Vol. 144, No. 1, 1987 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 14, 1987 Pages 367-374

ISOLATION AND SEQUENCE OF A cDNA CLONE FOR THE RAT PULMONARY SURFACTANT-ASSOCIATED PROTEIN (PSP-A)*

Kimihiko Sano[§,1], James Fisher[2], Robert J. Mason[1], Yoshio Kuroki[1], James Schilling[3], Bradley Benson[3], and Dennis Voelker[1]

[1] *Department of Medicine, National Jewish Center for Immunology and Respiratory Medicine, Denver CO 80206*

[2] *Eleanor Roosevelt Institute for Cancer Research and Department of Medicine, University of Colorado Health Science Center, Denver, CO 80220*

[3] *Calfornia Biotechnology,Inc, Mountain View, CA 94043*

Received February 13, 1987

SUMMARY: Pulmonary surfactant is composed mainly of phospholipid and two groups of apoproteins. One of these apoproteins is a family of glycoproteins (pulmonary surfactant-associated protein A,PSP-A). We have isolated and sequenced a cDNA clone encoding for rat PSP-A and the full amino acid sequence has been deduced from the nucleotide sequence. The sequence of 56 amino acids at the N-terminus of PSP-A isolated from rats treated with silica was determined independently, and there is complete agreement with the sequence deduced from the cDNA. Isolated rat alveolar type II cells contain two species of mRNA for this protein.

The low surface tension in the alveoli of the lung is maintained by pulmonary surfactant which is a complex of phospholipids, cholesterol, and two families of apoproteins. The major family of proteins is a glycoprotein (PSP-A) which has a molecular weight ranging from 26 to 38 kDa (1-4). One family of apoproteins is a group of very hydrophobic proteins which have molecular weight ranging from 6 to 14 KDa (1,5,6). The protein components of surface active material are thought to be required for the full biological activity of pulmonary surfactant (1). PSP-A is synthesized and secreted by alveolar type II epithelial cells and has been reported

* This investigation has been supported by National Heart,Lung, and Blood Institute Grant HL-29891 and HL-27353.

§ Present address: Department of Pediatrics, Kobe University School of Medicine, Kobe, 650, Japan.

The abbreviations used are: PSP-A, pulmonary surfactant-associated protein A; SDS, sodium dodecyl sulfate.

0006-291X/87 $1.50

to have an affinity for phospholipids and enhance the adsorption of surfactant to an air-liquid interface (3,7). From in vitro translation studies the primary translation products of rat PSP-A have an apparent molecular weight of 26 kDa (3,4). This precursor is modified post-translationally by N-linked glycosylation to form the higher molecular weight forms which are recovered in bronchoalveolar lavage fluid (3,4). Recently PSP-A has been discovered to inhibit secretion of pulmonary surfactant from isolated alveolar type II cells in vitro (8). Although the primary structures of canine and human PSP-A have been determined (9-11), this is the first report of the structure of rat PSP-A. The characterization of rat PSP-A is important because most of the in vitro studies on the synthesis and secretion of pulmonary surfactant are preformed with isolated rat alveolar type II cells. In addition, the antibodies to human and canine PSP-A do not react well with rat PSP-A. Therefore, we determined the amino acid sequence of rat PSP-A. These data also permit comparison of the structual similarities between species and can provide insights regarding the function of this protein.

EXPERIMENTAL PROCEDURES

A library of rat lung cDNA constructed in lambda gt10 was screened on nitrocellulose filters with ^{32}P-labelled human PSP-A cDNA clone as a probe (10). The filters were prehybridized in 6x SSC (1x SSC=0.15M sodium chloride and 0.015M sodium citrate,pH 7), 2.5x Denhardt's solution, 0.5 mg/ml salmon sperm DNA, and 50% formamide at 42°C overnight and then hybridized in 6x SSC, 1x Denhardt's solution, and 0.5 mg/ml salmon sperm DNA, 50 % formamide, and 10% dextran sulfate with probe (1 x 10^6 cpm/ml) at 42°C overnight. Filters were washed 4 times for 5 min with 2x SSC and 0.1% SDS at room temperature and then 4 times for 15 min with 0.1x SSC and 0.1% SDS at 45°C and exposed for autoradiography with an intensifyng screen at -70°C for 16 h. Positive plaques were purified by 4 succesive screenings. Phage DNA was purified on a CsCl density gradient (12) and then analyzed by Southern blot analysis using the same hybridization conditions as described above after digestion of phage DNA with EcoRI. DNA from positive clones was subcloned into M13 phage for DNA sequence analysis. Various sized deletions in the parent M13 clone were generated using T4 polymerase (13). These clones were sequenced by the dideoxy sequencing method. Rat PSP-A cDNA was also subcloned into pUC plasmid and used as a probe for Northern blot analysis. Total cellular RNA was extracted by the guanidine isothiocyanate method (13) from adult rat lung and from rat alveolar type II cells which were isolated by elastase digestion of lung and by centrifugation on a discontinuous metrizamide density gradient (14). RNA was

separated on formaldehyde- denaturing agarose gels and electrotransferred to Nytran®. The filter was prehybridized in 1M NaCl, 10% dextran sulfate, 50% formamide, and 1% SDS at 42°C overnight and then hybridized in 5x SSC, 1x Dernhardt's solution, 0.1 mg/ml salmon sperm DNA, 10% dextran sulfate, 45% formamide, and 0.5% SDS with 5 x 10^5 cpm/ml ^{32}P-labelled rat PSP-A cDNA at 42°C overnight. The filter was washed as described above except that final washing was carried out at 60°C and exposed for autoradiography with an intensifying screen at -70°C for 16h. PSP-A was isolated and purified from surfactant isolated from rats which had been previously treated with silica (15). The N-terminal amino acid sequence was determined on a gas-phase microsequencer (Applied Biosystems).

RESULTS AND DISCUSSION

When the rat lung cDNA library was screened with the ^{32}P-labelled human PSP-A probe(10), about 0.1% of the clones hybridized with the probe. Ten of the positive clones were randomly selected and analyzed by EcoRI digestion and Southern blot hybridization to the human PSP-A probe. Clones 2,3,5,6,8 were found to contain a 1.6 kb insert and clones 4,9,10 contained a 0.9 kb insert which hybridized with human PSP-A cDNA. The phage DNA of clones 1 and 7 was not digestable by EcoRI, and we did not attempt to purify the DNA of these clones further. Clones 2 and 10 had the highest molecular weight in each group. To confirm that each insert corresponded with a cellular message, Northern blots of total cellular RNA purified from whole lung and from isolated rat alveolar type II cells were probed using the insert DNA from clone 2. As shown in Fig.1 for whole lung two species of RNA were identified, one with a mass of 1.6 kb and the other with a mass of 0.9 kb. These two different size of RNAs also hybridized with the insert DNA from clone 10 (data not shown). Similar results were obtained with RNA isolated from alveolar type II cells (data not shown). It is not known whether or not these two mRNA species are the transcripts of the same or two different genes. The insert DNA from clone 10 was subcloned into M13 phage and was sequenced in both orientations. Fig.2 shows the DNA sequence of the insert and predicted amino acid sequence. This clone contained the entire coding region and some untranslated sequences. The sequence of 56 amino acids at the N-terminus of PSP-A isolated from lung lavage of silicotic rat lungs corresponds exactly with the sequence predicted from cDNA from clone 10. Rat PSP-A is preceded by a signal peptide

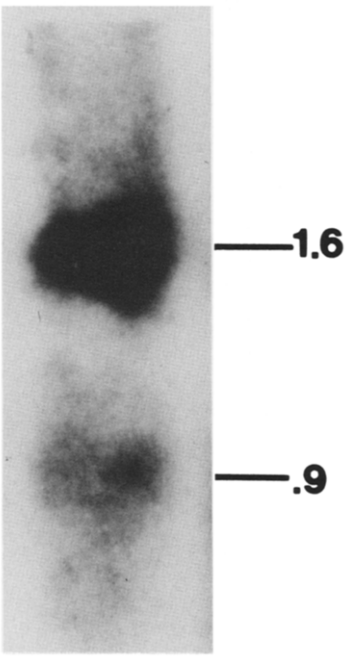

Fig.1. Northern blot of 10 ug of total cellular RNA purified from adult rat lung and probed with nick translated rat PSP-A cDNA clone 2. The hybridized bands were estimated to be 1.6 kb and 0.9 kb by comparison with RNA standards of 4.4, 2.4, 1.4, and 0.3 kb from Bethesda Research laboratories.

of 20-29 amino acids which is removed from the mature protein. From the data in fig. 2 it is apparent that there are two potential translation start sites corresponding to methionine in postions -29 and -20. The sequence 5' to the AUG coding for methionine in position -20 most closely approximates the consensus sequence of Kozak (16) for the likely start site. The sequence of the signal peptide of rat PSP-A is not homologous to canine or human PSP-A except that all residues are hydrophobic. The sequence of rat PSP-A itself is highly homologous to canine and human PSP-A as shown in Fig.3. Overall 64 % of the amino acid sequences are conserved in all species. As reported for canine and human PSP-A, rat PSP-A consists of a short N-terminal segment (7 amino acids residues), a collagen-like domain of 24 Gly-X-Y repeats, and finally a C-terminal domain of 148 amino acids.

The primary in vitro translation products of rat PSP-A have been reported to have an apparent molecular weight of 26 kDa (3,4) or 35 kDa (17). The molecular weight of rat PSP-A calculated from its predicted amino acid sequence is 28,180 daltons, and this is likely to be a slight underestimate since some of the proline

```
                    -29                                     -20
                    Met Trp Lys Pro Leu Gly Ile Val Ala Met Ser Leu Cys Ser
acacagcctgcaggtctgt ATG TGG AAG CCA CTG GGG ATA GTA GCC ATG TCA CTG TGT TCT

                    -10                                   -1   1
Leu Ala Phe Thr Leu Phe Leu Thr Val Val Ala Gly Ile Lys Cys[Asn*Val Thr Asp
TTG GCC TTC ACC CTC TTC TTG ACT GTT GTC GCT GGT ATC AAG TGC AAT GTG ACA GAC

                     10                                         20
Val Cys Ala Gly Ser Pro Gly Ile Pro Gly Ala Pro Gly Asn His Gly Leu Pro Gly
GTT TGT GCT GGA AGC CCT GGG ATC CCT GGA GCT CCT GGA AAC CAT GGT CTG CCT GGC

                        30                                      40
Arg Asp Gly Arg Asp Gly Val Lys Gly Asp Pro Gly Pro Pro Gly Pro Met Gly Pro
AGA GAC GGG AGA GAC GGT GTC AAA GGA GAC CCT GGA CCT CCA GGT CCC ATG GGC CCT

     X                        50  X                                  60
Pro Gly Gly Met Pro Gly Leu Pro Gly Arg Asp Gly Leu Pro Gly Ala Pro Gly Ala
CCT GGA GGA ATG CCA GGT CTT CCT GGA CGC GAT GGG CTG CCC GGA GCC CCT GGT GCA

                                70                                      80
Pro Gly Glu Arg Gly Asp Lys Gly Glu Pro Gly Glu Arg Gly Leu Pro Gly Phe Pro
CCT GGA GAA CGT GGA GAC AAG GGA GAG CCT GGA GAA AGG GGC CTG CCA GGA TTT CCA

                                    90
Ala Tyr Leu Asp Glu Glu Leu Gln Thr Glu Leu Tyr Glu Ile Lys His Gln Ile Leu
GCT TAC CTG GAT GAG GAG CTC CAG ACT GAA CTC TAT GAG ATC AAA CAT CAG ATT CTG

100                                     110
Gln Thr Met Gly Val Leu Ser Leu Gln Gly Ser Met Leu Ser Val Gly Asp Lys Val
CAA ACA ATG GGA GTC CTC AGC TTG CAA GGA TCC ATG CTG TCA GTG GGG GAT AAA GTC

    120                                     130
Phe Ser Thr Asn Gly Gln Ser Val Asn Phe Asp Thr Ile Lys Glu Met Cys Thr Arg
TCC ACC AAT GGG CAG TCA GTC AAC TTT GAT ACC ATT AAA GAG ATG TGT ACC TTT AGA

        140                                     150
Ala Gly Gly Asn Ile Ala Val Pro Arg Thr Pro Glu Glu Asn Glu Ala Ile Ala Ser
GCA GGA GGC AAC ATT GCT GTC CCG AGG ACT CCT GAG GAG AAC GAG GCC ATT GCA AGT

            160                                     170
Ile Ala Lys Lys Tyr Asn Asn Tyr Val Tyr Leu Gly Met Ile Glu Asp Gln Thr Pro
ATT GCG AAG AAG TAC AAC AAC TAT GTC TAC TTG GGC ATG ATT GAA GAC CAG ACT CCT

                180                                     190
Gly Asp Phe His Tyr Leu Asp Gly Ala Ser Val Asn*Tyr Thr Asn Trp Tyr Pro Gly
GGA GAC TTC CAC TAC CTG GAT GGG GCT TCT GTG AAC TAC ACC AAC TGG TAC CCA GGA

                    200                                     210
Glu Pro Arg Gly Gln Gly Lys Glu Lys Cys Val Glu Met Tyr Thr Asp Gly Thr Trp
GAA CCC AGG GGT CAG GGC AAA GAA AAG TGT GTA GAA ATG TAT ACA GAT GGG ACA TGG

                        220                                 228
Asn Asp Arg Gly Cys Leu Gln Tyr Arg Leu Ala Val Cys Glu Phe @@@
AAT GAT AGG GGC TGC CTG CAG TAC CGG CTG GCT GTT TGT GAA TTT TGA  tcaagcaatt

agacgaaaagatgaaccctcacactgcctctatcctgatgattcatctggtctgtaaaaccctgcaactaccttt

caagcaaatttacttgtggccttcg
```

Fig.2. Nucleotide sequence and predicted amino acid sequence of rat PSP-A cDNA. The peptides of the N-terminus that match predicted amino acids are indicated by underlining. Possible oligosaccharide binding sites to asparagine residues are shown by the asterisks. The superscript X over glycines in positions 44 and 51 indicate two amino acids whose identity was indeterminate in the initial protein sequence.

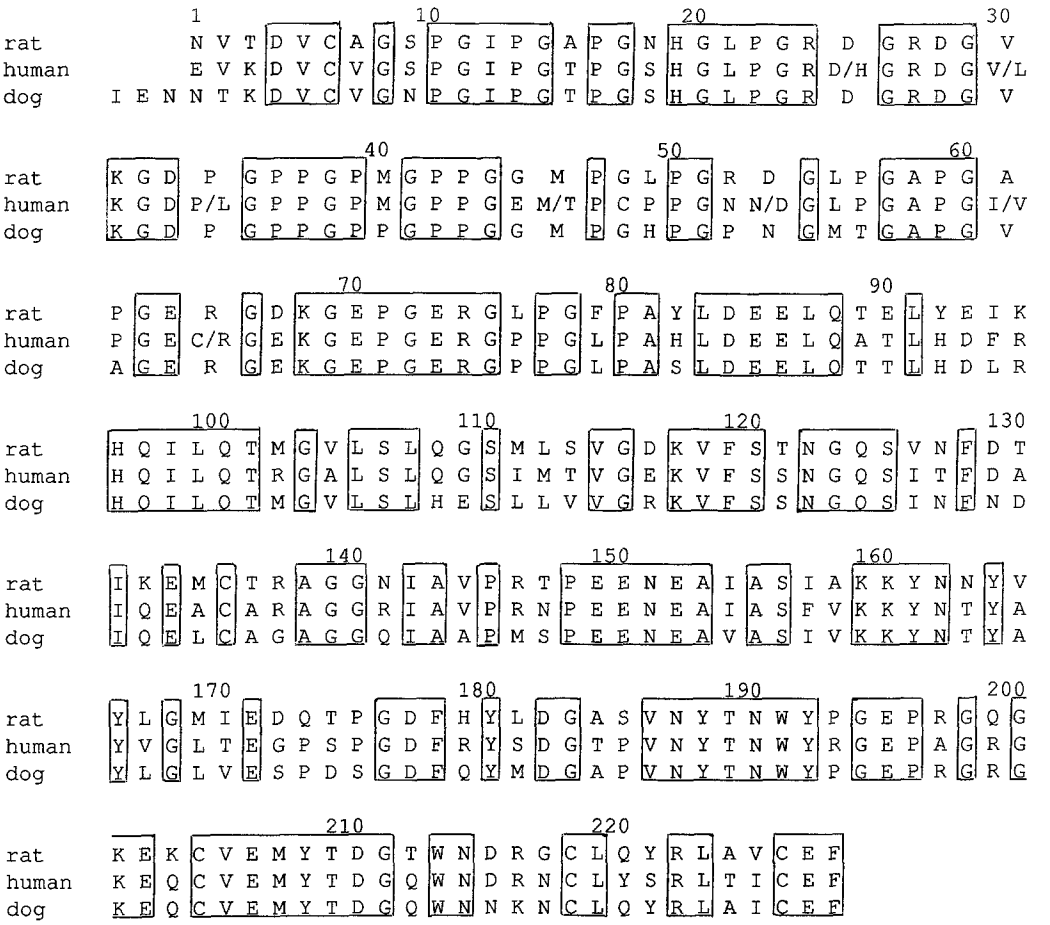

Fig.3. Comparison of rat, human, and canine PSP-A sequences. Sequences of PSP-A of the three species are compared. Sequences of human and canine PSP-A are taken from Refs.10 and 11, and 9, respectively. Residues which are identical in three species are boxed to emphasize homology.

residues in the collagen like domain are likely to be hydroxylated. By the protein sequence analysis prolines at positions 10, 13, 16, 22, 34, 37, 43, 47, 50, and 56 are hydroxylated. The calulated molecular weight of the longest signal peptide is 3,620 daltons. Theoretically the primary translation product should have a minimal molecular weight of 31,800 daltons with the signal peptide or 28,180 daltons after cleavage of signal peptide. The reason for the discrepancy of molecular weight from those reported from in vitro translation by other investigators is uncertain.

A portion of the rat PSP-A sequence forms a collagen-like region. Such domains have been found in other noncollagen proteins including several carbohydrate binding proteins such as the rat

liver mannose-binding proteins (18). Of particular interest is the presence of a single interruption in the Gly-X-Y repeats structure at Gly-44 in rat PSP-A. This interruption has also been observed for both human and canine PSP-A and for mannose-binding proteins (18). In all other portions of the collagen-like domain, the sequences resemble the triple helix forming segments of collagen in that a large number of the X and Y positions are occupied by proline residues. The cysteine residues in the noncollagen-like region at positions 135, 204, 218, and 226 in rat PSP-A are conserved in human and canine PSP-A and in two mannose-binding proteins. These residues may be involved in the formation of interchain or intrachain disulfide bonds. Benson *et al.* suggested that only Cys-9 of canine PSP-A (Cys-6 in rat PSP-A) is involved in an interchain disulfide bond formation because the collagenase-resistant domain did not form dimers in the absence of a reducing agent whereas native canine PSP-A formed dimers under these conditions (6). Since native rat PSP-A forms dimers, trimers, and tetramers under denaturing non-reducing conditions (19), at least two different cysteine residues are suggested to be involved in disulfide bonds in rat PSP-A. In human PSP-A cysteine residues at positions 135, 204, 218, and 226 are involved in intrachain disulfide bridges (Schilling and Benson, unpublished observations).

When rat PSP-A recovered from bronchoalveolar lavage fluid was analyzed by SDS polyacrylamide gel electrophoresis under reducing conditions, three different proteins having molecular weights of 26 kDa, 32 kDa, and 36 kDa were observed (19). The latter two species are thought to be glycosylated forms of the 26 kDa protein because 26 kDa is the size of primary translation product and mRNA translation in the presence of pancreatic microsomal membranes yields the two higher molecular weight forms (3,4). There are two possible N-glycosylation sites, Asn-1 and Asn-187, both of which have the consensus sequence of Asn-X-Thr/Ser. However, based on the molecular yields of recovered phenylthiohydantoin amino acids during the protein sequencing, it is unlikely that Asn-1 is glycosylated to any significant degree in rat PSP-A. Therefore, we attribute the difference in molecular weight of the 32 kDa and the 36 kDa proteins to a different degree of glycosylation at Asn-187 and/or to other post-translational modifications.

ACKNOWLEDGMENTS: Authors are grateful to Drs. Nakane, A. Kuno, and F. Murad (Stanford University) for rat lung cDNA library and to Dr. T. White (California Biotechnology Institute) for human PSP-A cDNA. Skillful technical assistance of Ms. T. Hofmann is greatly appreciated.

REFERENCES

1. King,R.J.,Klass,D.J.,Gikas,E.G., and Clements,J.A. (1975) Am. J. Physiol. 224,788-795.
2. King,R.J. (1982) J.Appl.Phys. 53, 1-8.
3. Whitsett,J.A.,Ross,G.,Weaver,T.,Rice,W.,Dion,C., and Hull,W. (1985) J.Biol.Chem. 260, 15273-15279.
4. Floros,J.,Phelps,D.S.,Kourembanas,S.,and Taeusch,H.W. (1986) J.Biol.Chem. 261, 828-831.
5. Claypool,W.D.,Wang,D.L.,Chander,A.,and Fisher,A. (1984) J.Clin.Invest. 74,677-684.
6. Whitsett,J.A.,Ohning,B.L.,Ross,G.,Meuth,J.,Weaver,T.,Holm,B.A., Shapiro,D.L.,and Notter,R.H. (1986) Pediatr.Res. 20,460-467.
7. King,R.J. and Macbeth,M.C. (1979) Biochim.Biophys.Acta 557, 86-101.
8. Dobbs,L.G.,Wright,J.R., and Hawgood,S. (1986) Am.Rev.Resp.Dis. 133, A118.
9. Benson,B.,Hawgood,S.,Schilling,J.,Clements,J.,Damm,D.,Cordell, B.,and White, R.T. (1985) Proc.Natl.Acad.Sci.USA 82, 6379-6383.
10. White,R.T.,Damm,D.,Miller,J.,Spratt,K.,Schilling,J., Hawgood,S. ,Benson,B., Cordell,B. (1985) Nature 317, 361-363.
11. Floros,J.,Steinbrink,R.,Jacobs,K.,Phelps,D.,Kriz,R.,Recny,M., Sultzman,L.,Jones,S.,Taeusch,H.W.,Frank,H.A., and Fritsch,E.F. (1986) J.Biol.Chem. 261, 9029-9033.
12. Maniatis,T.,Fritsch,E.F., and Sambrook,J. (1982) Molecular cloning: a loboratory manual (Cold Spring Harbor Laboratory, Cold Spring Harbor,NY).
13. Dale,R.M.K.,McClure,B.A., and Houchins,J.P. (1985) Plasmid 13, 31-40.
14. Dobbs,L.G., and Mason,R.J. (1979) J.Clin.Invest. 63, 378-387.
15. Dethloff,L.A.,Gilmore,L.B.,Brody,A.R., and Hook,G.E.R. (1986) Biochem.J. 233,111-118.
16. Kozak,M. (1984) Nucl.Acid Res. 12,857-872.
17. Katyal,S.L. and Singh,G. (1985) Biochem.Biophys.Res.Commun. 127, 106-111.
18. Drickamer,K.,Dordal,M.S., and Reynolds,L. (1986) J.Biol.Chem. 261, 6878-6887.
19. Weaver,T.E.,Hull,W.M.,Ross,G.F., and Whitsett,J.A. (1985) Biochim.Biophys.Acta 827, 260-267.