Two SP-C Genes Encoding Human Pulmonary Surfactant Proteolipid*

(Received for publication, October 9, 1987)

Stephan W. Glasser†, Thomas R. Korfhagen, Charles M. Perme, Tami J. Pilot-Matias‡, Sharon E. Kister§, and Jeffrey A. Whitsett¶

From the Divisions of Pulmonary Biology, Children’s Hospital, Cincinnati, Ohio 45267-0541 and §Corporate Molecular Biology, Abbott Laboratories, Abbott Park, Illinois 60064

Human pulmonary surfactant proteolipid of $M_r = 5,000$, now termed surfactant protein C (SP-C), is produced by proteolytic processing of an $M_r = 22,000$ precursor. The active hydrophobic peptide imparts surface active properties to pulmonary surfactant phospholipids. We have determined the entire nucleotide sequence of two distinct genes encoding SP-C from a genomic library prepared from human leukocytes. SP-C genes were encoded by approximately 3.0 kilobase pairs of DNA containing six exons and five introns. In both genes, the active hydrophobic region of the polypeptide was located in the second exon that encodes a peptide of 53 amino acids. The entire nucleotide sequences of the two classes of SP-C genes differed by only 1%. Two cDNAs encoding SP-C were distinguished on the basis of an 18-nucleotide deletion at the beginning of the fifth exon; no such deletion was detected within the two classes of SP-C genes. Comparison of the 3' untranslated regions of SP-C cDNA clones and the two classes of genomic clones demonstrated that cDNAs with and without the 18-base pair deletion could be derived from both of the genes. This 18-base pair deletion occurs in nucleotide sequences compatible with two distinct RNA splice sites. One additional cDNA clone showed the addition of an 8-base pair insert at the end of exon 5, which was also compatible with two distinct splice sites. Both classes of SP-C genes were represented by cDNAs, demonstrating that both classes of genes are actively transcribed. The two SP-C genes were readily distinguished on the basis of their nucleotide sequences and restriction fragment analyses of their flanking DNA. Two distinct classes of human SP-C genes are transcribed, and the heterogeneity in the SP-C RNAs appears to result from differential splicing.

Pulmonary surfactant is a complex mixture of phospholipids and proteins in the alveoli of the lung and confers the surface tension lowering properties essential for normal breathing. Hydrophobic proteins have been isolated from pulmonary surfactant, migrating in SDS-polyacrylamide gel electrophoresis with $M_r = 5,000-18,000$, depending upon the species and presence or absence of sulphydryl reducing agents (1-8). Recently, two distinct hydrophobic polypeptides have been more fully characterized. These peptides are soluble in relatively nonpolar solvents and enhance the surface adsorption properties of surfactant phospholipids. Both proteins are components of bovine-based surfactant replacement preparations being utilized for treatment of hyaline membrane disease (9-11). cDNA encoding human surfactant proteolipid with amino terminus of phenylalanine, now to be termed SP-B, and cDNA encoding a related canine protein, termed SP-18, were recently described (12-14). SP-B is derived by proteolytic processing of a 40,000-dalton precursor, resulting in an $M_r = 6,000-8,000$ (reduced) or $M_r = 18,000$ (unreduced) hydrophobic peptide (12-14). Likewise, several distinct cDNAs encoding a second hydrophobic surfactant protein, SP-C (termed surfactant proteolipid SLP[pVal]), were recently isolated (15, 17). SP-C arises by proteolytic processing of an $M_r = 22,000$ precursor. Expression of both SP-C and SP-B increased with advancing gestation in human fetal lung, and expression of SP-C and SP-B mRNA was enhanced by exposure to glucocorticoid and to a lesser extent by 8-Br-CAMP in fetal lung explants (15, 17). DNA sequences encoding SP-C were assigned to human chromosome 8 (15). In the present work, we report the isolation, sequence, and characterization of two distinct genes encoding human SP-C and their relationship to two classes of human SP-C cDNAs.

MATERIALS AND METHODS

Reagents and Bacterial Strains—T4 DNA ligase and DNA restriction endonucleases were obtained from New England Biolabs and used according to manufacturer’s recommendations. Reverse transcriptase was obtained from Life Sciences, Inc., St. Petersburg, FL. Plasmids pUC18 and pUC19 and bacteriophage M13mp18 and mp19 were purchased from Pharmacia LKB Biotechnology Inc. and were used for subcloning and DNA sequencing according to modifications of the dideoxy methods as described by Bruner et al. (18). The EMBL3 human leukocyte genomic library was purchased from Clontech, Inc., Palo Alto, CA. Nick translation kits were purchased from Bethesda Research Laboratories Life Technologies, Inc. Escherichia coli JM109 was used as host strain for pUC plasmids and M13 clones. E. coli DP50supF was used as host strain for the EMBL3 library and propagation of a phage.

Screening of the Genomic Library—The cDNA insert of SP-C clone 334.2 previously described (15) was used to screen the EMBL3 genomic library. It was labeled with [α-32P]dCTP using the nick translation kit according to manufacturer’s recommendations. Approximately 2.5 x 10⁶ phage were plated at a density of 30,000 plaque-forming units per 100 x 10 mm plate. Duplicate nitrocellulose filters were lifted from each plate. Filters were prehybridized overnight at 61 °C in 6 x SSC (1 x SSC is 150 mM NaCl, 15 mM sodium citrate), 1 x blocking buffer, and 0.25 mg/ml denatured bacterial chromosome. Filters were then hybridized at 61 °C in 6 x SSC, 0.25% SDS, 100 µg/ml denatured bacterial chromosome, 1 x blocking buffer, and 10% formamide for 48-72 h. Filters were washed 3 times for 5 min each in 2 x SSC, 0.1% SDS at 61 °C. Blots were exposed to an X-ray film using an intensifying screen.
10 × Denhardt's solution (1 × Denhardt's is 0.02% bovine serum albumin, 0.02% polyvinylpyrrolidone, 0.02% Ficoll), 0.1% SDS, and 100 μg/ml heat-denatured salmon sperm DNA. Filters were hybridized with the 32P-labeled cDNA probe at about 1 ng/ml (2 × 10⁶ cpm/μg) at 61 °C for 42 h in fresh buffer. Filters were washed as follows: briefly in 3 × SSC, 0.1% SDS at room temperature; 45 min in 6 × SSC, 0.1% SDS at 60 °C; 45 min in 1 × SSC, 0.1% SDS; and 120 min in 1 × SSC, 0.1% SDS at 60 °C. Filters were exposed to Kodak XAR film with intensifying screens for 96 h. Second and third screens to obtain plate purity were performed at 10- and 100-fold lower phase dilutions, respectively. Phage lysates and plasmid DNA were prepared from plate lysates or from liquid lysates essentially as described by Maniatis et al. (18). To identify genomic fragments containing SP-C sequences, cloned genomic DNA was digested with restriction endonucleases and characterized by Southern hybridization using the 32P-labeled cDNA insert of clone 334.2 as a probe. The probe was labeled by the oligonucleotide primer labeling technique using a kit purchased from Pharmacia LKB Biotechnology Inc. DNA was transferred to nitrocellulose and hybridized with the probe (1.5 ng/ml) in 5 × SSC, 5 × Denhardt's solution, 0.1% SDS, 50 μg/ml heat-denatured salmon sperm DNA at 65 °C. The filters were rinsed at room temperature in 2 × SSC, 0.1% SDS and then twice in 2 × SSC, 0.1% SDS at 65 °C and four times in 0.2 × SSC, 0.1% SDS at 65 °C and subjected to autoradiography.

Nucleotide Sequence Analysis—Oligonucleotides were synthesized by a phosphoramidite procedure utilizing an Applied Biosystems nucleic acid synthesizer (20). Oligonucleotide probes for identification of the first exon were labeled with [32P]dATP using polynucleotide kinase essentially as described by Maniatis et al. (21). Sequence analysis was performed by the dyeex method of Brunner et al. (18) as modified for use with primer oligonucleotides. Combined electrophoresis on 80 cm (6%) gels and 100 cm (5%) gels for 24 and 48-72 h, respectively, was routinely utilized to extend the reading to 700 bp per reaction. Primer extension sequence analysis was performed utilizing a complementary oligonucleotide which was synthesized on the basis of the most 5'-cDNA sequence available. The oligonucleotide was used as a primer to human adult lung poly(A)+ RNA. A modification of the RNA-directed sequencing method described by Geliebter et al. (22) was used.

The following oligonucleotides were used for blot hybridization or sequencing reactions:

5'-GGGCAGCGGGGAATGCGCAATCGGC-3' (exon 2 (-)-strand) (1)
5'-AGCAAGTAGGATGCGGCCG-3' (exon 1 (+)-strand) (2)
5'-CTGTGGTCATAC-3' (exon 4 (+)-strand) (3)
5'-CTTCTGGGCGGTCG-3' (exon 5 (+)-strand) (4)
5'-CACGCCCATGCGCA-3' (exon 5 (-)-strand) (5)
5'-AGGACTTGCCGAGAGCGA-3' (intron 2 (+)-strand) (6)
5'-TAGACGTATGGCCT-3' (exon 5 (-)-strand) (7)
5'-GTTCGAGGATGACC-3' (exon 3 (+)-strand) (8)

RESULTS

Twelve genomic clones were identified by their hybridization with SP-C cDNA clone 334.2. Restriction endonuclease mapping of the 12 clones demonstrated two distinct patterns of DNA fragments. Seven clones were represented by clone AVG519 and five clones were represented by clone AVG524. Restriction endonuclease fragments containing SP-C coding regions were identified by their hybridization with cDNA clone 334.2. A distinct 1.8-kb HindIII/EcoRI fragment was identified in clone AVG519, whereas a 4-kb HindIII fragment was identified in clone AVG524. Both bands were reduced by 250 bp after digestion with ApaLI which cuts the SP-C cDNAs in the region encoding the active hydrophobic peptide isolated from surfactant (15). These preliminary analyses delineated the restriction endonuclease fragments encoding SP-C and were consistent with the restriction analysis of the cDNA clones previously characterized for SP-C (15).

The restriction map and organization of the two SP-C clones are represented by Fig. 1. The flanking regions of the genomic clones contained restriction site differences identified with Smal, HindIII, BamHI, and ApaLI. Restriction maps of the two classes of SP-C genes were readily distinguishable by this analysis. Nucleotide sequence analysis revealed that both SP-C genes were composed of six exons and five introns spanning 2687 base pairs from TATAA to the end of the sixth exon, Fig. 2. Part of the first, fifth, and the entire sixth exons of SP-C are untranslated. The Mᵦ, 22,000 polypeptide precursor is encoded by exons 2, 3, 4, and 5. The most hydrophobic region of the peptide (beginning NHZ-Ile-Pro-Cys-Cys-Pro-Val . . .) is located within exon 2 which encodes a peptide of 53 amino acids.

Because we were unable to identify complete cDNAs encoding SP-C, RNA-directed sequencing was performed to identify the nucleotide sequence upstream from the available cDNA sequence to the transcription initiation site. A strong reverse transcriptase stop was detected in the RNA-directed sequence analysis, Fig. 3. It is inferred that this stop demonstrates the site of transcription initiation. To identify the 5' untranslated exon and upstream flanking sequences of the SP-C genes, an oligonucleotide spanning the 5' most cDNA and eight base pairs of extended RNA-directed sequence was synthesized (oligonucleotide 2) and used to analyze Southern blots of AVG519 and AVG524, Fig. 4. Identical restriction fragment patterns were detected in both classes of genomic clones with this oligonucleotide after digestion with KpnI and KpnI/HindIII. The 1.8-kb KpnI and 1.1-kb KpnI/HindIII fragments were sequenced and found to include nucleotide sequences from the 5' most cDNA clone (clone RJ2-1) and the 5'-nucleotide sequence derived from mRNA-directed sequence analysis, Fig. 3. As shown in Fig. 2, a sequence TATAA was located 34 base pairs 5' to the predicted transcription initiation site in both clones. The sequence TACACCTCT matches a consensus eucaryotic promoter sequence TCAATCT in six of eight nucleotides (23). The sequence GCAGGAAA (nucleotide -549) and TGTGTTGGC (nucleotide +75) match with...
Two Genes Encoding Human Pulmonary SP-C

Fig. 2. Nucleotide sequence of XVG519 and predicted translation product. Nucleotide sequence analysis was performed as described by Brunner et al. (16). The nucleotide numbered +1 indicates the transcription initiation site determined as described in the text. Nucleotides prime to this site are labeled as + and 5' are labeled as -.

The translated peptide sequence was predicted from the cDNA sequence for SP-C (15).

Differences in nucleotide sequence between XVG519 and XVG524 are written above the XVG519 clone sequence; asterisks indicate single base insertions; dashes indicate single base deletions relative to XVG524; and letters indicate substitutions.

The identified enhancers of human c-fos and the Herpes virus, respectively (24).

As shown in Fig. 2, the two SP-C genes demonstrate few but clear differences in nucleotide sequence. However, the predicted protein sequence from the exons encoded by the XVG519 and XVG524 were identical. No differences were observed in the first five exons of the SP-C genes, and two nucleotide differences were detected in the sixth exon which is in the 3' untranslated region. Sequences of the 3' untranslated regions of SP-C cDNA clones identified their relationship to the two classes of SP-C genes. Five cDNA clones representing the XVG519 gene and two representing XVG524 gene were detected, demonstrating the transcription of both classes of SP-C genes.

Two distinct types of SP-C cDNA clones were previously isolated (15), differing by the presence (+18) or absence (−18) of 18 base pairs. This 18-base pair deletion results in the removal of six amino acids with no change in reading frame. These 18 base pairs are found in both classes of SP-C genomic clones, at the beginning of exon 5 (nucleotides 2060–2078), and bracket the untranslated portions of exon 5 and all of exon 6.

Two distinct genes encoding the human surfactant proteolipid, SP-C precursor, have been isolated and characterized.
were analyzed on a 6% polyacrylamide/urea gel and subjected to RNA sequencing was performed as described under “Materials and Methods” according to the method of Geliebter (22). Reactions were analyzed on a 6% polyacrylamide/urea gel and subjected to autoradiography. Sequence from the RNA template is as follows: AGGTGCTATGCTCTCCT, CGGGCTCTCCATCAGGACCTCTTTGCTGCCCACATCCATCTTGCTGC-

The nucleotide sequences of both genes demonstrate close homology, with the differences confined to exon 6, introns, and flanking regions of the genes. Comparison of the SP-C cDNA and genomic clones demonstrated the organization of six exons and five introns which comprise the SP-C genes.

There was complete identity of the exons encoding the polypeptide in both classes of SP-C genomic clones. Therefore, the SP-C genomic clones did not account for the differences in nucleotide sequence of two classes of SP-C cDNAs previously described. SP-C mRNAs representing both classes of genes were detected by analysis of the nucleotide sequence of the 3′-untranslated region of SP-C cDNAs.

Sequences associated with the consensus eucaryotic promoter, TATAAA (23), are located at 34 base pairs upstream of the transcription initiation site in both clones as predicted from the RNA sequencing studies. The sequence TCACCTCTCT matches the upstream eucaryotic promoter TCAATCT (23) in 6 of 8 base pairs and is located 62 base pairs upstream of the transcription initiation site. Intron-exon junctions are well demarcated by consensus splice sequences AG and GT (23, 25). The 5′-untranslated region is only 25 bp, while the 3′ portion of exon 5 and the entire sixth exon constitute the 3′-untranslated region. Short 5′-untranslated regions have been described in other genes such as those encoding histones (26). The significance of such a short 5′-untranslated sequences is not known.

Hydropathy analysis of the predicted SP-C precursor peptide demonstrates that the hydrophobic, potentially membrane-associated domain of the Mₛ = 22,000 precursor protein is contained within the second exon of both SP-C genes. The predicted peptide domain derived from exons 3 and 4 are rich in charged amino acids which are not compatible with the amino acid composition of the SP-C precursors. The precise carboxyl terminus of SP-C peptide has not been reported; however, its migration with Mₛ = 5,000-6,000 and its hydrophobic properties including association with lipids and its solubility in organic solvents are consistent with its derivation primarily from the peptide encoded by the second exon. There is no readily discernible signal peptide sequence in the amino terminus of the predicted Mₛ = 22,000 polypeptide, and the nature of the processing and secretion of SP-C are unclear at present. Proteolytic processing in both amino and carboxyl termini must account for the generation of the smaller peptide detected in pulmonary surfactant. The lack of a signal peptide sequence differs from that for other surfactant proteins, SP-A and SP-B, which have amino-terminal sequences compatible with that of signal peptides (13, 27).

Pulmonary surfactant synthesis and secretion is highly regulated during perinatal development and accompanies the morphologic differentiation of respiratory epithelial cells to Type II cells. The Type II cell is the major synthetic site for the production of surfactant. Surfactant phospholipid synthesis is enhanced by numerous humoral agents including glucocorticoids and cyclic nucleotides (28–30). Recent work from our laboratory has demonstrated the induction of both SP-C and SP-B during development of human fetal lung. Addition of dexamethasone enhanced the RNA level of both SP-C and SP-B in fetal lung organ culture and concomitantly decreased that of SP-A (17). Both stimulatory and inhibitory effects of glucocorticoids in SP-A (SAP-35) expression have been recently reported (31, 32). Stimulation of SP-C RNA synthesis by 8-Br-cAMP was also detected in this culture system; however, 8-Br-cAMP was not as effective as dexamethasone in the in vitro system. The molecular mechanisms involved in the regulation of SP-C expression remain to be elucidated. However, the sequences TGATGTCA and TGATGGTCA were located 512 and 398 base pairs upstream from the transcription initiation site in both SP-C genomic clones. These sequences are homologous with the sequence, TGACGTCA,

519 524

-1.8 -1.1

519 524

-69 -50 -45

Fig. 3. RNA-directed primer extension sequence of exon 1. RNA sequencing was performed as described under “Materials and Methods” according to the method of Geliebter et al. (22). Reactions were analyzed on a 6% polyacrylamide/urea gel and subjected to autoradiography. Sequence from the RNA template is as follows: CGGGCTCTCCATCAGGACCTCTTTGCTGCCCACATCCATCTTGCTGC-

Fig. 4. Localization of exon 1 by Southern blot hybridization with oligonucleotide probe 2 described in the text. Lanes a, b, c, and f are digests of XVG519. Lanes c, d, g, and h are digests of XVG524. Lanes a and c are KpnI digests; lanes b and d are KpnI/HindIII double digests indicating homology of the immediate 5′ regions of the two genes. Lanes e and g are Smal digests; lanes f and h are HindIII digests indicating differences in the 5′-flanking sequences.

5S. W. Glasser, T. R. Korfhagen, C. M. Perme, T. J. Pilot-Matias, S. E. Kister, and J. A. Whitsett, unpublished observations.
which has been previously associated with cAMP responsiveness (33). The SP-C sequences comprising a possible cAMP regulatory element most closely fit those described for the rat somatostatin gene (33). No clear glucocorticoid regulatory element was discerned within 500 base pairs upstream of the transcription initiation site in the SP-C sequence analysis. A repetitive sequence with G-T nucleotides was noted in the 3′ region of both SP-C genes and is compatible with a Z-DNA element (34). The possible regulatory role of Z-DNA has been discussed but remains poorly clarified at present.

The 18-base pair deletion detected in cDNAs encoding SP-C previously reported (15) does not arise from nucleotide differences in the two genomic clones presently isolated for SP-C. Both AVG524 and AVG519 contain the 18 base pairs of sequence deleted in one class of cDNAs. The deleted sequence is located at the start of exon 5 and is preceded by a 6-base pair sequence which is also found at the end of exon 4. Sequences at the 5′ proximal end of exon 5 show two sets of sequence that conform to an intron-exon splice site consisting of a 10–15-base pair polypyrimidine tract followed by the dinucleotide AG, signaling the end of intronic sequence (25). The base pairs (5′-CTCACTTCTACATTC-3′) comprise a 17-base pair tract preceding the AG of the major cDNA species. A second sequence (5′-TGCTCTGC-3′) precedes the AG at the end of the deleted 18-base pair sequence and may represent an alternate splice site which may account for the two observed SP-C cDNA clones. Because of its downstream location the second site is less favorable for RNA splicing. This would result in the minor species lacking the 18 base pairs. This deletion alters the structure of the SP-C precursor but does not alter the structure of the active hydrophobic peptides encoded by exon 2. The physiologic relevance of this heterogeneity is unknown at present.

An additional site where alternative splicing may occur was located at the end of exon 5. A cDNA clone corresponding to the AVG524 3′-polymorphisms, TP13-1, was found to be +18 and also to contain an 8-base pair insertion at the end of exon 5 (nucleotides 2230–2237, Fig. 2). The sequences surrounding both the 5′ and 3′ ends of the 8-base pair sequence conform to the exon-intron splice sequence (25); however, only one of the 11 SP-C cDNAs analyzed contained this insert. Perhaps because of its downstream location the second site is less favored, resulting in a minor RNA species containing the 8-base pair insertion. Another cDNA clone corresponding to the AVG524 3′-polymorphisms, TP14-1, was found to be +18. In addition, cDNA clones which were both +18 and −18 were found, which showed sequence identity corresponding to the polymorphisms found in the 3′ end of SP-C genomic clone AVG519. This indicates that the spliced mRNAs can be transcribed from either copy of the SP-C gene and that both SP-C genes are expressed. Whether both of the SP-C genes are coordinately regulated remains to be clarified.

The nucleotide sequences of AVG519 and AVG524 were entirely conserved in the amino acid coding region. The genomic sequences were identical to the SP-C cDNA except for a single nucleotide difference in exon 5 in which leucine is encoded by CTG rather than TTG observed in the cDNAs. Nucleotide differences were noted in the exons encoding only the 3′-untranslated regions. The number of nucleotide sequence differences were similar in both introns and the 5′ and 3′ flanking sequences available for analysis (approximately 500 base pairs upstream from the TATAA). The 5′ most fragment of AVG519 (600 bp HindIII fragment) was not present in AVG524 by Southern blot hybridization, demonstrating differences in flanking sequences in that region.

The nearly complete conservation of the nucleotide sequence in the exons, the small divergence in the introns, and the differences in restriction mapping of the 3′ and 5′ regions are most consistent with the interpretation that these SP-C genomic clones represent two distinct gene loci encoding SP-C. The 334.2 probe used for blot analysis of total genomic DNA from two different human fetal lungs digested with EcoRI, BamHI and HindIII identified bands consistent with both of two SP-C genes. Another band was detected which cannot be accounted for on VG519 or VG524, suggesting another SP-C related sequence (data not shown). We cannot exclude the possibility that these differences are due to allelic variation. We have previously demonstrated that SP-C cDNA hybridizes only to human chromosome 8 (15).

SP-C and SP-B are remarkably hydrophobic surfactant peptides that are associated with surface properties of surfactant phospholipid. Addition of canine and bovine SP-C to mixtures of phospholipids results in a mixture with absorption and surface tension properties virtually identical to those of whole surfactant (36). Such surface activity is vital to perinatal adaptation for air breathing at birth. Knowledge of the factor or factors mediating lung and developmental specific expression of these important proteins will provide information vital to our understanding of lung development and to the clinical care of infants who may be at risk for hyaline membrane disease.

Acknowledgments—We gratefully acknowledge the assistance of Michael D. Bruno, Mary Beth Imfeld, Vivian Chu, and Kelly Floyd.

REFERENCES

Two Genes Encoding Human Pulmonary SP-C


