Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 4007–4011, June 1987
Biochemistry

cDNA and deduced amino acid sequence of human pulmonary surfactant-associated proteolipid SPL(Phe)

(lung/hyaline membrane disease)

STEPHAN W. GLASSER*, THOMAS R. KORFHAGEN*, TIMOTHY WEAVER*, TAMI PILOT-MATIAS†, J. LAWRENCE FOX†, AND JEFFREY A. WHITSETT*‡

*Department of Pediatrics, University of Cincinnati, College of Medicine, Cincinnati, OH 45267-0541; and †Abbott Laboratories, Abbott Park, IL 60064

Communicated by David Shemin, February 19, 1987

ABSTRACT **Hydrophobic surfactant-associated protein of M_r 6000–14,000 was isolated from ether/ethanol or chloroform/methanol extracts of mammalian pulmonary surfactant. Automated Edman degradation in a gas-phase sequencer showed the major N-terminus of the human low molecular weight protein to be Phe-Pro-Ile-Pro-Leu-Pro-Tyr-Cys-Trp-Leu-Cys-Arg-Ala-Leu-. Because of the N-terminal phenylalanine, the surfactant protein was designated SPL(Phe). Antiserum generated against hydrophobic surfactant protein(s) from bovine pulmonary surfactant recognized protein of M_r 6000–14,000 in immunoblot analysis and was used to screen a λgt11 expression library constructed from adult human lung poly(A)$^+$ RNA. This resulted in identification of a 1.4-kilobase cDNA clone that was shown to encode the N-terminus of the surfactant polypeptide SPL(Phe) (Phe-Pro-Ile-Pro-Leu-Pro-) within an open reading frame for a larger protein. Expression of a fused β-galactosidase-SPL(Phe) gene in *Escherichia coli* yielded an immunoreactive M_r 34,000 fusion peptide. Hybrid-arrested translation with this cDNA and immunoprecipitation of [^{35}S]methionine-labeled *in vitro* translation products of human poly(A)$^+$ RNA with a surfactant polyclonal antibody resulted in identification of a M_r 40,000 precursor protein. Blot hybridization analysis of electrophoretically fractionated RNA from human lung detected a 2.0-kilobase RNA that was more abundant in adult lung than in fetal lung. The larger RNA and translation product indicates that SPL(Phe) is derived by proteolysis of a large polypeptide precursor. The amino acid sequence of the predicted protein, beginning Phe-Pro-Ile-Pro-Leu-Pro-Tyr-, comprises a hydrophobic peptide that is a major protein component of surfactant lipid extracts used successfully to treat hyaline membrane disease in newborn infants. These proteins, and specifically SPL(Phe), may therefore be useful for synthesis of replacement surfactants for treatment of hyaline membrane disease in newborn infants or of other surfactant-deficient states.**

Pulmonary surfactant is composed primarily of phospholipids (phosphatidylcholine and phosphatidylglycerol) and lesser amounts of surfactant-associated apoproteins. Two groups of surfactant apoproteins have been previously distinguished on the basis of their solubility in organic solvent systems. Protein of M_r 35,000 (SAP-35) has been identified in numerous mammalian surfactants and is relatively insoluble in ether/ethanol or chloroform/methanol solvent systems (1, 2). Human SAP-35 is a glycoprotein arising from an mRNA encoding a M_r 23,000 polypeptide that includes a hydroxyproline-containing, collagen-like N-terminal region of ≈70 amino acid residues (3, 4). Hydrophobic proteins or peptides soluble in organic solvents have also been isolated from mammalian surfactants (5–12). In recent work in this laboratory, these hydrophobic peptides were shown to be unrelated to SAP-35 or its fragments (11, 12). These preparations contained protein migrating primarily at M_r 6000–14,000 on NaDodSO$_4$/PAGE, and two distinct N-terminal amino acids were observed during amino acid sequencing (12). Reconstitution of hydrophobic surfactant peptides with synthetic phospholipids imparts virtually complete surfactant-like properties to synthetic phospholipids, including rapid surface adsorption and lowering of surface tension during dynamic compression (9, 10, 13).

Surfactant lipid extracts containing the two hydrophobic peptides as the sole apoproteins (11, 12) have been used for therapy in hyaline membrane disease in newborn infants (14–16). Two such preparations, CLSE and surfactant-TA, contain only the M_r 6000–14,000 proteins and do not contain the larger SAP-35 apoprotein (11, 12). The present work describes the identification and cloning of cDNA of one of the hydrophobic surfactant-associated proteolipids [SPL(Phe), so-called because of its N-terminal sequence Phe-Pro-Ile-Pro-Leu-Pro-Tyr-] and provides the deduced amino acid sequence for SPL(Phe) protein.

MATERIALS AND METHODS

Materials. Bacteriophage T4 DNA ligase and DNA-restriction endonucleases were obtained from New England Biolabs. Reverse transcriptase (RNA-directed DNA polymerase) was obtained from Life Sciences (St. Petersburg, FL) and used according to the manufacturer's recommendations. Horseradish peroxidase-conjugated goat anti-rabbit IgG was obtained from Miles. *Escherichia coli* strains Y1090 [Δ*lacU169 proA*$^+$ Δ*lon araD139 strA supF trpC22*::Tn*10* (pMC9)], and Y1088 [Δ*lacU169 supE supF hsdR hsdM*$^+$ *metB trpR tonA21 proC*::Tn*5* (pMC9 = pBR322)] were used as host strains for λgt11. These strains were described more fully by Young and Davis (17). *E. coli* strain JM103, described by Messing *et al.* (18), was used for growth of pUC plasmid and M13 phage subclones. The phage cloning vector λgt11 was used for construction of a cDNA library as described by Young and Davis (19); the library was constructed by Clontech (Palo Alto, CA). Plasmid pUC19 and bacteriophage M13mp11 were purchased from Pharmacia and were used for subcloning and DNA sequencing as described by Messing (20) and Yanisch-Perron *et al.* (21). The λgt11 library was constructed from human lung poly(A)$^+$ RNA prepared, according to the method of Chirgwin *et al.* (22), from tissue obtained from an adult male immediately at death. Tissue was provided by the National Diabetes Tissue Interchange (Washington, DC). Rabbit reticulocyte lysates were purchased from Promega Biotec (Madison, WI).

Protein Purification. Hydrophobic surfactant peptides were purified from ether/ethanol or chloroform/methanol extracts of human or bovine lung lavage surfactant as described (11, 12).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡To whom reprint requests should be addressed at: University of Cincinnati, College of Medicine, 231 Bethesda Avenue, Cincinnati, OH 45267-0541.

Bovine and canine proteins were obtained from lung lavage surfactant at sacrifice and human protein was isolated from lung lavage of human cadavers at the University of Cincinnati College of Medicine. Protein was delipidated by silicic acid chromatography in chloroform/methanol, followed by extensive dialysis in chloroform/methanol. After electrophoresis of these preparations, silver stain revealed a diffuse protein band at M_r 6000 as the primary protein; smaller amounts of protein of M_r 14,000 were also observed in these preparations, as reported previously (11, 12).

Amino Acid Sequence Analysis. Automated Edman degradation was performed on the delipidated surfactant extracts, using an Applied Biosystems (Foster City, CA) model 470A vapor-phase protein sequencer. In every analysis, samples were deposited on Polybrene-treated glass-fiber filters. Phenylthiohydantoin derivatives of amino acids were identified by HPLC on an Altex Ultrasphere ODS column (4.5 × 250 mm) preceded by a precolumn (4.6 × 45 mm) of the same resin, as previously described (23).

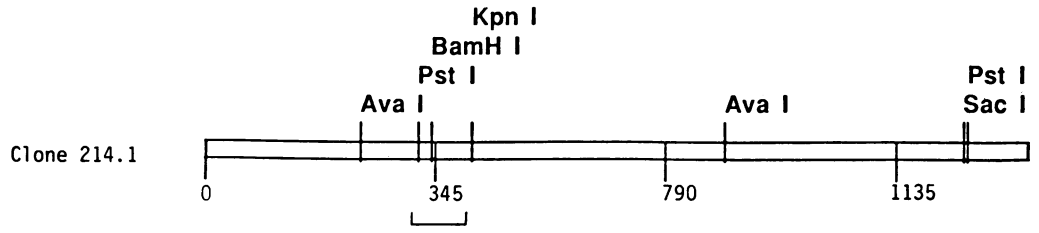
Immunologic Assays. Antisera were generated against the hydrophobic surfactant proteins in New Zealand White rabbits by repeated injection of a chloroform/methanol extract of bovine surfactant. The resulting antisera specifically recognized canine, human, and bovine surfactant protein by ELISA assay and immunoblot analysis and did not recognize serum proteins or SAP-35 (11, 12). The antibody titer was $>10^6$ against the bovine proteolipid in a single-phase ELISA assay. To optimize reactivity in the ELISA, the hydrophobic proteins were added to the assay plate (Falcon, polyvinylchloride) in 100% methanol and incubated overnight at −30°C. The human lung cDNA library was screened essentially as described by Young and Davis (19). To minimize nonspecific background in the screening procedure, the antiserum was preabsorbed with *E. coli* Y1090 lysate. The lysate was obtained by treatment of Y1090 cells with lysozyme and EDTA (pH 8.0) followed by sonication on ice with a Branson sonifier (six bursts at 50 watts, 10 sec per burst). The sonicate was centrifuged at low speed to clear the supernatant, which then was preincubated with the SAP-6 antiserum for 48 hr at 4°C. The library was plated at 20,000 plaque-forming units per 150-mm dish, grown for 5 hr at 42°C, and then overlaid with nitrocellulose filters soaked in 10 mM isopropyl β-D-thiogalactoside. Filters were incubated overnight at 37°C, and nonspecific binding sites were blocked by incubation for 1 hr at 4°C in 50 mM Tris·HCl, pH 7.4/150 mM NaCl containing powdered milk at 50 mg/ml. Primary antibody at a dilution of 1:1000 was incubated with the filters at 4°C overnight. Horseradish peroxidase-conjugated goat anti-rabbit IgG (the second antibody) at a dilution of 1:3000 was incubated with the filters at 4°C for 16 hr. Color was developed with 4-chloro-1-naphthol for 3–5 min, and reactions were terminated by serial washings in distilled H_2O to identify isolated plaques. For second and third screenings, phage were plated at 2000 and 200 plaque-forming units per 150-mm dish, respectively.

Oligonucleotide Probe Synthesis. An oligodeoxyribonucleotide probe was made by the phosphoramidite method described by Matteucci and Caruthers (24). The sequence was 5′ CCAICAITAIGGIATIGGIATIGGIAA 3′, where I stands for deoxyinosine. This analog was incorporated into positions that were ambiguous based on the reverse translation of the amino acid sequence, since deoxyinosine can base-pair with any of the four naturally occurring bases. This technique was more fully described by Ohtsuka *et al.* (25).

RESULTS

Peptide Sequence. Two distinct protein sequences were observed in amino acid sequencing of the human proteins. The sequence Phe-Pro-Ile-Pro-Leu-Pro-Tyr-Cys-Trp-Leu-Cys-Arg-Ala-Leu was detected as the major peptide, representing approximately two-thirds of the amino acid sequence obtained from the human proteolipid preparation. A second distinct N-terminal sequence, Ile-Pro-Cys-Cys-, was also observed; more extensive sequence analysis of this sequence from bovine and canine preparation resulted in definition of this as a distinct protein that did not overlap with the SPL(Phe) sequence. Detailed analysis of this second surfactant protein, its cDNA, and derived sequence will be described elsewhere.

Immunologic Screening. From approximately 220,000 clones, 22 positive plaques were detected in the first immunologic screening, of which 13 remained positive in the second screening. Of the 13 phage clones, 5 hybridized with the oligonucleotide probe synthesized on the basis of a reverse translation of the N-terminal sequence of SPL(Phe) as was known at the time. This sequence was Phe-Pro-Ile-Pro-Ile-Pro-Tyr-Cys-Trp. Subsequent amino acid sequence analysis revealed that the fifth residue was leucine rather than isoleucine. However, this change resulted in only a single base mismatch between the oligonucleotide probe and the actual DNA sequence, and thus the probe was still usable. This probe was used to further identify the clones isolated by immunologic screening. Two distinct clones that encode the SPL(Phe) peptide sequence were isolated initially: 214.1 (1.4 kilobases, Fig. 1) and 219.1 (0.4 kilobase). The nucleotide sequences of these two clones are virtually identical. Both represent open reading frames that indicate the Phe-Pro-Ile-Pro sequence was derived from cleavage of a Gln-Phe bond in a larger preprotein (Fig. 2). The upstream sequence of




```
1: TTC CCC ATT CCT CTC CCC TAT TGC TGG CTC TGC AGG GCT CTG ATC AAG CGG ATC ...

2: Phe Pro Ile Pro Leu Pro Tyr Cys Trp Leu Cys Arg Ala Leu Ile Lys Arg Ile ...

3: Phe Pro Ile Pro Leu Pro Try Cys Trp Leu Cys Arg Ala Leu
```

FIG. 1. (*Upper*) Restriction map of human SPL(Phe) clone 214.1. The SPL(Phe) cDNA of 1408 bases was subjected to restriction enzyme mapping. (*Lower*) The DNA sequence encoding the N-terminal sequence of SPL(Phe) is shown in line 1; the predicted amino acid sequence is shown in line 2 and compared with that determined for human SPL(Phe) by direct amino acid sequence analysis (line 3).

```
GAA TTC CGG GAG TGC AAC GTC CTC CCC TTG AAG CTG CTC ATG CCC CAG TGC AAC CAA GTG CTT GAC GAC TAC TTC CCC CTG GTC ATC GAC    90
Glu Phe Arg Glu Cys Asn Val Leu Pro Leu Lys Leu Leu Met Pro Gln Cys Asn Gln Val Leu Asp Asp Tyr Phe Pro Leu Val Ile Asp

TAC TTC CAG AAC CAG ACT GAC TCA AAC GGC ATC TGT ATG CAC CTG GGC CTG TGC AAA TCC CGG CAG CCA GAG CCA GAG CAG GAG CCA GGG   180
Tyr Phe Gln Asn Gln Thr Asp Ser Asn Gly Ile Cys Met His Leu Gly Leu Cys Lys Ser Arg Gln Pro Glu Pro Glu Gln Glu Pro Gly

ATG TCA GAC CCC CTG CCC AAA CCT CTG CGG GAC CCT CTG CCA GAC CCT CTG CTG GAC AAG CTC GTC CTC CCT GTG CTG CCC GGG GCC CTC   270
Met Ser Asp Pro Leu Pro Lys Pro Leu Arg Asp Pro Leu Pro Asp Pro Leu Leu Asp Lys Leu Val Leu Pro Val Leu Pro Gly Ala Leu

CAG GCG AGG CCT GGG CCT CAC ACA CAG GAT CTC TCC GAG CAG CAA TTC CCC ATT CCT CTC CCC TAT TGC TGG CTC TGC AGG GCT CTG ATC   360
Gln Ala Arg Pro Gly Pro His Thr Gln Asp Leu Ser Glu Gln Gln Phe, Pro Ile Pro Leu Pro Tyr Cys Trp Leu Cys Arg Ala Leu Ile

AAG CGG ATC CAA GCC ATG ATT CCC AAG GGT GCG CTA CGT GTG GCA GTG GCC CAG GTG TGC CGC GTG GTA CCT CTG GTG GCG GGC GGC ATC   450
Lys Arg Ile Gln Ala Met Ile Pro Lys Gly Ala Leu Arg Val Ala Val Ala Gln Val Cys Arg Val Val Pro Leu Val Ala Gly Gly Ile

TGC CAG TGC CTG GCT GAG CGC TAC TCC GTC ATC CTG CTC GAC ACG CTG CTG GGC CGC ATG CTG CCC CAG CTG GTC TGC CGC CTC GTC CTC   540
Cys Gln Cys Leu Ala Glu Arg Tyr Ser Val Ile Leu Leu Asp Thr Leu Leu Gly Arg Met Leu Pro Gln Leu Val Cys Arg Leu Val Leu
                                                                          63

CGG TGC TCC ATG GAT GAC AGC GCT GGC CCA AGG TCG CCG ACA GGA GAA TGG CTG CCG CGA GAC TCT GAG TGC CAC CTC TGC ATG TCC GTG   630
Arg Cys Ser Met Asp Asp Ser Ala Gly Pro Arg Ser Pro Thr Gly Glu Trp Leu Pro Arg Asp Ser Glu Cys His Leu Cys Met Ser Val

ACC ACC CAG GCC GGG AAC AGC AGC GAG CAG GCC ATA CTA CAG GCA ATG CTC CAG GCC TGT GTT GGC TCC TGG CTG GAC AGG GAA AAG TGC   720
Thr Thr Gln Ala Gly Asn Ser Ser Glu Gln Ala Ile Leu Gln Ala Met Leu Gln Ala Cys Val Gly Ser Trp Leu Asp Arg Glu Lys Cys

AAG CAA TTT GTG GAG CAG CAC ACG CCC CAG CTG CTG ACC CTG GTG CCC AGG GGC TGG GAT GCC CAC ACC ACC TGC CAG GCC CTC GGG GTG   810
Lys Gln Phe Val Glu Gln His Thr Pro Gln Leu Leu Thr Leu Val Pro Arg Gly Trp Asp Ala His Thr Thr Cys Gln Ala Leu Gly Val

TGT GGG ACC ATG TCC AGC CCT CTC CAG TGT ATC CAC AGC CCC GAC CTT TGA TGA GAA CTC AGC TGT CCA GCT GCA AAG GAA AAG CCA AGT   900
Cys Gly Thr Met Ser Ser Pro Leu Gln Cys Ile His Ser Pro Asp Leu Non Non

GAG ACG GGC TCT GGG ACC ATG GTG ACC AGG CTC TTC CCC TGC TCC CTG GCC CTC GCC AGC TGC CAG GCT GAA AAG AAG CCT CAG CTC CCA   990

CAC CGC CCT CCT CAC CGC CCT TCC TCG GCA GTC ACT TCC ACT GGT GGA CCA CGG GCC CCC AGC CCT GTG TCG GCC TTG TCT GTC TCA GCT  1080

CAA CCA CAG TCT GAC ACC AGA GCC CAC TTC CAT CCT CTC TGG TGT GAG GCA CAG CGA GGG CAG CAT CTG GAG GAG CTC TGC AGC CTC CAC  1170

ACC TAC CAC GAC CTC CCA GGG CTG GGC TCA GGA AAA ACC AGC CAC TGC TTT ACA GGA CAG GGG GTT GAA GCT GAG CCC CGC CTC ACA CCC  1260

ACC CCC ATG CAC TCA AAG ATT GGA TTT TAC AGC TAC TTG CAA TTC AAA ATT CAG AAG AAT AAA AAA TGG GAA CAT ACA GAA CTC TAA AAG  1350

ATA GAC ATC AGA AAT TGT TAA GTT AAG CTT TTT CAA AAA ATC AGC AAT TCC GGA ATT C                                            1408
```

FIG. 2. Nucleotide sequence and predicted amino acid sequence of clone 214.1. SPL(Phe) begins at nucleotide 316 of the clone (underlined). Non, nonsense (termination) codon.

214.1 differs at a single nucleotide (position 107), so that the isoleucine codon ATT in 219.1 is changed to the threonine codon ACT in 214.1 (Fig. 2). The cDNA encoding SPL(Phe) is contained within a larger cDNA encoding a precursor protein predicted from the open reading frame of the cDNA in both clones. The predicted amino acid sequence and the N-terminal amino acid sequence of the human protein are identical (Fig. 1). The C-terminus of the peptide is predicted from the presence of stop codons in the cDNA, which would result in the synthesis of a peptide of 181 amino acid residues. There is no detectable protein of this size in the surfactant extracts, suggesting that proteolytic processing at both the C- and the N-terminal regions of the preprotein accounts for the presence of the M_r 6000–14,000 material observed upon $NaDodSO_4$/PAGE.

Expression of a Fused β-Galactosidase-SPL(Phe) Peptide. The 1.4-kilobase *Eco*RI insert of clone 214.1 was subcloned into the plasmid pUC19. The reading frame was altered by cutting with restriction endonuclease *Sal* I within the plasmid multicloning site, filling in the ends with T4 DNA polymerase and deoxynucleotides, and religating (Fig. 3 *Right*). This construction gave a clone that expressed a M_r 32,000–34,000 fusion protein identified by immunoblot analysis with the antibody (Fig. 3, lanes c and d). The 214.1 insert in the same expression plasmid but out of reading frame (Fig. 3, lane b) showed no detectable immunoreactive protein. The size of this immunoreactive fusion protein indicates that clone 214.1 generates a larger precursor peptide containing the SPL(Phe) protein.

Primary Translation Product. Adult human lung poly(A)$^+$ RNA was isolated by oligo(dT) column chromatography, using methods described by Aviv and Leder (26), and was used in a cell-free translation assay. The [^{35}S]methionine-labeled protein migrated at M_r 40,000, and immunoprecipitation of the M_r 40,000 protein was specifically inhibited by the addition of excess bovine hydrophobic protein. Translation of this M_r 40,000 protein was inhibited by the addition of cDNA from clone 214.1 (Fig. 4). Thus, the cDNA sequence, the size of the fusion protein, and the size of the immunoreactive primary translation product are consistent with the synthesis of a larger precursor peptide for SPL(Phe). Synthesis of the M_r 40,000 immunoreactive translation product was not observed after *in vitro* translation of rat hepatic poly(A)$^+$ RNA (data not shown).

Blot Hybridization Analysis. RNA was extracted from human lung by the method of Chirgwin *et al.* (22). Blot analysis of electrophoretically fractionated adult human lung RNA resulted in detection of a 2.0-kilobase RNA and faint hybridization to a 1.7-kilobase RNA. Human fetal lung (approximately 19 weeks of gestation) contained much less SPL(Phe) RNA than adult lung (data not shown).

Hydropathy Plot of the Deduced Amino Acid Sequence. Hydropathy analysis, by the method of Kyte and Doolittle (27), of the predicted amino acid sequence is represented by Fig. 5. The sequence beginning Phe-Pro-Ile-Pro represents an N-terminal charged domain followed by an extremely hydrophobic domain compatible with a transmembrane or membrane-associated protein. The region from Trp-9 through Val-29 readily fits a model of an amphipathic helix, the charged residues lying in a distinct axis from the hydrophobic residues (28). Amino acid residues Gly-44 to Arg-64 and Arg-72 to Trp-92 also fit into an amphipathic helical model.

DISCUSSION

We have isolated and partially characterized two cDNAs for one of the small hydrophobic proteolipid-like proteins present in extracts of human surfactant, herein termed surfac-

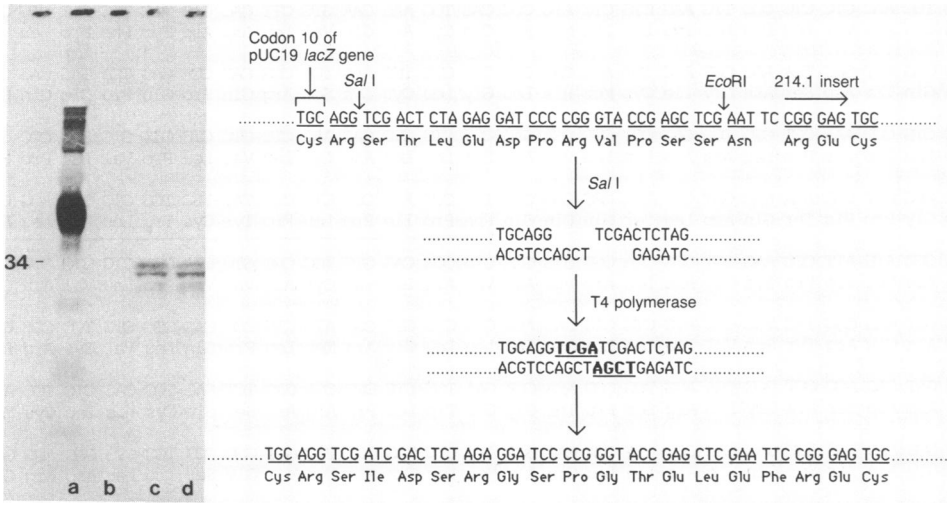


FIG. 3. (*Left*) Immunoblot analysis of the fusion protein of SPL(Phe) clone 214.1, using rabbit antiserum to bovine surfactant extract. Lane a: standard proteins used for molecular weight calibration. Lanes c and d: fusion protein (M_r 34,000) expressed from pUC19 construct. Lane b: the same construct but with the 214.1 sequence inserted out of reading frame. (*Right*) Construction of the pUC19 expression plasmid.

tant-associated proteolipid SPL(Phe) on the basis of the N-terminal amino acid residue phenylalanine. The size and deduced amino acid sequence of the human cDNA predicts a small, hydrophobic peptide derived from a larger primary translation product. SPL(Phe) is one of the hydrophobic, low molecular weight peptides recently detected in surfactant extract preparations and associated with surfactant-like activity of rapid surface absorption and surface-tension lowering when mixed with synthetic lipids (11, 12). SPL(Phe) therefore has potential utility in the synthesis of protein/phospholipid mixtures for therapy of surfactant-deficient states.

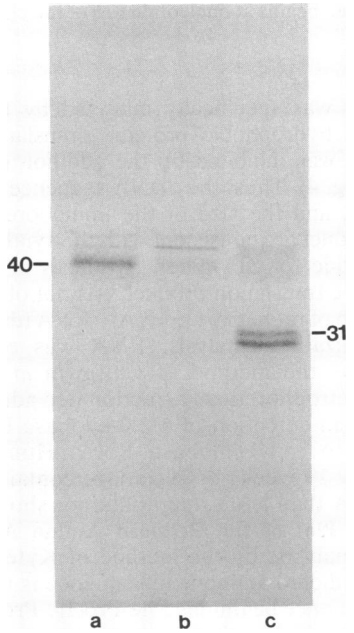


FIG. 4. Hybrid-arrested and -selected translation. Poly(A)$^+$ RNA from adult human lung was hybridized with ≈5 μg of SPL(Phe) cDNA 214.1 for 2 hr at 50°C in 80% formamide/10 mM Pipes, pH 6.4/0.25 mM EDTA/0.4 M NaCl. RNA·DNA hybrids were aliquoted to separate tubes; in one, the hybrid was "melted" by heating at 100°C for 1 min followed by rapid chilling. Immunoprecipitation of the hybrid-arrested and hybrid-selected products of *in vitro* translation with rabbit reticulocyte lysate identified a M_r 40,000 protein in the sample that had been hybridized and melted (lane a) and the virtual absence of that protein in the hybrid-arrested form (lane b). Immunoprecipitation with antiserum directed against another surfactant-associated protein, SAP-35 (lane c), demonstrated the specificity of the SPL(Phe) cDNA for the M_r 40,000 protein.

The primary sequence beginning Phe-Pro-Ile-Pro-Leu-Pro- is generated by proteolytic cleavage of a larger propeptide, with cleavage occurring at a Gln-Phe peptide bond. The exact location of the C-terminus of SPL(Phe) peptides in surfactant is unclear at present, since the cDNA sequence (from the N-terminal phenylalanine codon to the stop codon) predicts a polypeptide of 181 amino acids, larger than the heterogeneous proteins of M_r 6000–14,000 seen by NaDodSO$_4$/PAGE. The predicted human SPL(Phe) protein is 77% homologous with the first 60 amino acids obtained from bovine SPL(Phe) [S. Yu, F. Possmayer, and R. Olafson (London, ON), personal communication]. In our own laboratory, most of the proteins in the surfactant extract preparations migrate with M_r 6000; smaller amounts of protein are detected at M_r 14,000, 18,000, and 26,000. Increased amounts of the higher molecular weight proteins were detected in the absence of 2-mercaptoethanol, suggesting the presence of disulfide-linked oligomers or aggregates of the peptides. Since two distinct N-terminal sequences have been identified, it is unclear whether the size heterogeneity represents distinct peptides of M_r 6000 and 14,000 or various peptides [containing SPL(Phe)] that are generated by variable proteolytic cleavage. Alternative cleavage of the precursor propeptide might result in the heterogeneity observed by electrophoretic analysis of surfactant proteins.

Only one significant sequence difference was noted in the two clones isolated for SPL(Phe). Clone 219.1 is 398 bases long and encodes the SPL(Phe) protein sequence with 321 bases of 5′ nucleotide sequence. In clone 214.1, the 5′ sequence exists as an open reading frame and differs from the upstream 5′ nucleotide sequence from clone 219.1 at a single nucleotide (position 107 in 214.1) resulting in a threonine codon in place of the isoleucine codon found in 219.1. This single nucleotide change is striking in that it results in a loss of an Asn-Gln-Thr amino acid sequence that comprises a consensus sequence for asparagine-linked glycosylation.

Use of the SPL(Phe) clone 214.1 DNA as a probe in blot hybridization analysis of poly(A)$^+$ RNA from adult lung identified a 2.0-kilobase RNA. Much less SPL(Phe) RNA was detected in fetal lung RNA, suggesting an ontogenic expression of this surfactant-associated protein. Morphologic differentiation of the type II cells and increased surfactant phospholipid

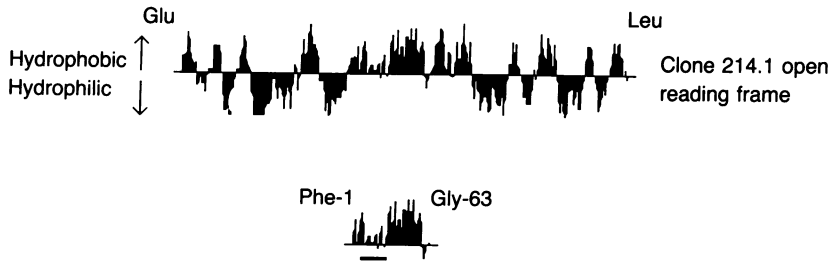


FIG. 5. Hydropathy plot of amino acid sequence predicted by clone 214.1. The plot was based on $n = 11$, using the assumptions of Kyte and Doolittle (27). The N-terminal region of SPL(Phe) (Phe-1 to Gly-63) is represented below the complete plot; the horizontal bar underlines the putative amphipathic helix (Trp-9 to Val-29).

synthesis accompany the perinatal maturation of the lung prior to birth (30). Factors controlling the developmental expression of SPL(Phe) synthesis remain to be more fully determined.

The predicted amino acid sequence of SPL(Phe) completely distinguishes this hydrophobic protein from SAP-35 (3, 4), a relatively abundant glycoprotein present in the mammalian surfactants. The predicted SPL(Phe) protein is remarkably hydrophobic in nature, consistent with its strong association with phospholipids and its solubility in chloroform/methanol or ether/ethanol solvent systems. SPL(Phe) has features similar to those previously described for proteolipids (29). Extensive dialysis of the surfactant preparations with chloroform/methanol results in a preparation with little remaining phospholipid (13). The possibility that the hydrophobic proteins contain covalently attached phospholipid remains unclarified, however. Hydropathy plots and protein conformations based on the predicted amino acid sequence demonstrate a highly hydrophobic transmembrane domain, with the more hydrophilic regions in the N-terminal and the C-terminal domains. Based on the end of the hydrophobic domain and the size of the protein as estimated by $NaDodSO_4$/PAGE, we predict a C-terminus resulting in a peptide of ≈60 amino acids (Phe-1 to approximately Gly-63). The C- and N-terminal domains are both consistent with a model of an amphipathic helix (28). An N-terminal helix of 21 amino acids can be modeled beginning with Trp-9 and terminating with Val-29. Following the possible amphipathic helical domain, a hydrophobic region rich in valine and leucine comprises a domain likely conferring close association with surfactant phospholipids. The apposition of an amphipathic helical domain and a hydrophobic domain is also apparent in another surfactant-associated protein, SAP-35 (3, 23). Recent work in this laboratory showed that this region of SAP-35 is the phospholipid binding site. We have been unable to determine the C-terminus of the SPL(Phe) protein by direct sequence analysis, and numerous attempts to isolate proteolytic or cyanogen bromide-generated fragments of SPL(Phe) have been unsuccessful.

Surfactant extracts used in therapy of hyaline membrane disease in premature infants contain only small hydrophobic proteins of M_r 6000–14,000. Reconstitution of synthetic phospholipids with purified hydrophobic proteins results in a mixture whose biophysical properties are virtually identical to those of whole pulmonary surfactant, including virtually instantaneous surface absorption and reduced surface tension during dynamic compression (7, 9, 11, 13). SPL(Phe) and/or related peptides likely interact with surfactant phospholipids to impart this remarkable surfactant-like activity to the phospholipids. Thus, the SPL(Phe) and/or related peptides may be potentially useful for the treatment of surfactant-deficient states.

We wish to acknowledge the assistance of Jean Clark, Joseph Meuth, and William Hull. This work was supported in part by Research Career Development Award HL10124 (J.A.W.) and Grants HL28623 (J.A.W.), HL36055 (T.W.) and GM10799 (T.R.K.) from the National Institutes of Health and by Ross Laboratories, Columbus, Ohio. S.W.G. is a graduate student in the Developmental Biology Program, University of Cincinnati.

1. King, R. J., Klass, D. J., Gikas, E. & Clements, J. A. (1973) *Am. J. Physiol.* **224,** 788–795.
2. Ross, G. F., Meuth, J., Ohning, B., Yangkil, K. & Whitsett, J. A. (1986) *Biochim. Biophys. Acta* **870,** 267–278.
3. White, R. T., Damm, D., Miller, J., Spratt, K., Schilling, J., Hawgood, S., Benson, B. & Cordell, B. (1985) *Nature (London)* **317,** 361–363.
4. Whitsett, J. A., Hull, W., Ross, G. & Weaver, T. (1985) *Pediatr. Res.* **19,** 501–508.
5. Phizackerly, P. J. R., Town, M. H. & Neuman, G. E. (1979) *Biochem. J.* **183,** 731–736.
6. Tanaka, Y., Takei, T. & 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. & Fisher, A. B. (1984) *J. Clin. Invest.* **74,** 677–684.
9. Yu, S. & Possmayer, F. (1986) *Biochem. J.* **236,** 85–89.
10. Takahashi, A. & Fujiwara, T. (1986) *Biochem. Biophys. Res. Commun.* **135,** 527–532.
11. Whitsett, J. A., Notter, R. H., Ohning, B. L., Ross, G., Meuth, J., Holm, B. A., Shapiro, D. L. & Weaver, T. E. (1986) *Pediatr. Res.* **20,** 460–467.
12. Whitsett, J. A., Hull, W. M., Ohning, B. L., Ross, G. & Weaver, T. E. (1986) *Pediatr. Res.* **20,** 744–749.
13. Notter, R. H., Shapiro, D. L., Ohning, B. & Whitsett, J. A. (1987) *Chem. Phys. Lipids*, in press.
14. Fujiwara, T., Maeta, H., Chida, S., Morita, T., Watabe, Y. & Ake, T. (1980) *Lancet* **i,** 55–59.
15. Enhorning, G., Shennan, A., Possmayer, F., Dunn, M., Chen, C. P. & Milligan, J. (1985) *Pediatrics* **76,** 145–153.
16. Kwong, M. S., Egan, E. A., Notter, R. H. & Shapiro, D. L. (1985) *Pediatrics* **76,** 585–592.
17. Young, R. A. & Davis, R. W. (1983) *Science* **222,** 778–782.
18. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9,** 309–321.
19. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1194–1198.
20. Messing, J. (1983) *Methods Enzymol.* **101,** 70–78.
21. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33,** 103–119.
22. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
23. Ross, G. F., Notter, R. H., Meuth, J. & Whitsett, J. A. (1986) *J. Biol. Chem.* **261,** 14283–14291.
24. Matteucci, M. D. & Caruthers, M. H. (1981) *J. Am. Chem. Soc.* **103,** 3185–3191.
25. Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y. & Matsubara, K. (1985) *J. Biol. Chem.* **260,** 2605–2608.
26. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
27. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
28. Segrest, J. P., Jackson, R. L., Morrisett, J. D. & Gotto, A. M. (1974) *FEBS Lett.* **38,** 247–253.
29. Folch, J., Lees, M. & Sloane-Stanley, G. H. (1957) *J. Biol. Chem.* **226,** 497–509.
30. Rooney, S.A. (1985) *Am. Rev. Respir. Dis.* **131,** 439–460.