cDNA, Deduced Polypeptide Structure and Chromosomal Assignment of Human Pulmonary Surfactant Proteolipid, SPL(pVa1)*

(Received for publication, February 25, 1987)

Stephan W. Glasser†, Thomas R. Korfhagen, Timothy E. Weaver, Jean C. Clark, Tami Pilot-Matiasj, Joseph Muthij, J. Lawrence Foxj, and Jeffrey A. Whitsett*

From the University of Cincinnati, College of Medicine, Department of Pediatrics, Division of Neonatology, Cincinnati, Ohio 45267-0541 and §Abbott Laboratories, Abbott Park, Illinois 60064

In hyaline membrane disease of premature infants, lack of surfactant leads to pulmonary atelectasis and respiratory distress. Hydrophobic surfactant proteins of \( M_r = 5,000-14,000 \) have been isolated from mammalian surfactants which enhance the rate of spreading and the surface tension lowering properties of phospholipids during dynamic compression. We have characterized the amino-terminal amino acid sequence of pulmonary proteolipids from ether/ethanol extracts of bovine, canine, and human surfactant. Two distinct peptides were identified and termed SPL(pVa) and SPL(pPhe). An oligonucleotide probe based on the valine-rich amino-terminal amino acid sequence of SPL(pVa) was utilized to isolate cDNA and genomic DNA encoding the human protein, termed surfactant proteolipid (SPL(pVa)) on the basis of its unique polyvaline domain. The primary structure of a precursor protein of 20,870 daltons, containing the SPL(pVa) polypeptide, was deduced from the nucleotide sequence of the cDNAs. Hybrid-arrested translation and immunoprecipitation of labeled translation products of human mRNA demonstrated an \( M_r = 22,000 \) precursor protein, the active hydrophobic peptide being produced by proteolytic processing to \( M_r = 5,000-6,000 \). Two classes of cDNAs encoding SPL(pVa) were identified and mRNAs of approximately 900 bases were identified on Northern analysis of fetal and adult RNA. Human SPL(pVa) mRNA was more abundant in the adult than in fetal lung. The SPL(pVa) gene locus was assigned to chromosome 8.

Pulmonary surfactant is composed primarily of the phospholipids phosphatidylcholine, phosphatidylglycerol and lesser amounts of surfactant-associated proteins. Two groups of surfactant proteins have been distinguished on the basis of differential solubility in organic solvent systems. Surfactant-associated protein of \( M_r = 35,000 \) (SAP-35)\(^1\) has been identified as an abundant glycoprotein present in numerous mammalian surfactants (1). SAP-35 is insoluble in ether/ethanol or chloroform/methanol (1, 2) and arises from an mRNA encoding a 23,000-dalton polypeptide containing an approximately 70-amino acid, collagen-like amino-terminal domain (2-4). Hydrophobic small molecular weight proteins soluble in organic solvents have also been detected in a variety of mammalian surfactants (5-12). In recent work from this laboratory, we identified protein of \( M_r = 6,000-14,000 \) in the ether/ethanol extracts of surfactant which were unrelated to SAP-35 or its fragments (11, 12). These same proteins were detected in surfactant extract preparations used clinically for treatment of hyaline membrane disease (11-15). cDNAs encoding one of these proteins, human surfactant proteolipid with amino terminal of phenylalanine, SPL(Phe), an \( M_r = 7,500 \) peptide derived from an \( M_r = 40,000 \) precursor, were recently reported by this laboratory (16). cDNA encoding a protein homologous to SPL(Phe), SP-18, was recently isolated from canine lung (17). Reconstitution of small molecular weight surfactant proteins with synthetic phospholipids in vitro partially complete surfactant-like properties to the mixture, including rapid surface absorption and surface tension lowering during dynamic compression (9, 10, 18). In the present work, we have identified a novel, small molecular weight hydrophobic surfactant protein, herein termed surfactant proteolipid with a polyvaline domain or SPL(pVa).

MATERIALS AND METHODS

T4 DNA ligase and DNA restriction endonucleases were obtained from New England Biolabs, Beverly, MA. Reverse transcriptase was obtained from Life Sciences Inc., St. Petersburg, FL, and used according to manufacturer's recommendations. Escherichia coli strain Y1090 (\( \Delta \) lac U169 pro A\(^+\) \( \Delta \) lon ara D139 strA supF trpC 22::Tn10(mc9)) purchased from Clontech Laboratories, Palo Alto, CA, was used as the host strain for \( \lambda \) gt11. E. coli JM103 or JM109 (Pharmacia LKB Biotechnology Inc.) was used for growth of pUC plasmid and M13 phage subclones. The phage cloning vector \( \lambda \) gt11 was used for the construction of a cDNA library as described by Young and Davis (19). pUC 19 and M13 were used for subcloning and DNA sequencing as described by Messing (20). The \( \lambda \) gt11 library was constructed from human lung poly(A)\(^+\) RNA from an adult male from tissue obtained immediately at death as previously described (18). The tissue was provided by the National Diabetes Tissue Interchange, Washington, D.C. Wheat germ translation reagent was purchased from Promega Biotech Inc. and [\( ^35 \)S]methionine from Du Pont-New England Nuclear. A human genomic lymphocyte library in \( \lambda \) EMBL3 was purchased from Clontech Laboratories, Inc., Madison, WI.

Protein Purification—Hydrophobic surfactant proteins were purified from ether/ethanol extracts of human, canine, and bovine surfactant obtained by alveolar lavage as previously described (11, 12). Proteins were then delipidated by sillic acid chromatography in 1 M NaCl.

* This work was supported in part by National Institutes of Health Grants HL 10124 and HL28263 (to J. A. W.), GM10799 (to T. R. W.), and HL36055 (to T. R. W.), and by Children's Hospital Research Foundation and Ross Laboratories, Columbus, Ohio. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank\(^\text{TM}/\text{EMBL} \) Data Bank with accession number(s) J03617.

† Supported by the Graduate Program in Developmental Biology, University of Cincinnati.

‡ To whom correspondence should be addressed: University of Cincinnati College of Medicine, Department of Pediatrics, Division of Neonatology, 231 Bethesda Ave., Cincinnati, OH 45267-0541.

\(^1\) The abbreviations used are: SAP-35, surfactant-associated protein of \( M_r = 35,000 \); SPL(Phe), human surfactant proteolipid with amino terminal of phenylalanine; SPL(pVa), surfactant proteolipid of \( M_r \), approximately 7000-6000 containing polyvaline; SDS, sodium dodecyl sulfate; Pipes, 1,4-piperazinediethanesulfonic acid; PAGE, polyacrylamide gel electrophoresis.
chloroform/methanol, followed by extensive dialysis in chloroform/methanol. Silver stain analysis of these preparations after SDS-PAGE revealed protein of M, approximately 5,000-6,000 as the most abundant protein; progressively smaller amounts of protein of M, approximately 14,000, 18,000, and 26,000 were observed (increased amounts of larger proteins were observed in panels previously described by Griff et al. (28) using essentially identical methodology and were kindly performed by Webster Cavenee, Ludwig Institute, Montreal, Canada.

RESULTS AND DISCUSSION

Amino Acid Sequence—Amino acid sequence analysis of the human, bovine, and canine SPL(pVal) proteins resulted in the identification of the following sequences. Ile-Pro-Cys-Phe-Pro-Ser-Leu-Lys-Arg-Leu-Leu-Ile-Val-Val-Val-Val-Val-Val-Val-Val-Val-Val-Val-Val-Val (canine); Ile-Pro-Cys-Pro-Val-Asn-Ile-Lys-Arg-Leu-Ile-Val-Val-Val-Val-Val-Val-Val (bovine) and Ile-Pro-Cys-Pro-Val-Asn-Leu-Lys-Arg-Leu-Leu-Ile-Val-Val-Val-Val-Val-Val (human). Leucine and glycine were detected in the first cleavage in numerous analyses. A second amino acid sequence was observed during analysis of the amino terminus of the human protein: Phe-Pro-Ile-Pro etc.

cDNAs isolated for this distinct, hydrophobic surfactant protein, SPL(Phe), were recently reported by this laboratory (16).

Cloning of cDNA for Human SPL(pVal)—An oligonucleotide probe based on the polyvaline sequence was utilized to screen a agt11 expression library generated from human lung poly(A)+ RNA. Nucleotide sequence analysis of one clone (334.2) of 0.3 kilobases comprised an open reading frame predicting close identity to the amino acid sequence determined directly from the human protein and was used to isolate other clones from the same cDNA library (Fig. 1). Sequence analysis of nine unique clones resulted in a consensus sequence predicting a larger polypeptide precursor. Two distinct classes of cDNAs were detected by sequence analysis differing by the absence of 18 bases (463 to 480) in the 3’-coding region in clones 311.3 and 13.1 compared to other clones including TP11.2. The Ile-Pro-Cys-Cys peptide was found within the reading frame of a larger polypeptide suggesting that the hydrophobic peptide of M, = 5,000-6,000 arises from proteolytic processing of a precursor protein at both the amino and carboxyl terminus (Fig. 2). Both clone 334.2 and clone Tp11.2 hybridized to a single 850-9-kilobase RNA after Northern

Dideoxy Nucleotide RNA-directed Sequencing—Dideoxynucleotide RNA-directed sequencing was done by a modification of the procedure described by Geliebter et al. (27).

Chromosomal Assignment—The 32P-labeled SPL(pVal) cDNA (334.2) was hybridized to DNA obtained from the mouse-human chromosomal panels previously described by Griff et al. (28) using essentially identical methodology and were kindly performed by Webster Cavenee, Ludwig Institute, Montreal, Canada.

**Fig. 1.** Composite cDNA map of human SPL(pVal) clones. Clone 334.2 is the initial SPL(pVal) isolate and was used as a probe for rescreening of the cDNA library. Clones 311.3 (and a second clone, 13-1) have an 18-base pair deletion not seen in clones TP9-1 and TP11-2. The hatched box indicates the valine-rich hydrophobic domain of SPL(pVal) clones. A denotes Apa L1, a restriction endonuclease that cleaves phage lambda DNA infrequently and cuts at the start of the valine-rich domain of SPL(pVal). P, PstI; S, SmaI. * indicates sequence obtained from human lung mRNA-directed dideoxy sequencing utilizing 5’ SPL(pVal) oligo primers.
poly(A) + RNA in 80% deionized formamide, 10 mM Pipes, pH 6.4, translated in wheat germ extract. Translation of SPL(pVa1), terminated by the addition of 4 volumes of water containing 25 pg of bean trypsin inhibitor, blot analysis of human lung RNA. Abundance of the RNA arrested by SPL(pVa1) cDNA hybrid form at 0 °C. Both samples were ethanol-precipitated and one of which was melted, and the other which was preserved in the restricted SPL(pVa1) cDNA was hybridized with 5 pg of human lung SPL(pVa1) protein. Directed dideoxy sequencing using a synthetic oligonucleotide based on sequence from clone RJ-21 as a primer. This sequence is present mined using overlapping cDNA clones 311.3, 13-1, TPll-2, TP9-1, 22,000, was allowed to proceed in the melted sample produced amino acid sequence. The predicted sequence that is absent in cDNA clones 311.3 and 13-1. The predicted sequence and obtained sequence match at 16 of 17 amino acids, the sequence that is present underlined.

2. Identification of primary translation product of SPL(pVa1) RNA. Hybrid-arrested and -selected translation of SPL(pVa1) was utilized to identify the polypeptide precursor. EcoRI-restricted SPL(pVa1) cDNA was hybridized with 5 pg of human lung poly(A) + RNA in 80% deionized formamide, 10 mM Pipes, pH 6.4, 0.25 mM EDTA, and 0.4 M NaCl for 2 h at 50 °C. Hybridization was terminated by the addition of 4 volumes of water containing 25 μg of yeast tRNA. RNA-DNA hybrids were divided into two equal samples, one of which was melted, and the other which was preserved in the hybrid form at 0 °C. Both samples were ethanol-precipitated and translated in wheat germ extract. Translation of SPL(pVa1), M, = 22,000, was allowed to proceed in the melted sample (lane 1) but was arrested by SPL(pVa1) cDNA (lane 2). Lane 3, 14C-methylated soybean trypsin inhibitor, M, = 21,500.

2. FIG. 3. Identification of primary translation product of SPL(pVa1) RNA. Hybrid-arrested and -selected translation of SPL(pVa1) was utilized to identify the polypeptide precursor. EcoRI-restricted SPL(pVa1) cDNA was hybridized with 5 pg of human lung poly(A) + RNA in 80% deionized formamide, 10 mM Pipes, pH 6.4, 0.25 mM EDTA, and 0.4 M NaCl for 2 h at 50 °C. Hybridization was terminated by the addition of 4 volumes of water containing 25 μg of yeast tRNA. RNA-DNA hybrids were divided into two equal samples, one of which was melted, and the other which was preserved in the hybrid form at 0 °C. Both samples were ethanol-precipitated and translated in wheat germ extract. Translation of SPL(pVa1), M, = 22,000, was allowed to proceed in the melted sample (lane 1) but was arrested by SPL(pVa1) cDNA (lane 2). Lane 3, 14C-methylated soybean trypsin inhibitor, M, = 21,500.

3. FIG. 4. Hydropathy analysis of the SPL(pVa1) precursor. Hydropathy analysis using a span of 11 amino acids was performed on the entire predicted amino acid sequence according to the procedure of Kyte and Doolittle (30). Hydropophicity is plotted as a function of residue number from Met1 to the carboxyl terminus, lle229. Values indicating hydrophobic and hydrophilic regions are above and below 0.5, respectively, as indicated by the horizontal dotted line.
consistent in size with the hybrid-selected translation product of \( M_r = 22,000 \). Hydropathy analysis of the precursor protein was determined using the methods described by Kyte and Doolittle (30) (Fig. 4). There was no discernible signal peptide at the amino terminus, and the precursor polypeptide contained no asparagine-linked glycosylation sites, contrasting with the SPL(Phe) precursors which contain one or two potential asparagine-linked glycosylation sites (16). The SPL(pVal) peptide begins at Ile\(^{36}\) and the domain including amino acids Leu\(^{37}\) to Ser\(^{43}\) is compatible with a membrane-associated or spanning domain of 25 amino acids. This region contains the repeated valine residues. The precise carboxyl terminus of the SPL(pVal) has not been identified directly, and numerous attempts to isolate proteolytic or CNBr fragments of the canine or bovine proteolipid have been unsuccessful.

The identification of the polypeptide sequence of this hydrophobic protein completely distinguishes the SPL(pVal) from SPL(Phe) and surfactant-associated protein of 35,000 daltons, SAP-35. Small molecular weight protein with composition consistent with the protein previously termed "surfactant apolipoprotein B" (1) has recently been identified as a carboxyl-terminal domain of SAP-35 (21). Mixtures of phospholipids with SAP-35 or its fragments were found to have relatively weak surface active properties compared to the surfactant proteolipids \( M_r = 6,000-14,000 \) (11, 21). Canine and bovine SPL(pVal) alone conferred virtually full biophysical activity to phospholipid mixtures (18). Previous studies suggesting the importance of SAP-35 alone in conferring this surfactant activity may have been confounded by the presence of the surfactant proteolipids in lipid extracts used in those reconstitution studies (31). The present work supports previous views in which small molecular weight surfactant proteins were distinguished from SAP-35 on the basis of amino acid composition and immunoreactivity (11, 12, 32).

**Chromosomal Assignment**—The \(^{32}\)P-labeled SPL(pVal) clone (334.2) was hybridized to mouse-human chromosomal hybrids containing all human chromosomes as previously characterized (28). Hybridization was observed with hybrids containing only human chromosome 8 (data not shown).

**Summary**—The present work identifies a human hydrophobic surfactant polypeptide, SPL(pVal), its polypeptide precursor, and complete mRNA sequence. The SPL(pVal) peptide and the control of SPL(pVal) expression during lung development may be useful in diagnosis and treatment of hyaline membrane disease and other pulmonary disorders associated with surfactant deficiency.

**Acknowledgment**—We acknowledge Webster Cavenee, Ludwig Institute, Montreal, Canada, for assistance in chromosomal assignment.

**REFERENCES**

1. King, R. J., Klaas, D. J., Gikas, E., and Clemetsa, J. A. (1973) *Am. J. Physiol.* 224, 788-795
2. Ross, G. F., Meuth, J., Ohning, B., Yangkil, K., and Whitsett, J. A. (1986) *Biochem. Biophys. Acta* 870, 267-279
3. White, R. T., Damm, D., Miller, J., Spratt, K., Schilling, J., Hawgood, S., Benson, B., and Corbell, B. (1985) *Nature* 317, 361-362
4. Whitsett, J. A., Hull, W., Ross, G., and Weaver, T. (1985) *Pediatr. Res.* 19, 501-508
5. Phizackerley, P. J., Town, M. H., and Neuman, G. E. (1979) *Biochem. J.* 183, 731-736
6. Takeya, Y., Takai, T., and Kanazawa, Y. (1983) *Chem. Pharm. Bull.* 31, 4100-4109
7. Suzuki, Y. (1982) *J. Lipid Res.* 23, 62-69
8. Claypool, W. D., Wang, D. L., Chander, A., and Fisher, A. B. (1984) *J. Clin. Invest.* 74, 677-684
9. Yu, S., and Posmayer, F. (1986) *Biochem. J.* 236, 85-89
10. Takahashi, A., and Fujiwara, T. (1990) *Biochem. Biophys. Res. Commun.* 175, 527-532
11. Whitsett, J. A., Notter, R. H., Ohning, B. L., Ross, G., Meuth, J., Holm, B. A., Shapiro, D. L., and Weaver, T. E. (1986) *Pediatr. Res.* 20, 463-467
12. Whitsett, J. A., Hull, W. M., Ohning, B. L., Ross, G., and Weaver, T. E. (1986) *Pediatr. Res.* 20, 744-749
13. Fujiwara, T., Maeta, H., Chida, S., Morita, T., Watabe, Y., and Ake, T. (1980) *Lancer* I, 55-59
14. Enhorning, G., Shennan, A., Posmayer, P., Dunn, M., Chen, C. P., and Milligan, J. (1985) *Pediatrics* 76, 145-153
15. Kwong, M. S., Ren, E. A., Notter, R. H., and Shapiro, D. L. (1985) *Pediatrics* 76, 686-692
16. Glasser, S. W., Kortzagen, H., Weaver, T. E., Pilot, T., Fox, J. L., and Whitsett, J. A. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 4007-4011
17. Hawgood, S., Benson, B. J., Schilling, J., Damm, D., Clements, J. A., and White, R. T. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 168-170
18. Notter, R. H., Shapiro, D. L., Ohning, B., and Whitsett, J. A. (1987) *Chem. Phys. Lipids* 44, 1-17
19. Young, R. A., and Davis, R. W. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 1194-1198
20. Messing, J. (1983) *Methods Enzymol.* 101, 70-78
21. Gertler, F. S., and Cavenee, W. K. (1985) *Methods Enzymol.* 101, 70-78
22. Brown, M. S., Klaas, D. J., and Saltis, J. A. (1986) *Biochemistry* *25*, 5729-5735
23. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
24. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) *Biochemistry* *18*, 5294-5299
25. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, pp. 125-126, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
26. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, pp. 125-126, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
27. Kushner, J. Z., Zeff, R. A., Melvold, R. W., and Nathanson, S. G. (1990) *Proc. Natl. Acad. Sci. U. S. A.* 85, 3371
28. Griffin, C., Ensmann, B. S., Hansen, J. R., Cavenee, W. K., and Myers, J. R. (1987) *Proc. Natl. Acad. Sci. U. S. A.* 84, 512-516
29. Korace, M. (1994) *Nucleic Acids Res.* 22, 87-872
30. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* 157, 105-132
31. Hawgood, S., Benson, B. J., and Hamilton, R. L. (1986) *Biochemistry* *25*, 184-190
32. Yu, S., and Posmayer, F. (1986) *Biochem. J.* 236, 85-89