sequences. *Proc. natl. Acad. Sci. U.S.A.* **84,** 7915.
7. Lu J., Willis A.C. & Reid K.B.M. (1992) Purification, characterization and cDNA cloning of human pulmonary surfactant protein D. *Biochem. J.* **284,** 795.
8. Persson A., Chang D., Rust K., Moxley M., Longmore W. & Crouch E. (1989) Purification and biochemical characterization of CP4(SP-D): a collagenous surfactant-associated protein. *Biochemistry*, **27,** 6361.
9. Persson A., Chang D. & Crouch E. (1990) Surfactant protein D (SP-D) is a divalent cation dependent carbohydrate binding protein. *J. biol. Chem.* **265,** 5755.
10. Rust K., Grosso L., Zhang V., Chang D., Persson A., Longmore W., Cai G-Z. & Crouch E. (1991) Human surfactant protein D: SP-D contains a C-type lectin carbohydrate recognition domain. *Arch. biochem. Biophys.* **290,** 116.
11. Hawgood S., Benson B.J., Schilling J., Damm D., Clements J.A. & White R.T. (1987) Nucleotide and amino acid sequences of pulmonary surfactant protein SP 18 and evidence for cooperation between SP 18 and SP 28–36 in surfactant lipid adsorption. *Proc. natl. Acad. Sci. U.S.A.* **84,** 66.
12. Drickamer K. (1988) Two distinct classes of carbohydrate-recognition domains in animal lectins. *J. biol. Chem.* **263,** 9557.
13. Shimizu H., Fisher J.H., Papst P., Benson B., Lau K., Mason R.J. & Voelker D.R. (1992) Primary structure of rat pulmonary surfactant protein D. *J. biol. Chem.* **267,** 1853.
14. Lee Y.M., Leiby K.R., Allar J., Paris K., Lerch B. & Okarma, T.B. (1991) Primary structure of bovine conglutinin, a member of the C-type animal lectin family. *J. biol. Chem.* **266,** 2715.
15. Loveless R.W., Feizi T., Childs R.A., Mizuochi T., Stoll M.S., Oldroyd R.G. & Lachmann P.J. (1989) Bovine serum conglutinin is a lectin which binds non-reducing terminal N-acetylglucosamine, mannose and fucose residues. *Biochem. J.* **258,** 109.
16. Malhotra R., Thiel S., Reid K.B.M. & Sim R.B. (1990) Human leukocyte C1q receptor binds other soluble proteins with collagen domains. *J. exp. Med.* **172,** 955.
17. Anderson O., Sorensen A.M., Svehag S.E. & Fenouillet E. (1992) Conglutinin binds the HIV-1 envelope glycoprotein gp160 and inhibits its interaction with cell membrane CD4. *Scand. J. Immunol.* (in press).
18. Friis-Christiansen P., Thiel S., Svehag S.E., Dessau R., Svendsen P., Andersen O., Laursen S.B. & Jensenius J.C. (1990) *In vivo* and *in vitro* antibacterial activity of conglutinin, a mammalian plasma lectin. *Scand. J. Immunol.* **31,** 453.
19. Kuan S.F., Rust K. & Crouch E. (1992) Interactions of surfactant protein D with bacterial lipopolysaccharides. *J. Clin. Invest.* **90,** 97.
20. Heery D.M., Gannon F. & Powell R. (1990) A simple method for subcloning DNA fragments from gel slices. *Trends Genet.* **6,** 173.
21. Sambrook J., Fritsch E.F. & Maniatis T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbour Laboratory Press, New York.
22. Chromczynski P. & Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162,**156.
23. Barondes S.H. (1988) Bifunctional properties of lectins: lectins redefined. *TIBS*, **13,** 480.
24. Sano K., Fisher J., Mason R.J., Kuroki Y., Schilling J., Benson B. & Voelker D. (1987) Isolation and sequence of a cDNA clone for the rat pulmonary-associated protein (PSP-A). *Biochem. biophys. Res. Comm.* **144,** 367.
25. Drickamer K., Dordal M.S. & Reynolds L. (1986) Mannose-binding proteins isolated from rat liver contain carbohydrate-recognition domains linked to collagenous tails. Complete primary structures and nomology with pulmonary surfactant apoprotein. *J. biol. Chem.* **261,** 6878.
26. Rucoslahti E. & Pierschbacher M.D. (1987) New perspectives in cell adhesion: RGD and integrins. *Science*, **238,** 491.
27. Dedhar S., Ruoslahti E. & Pierschbacher M.D. (1987) A cell surface receptor complex for collagen type I recognizes the Arg-Gly-Asp sequence. *J. Cell Biol.* **104,** 585.
28. Crouch E., Rust K., Persson A., Marienchеck W., Moxley M. & Longmore W. (1991) Primary translation products of pulmonary surfactant protein D. *Am. J. Physiol.* **260,** L247.
29. Weis W.I., Kahn R., Fourme R., Drickamer K. & Hendrickson W.A. (1992) Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD-phasing. *Science*, **254,** 1608.
30. Strang C.J., Slayter H.S., Lachmann P.J. & Davis A.E. (1986) Ultrastructure and composition of bovine conglutinin. *Biochem. J.* **234,** 381.
31. Persson A., Rust K., Chang D., Moxley M., Longmore W. & Crouch E. (1988). CP4. A pneumocyte-derived collagenous surfactant-associated protein. Evidence for heterogeneity of collagenous surfactant proteins. *Biochemistry*, **27,** 8576.
32. Holmskov U., Teisner B., Pedersen N.T., Laursen S.B., Rasmussen H.B. & Jensenius J.C. (1992) Tissue localization of conglutinin, a bovine C-type lectin. *Immunology*, **76,** 169.
33. Baatrup G., Thiel S., Isager H., Svehag S.-E. & Jensenius J.C. (1987) Demonstration in human plasma of a lectin activity analogous to that of bovine conglutinin. *Scand. J. Immunol.* **26,** 355.