

# Novel, Non-Radioactive, Simple and Multiplex PCR-cRFLP Methods for Genotyping Human SP-A and SP-D Marker Alleles

Susan DiAngelo<sup>1</sup>, Zhenwu Lin<sup>1</sup>, Guirong Wang<sup>1</sup>, Scott Phillips<sup>1</sup>, Mika Ramet<sup>1</sup>, Junming Luo<sup>1</sup> and Joanna Floros<sup>1,2,#</sup>

<sup>1</sup>*Department of Cellular and Molecular Physiology, Pennsylvania State University College of Medicine, Hershey, PA, USA*

<sup>2</sup>*Department of Pediatrics, Pennsylvania State University College of Medicine, Hershey, PA, USA*

**ABSTRACT:** We have previously identified an allele of the human SP-A2 gene that occurs with greater frequency in an RDS population [12]. Because of the importance of SP-A in normal lung function and its newly emerging role in innate host defense and regulation of inflammatory processes, we wish to better characterize genotypes of both SP-A1 and SP-A2 genes. It has been determined that SP-D shares similar roles in immune response. Therefore, in this report we 1) describe a novel, non radioactive PCR based-cRFLP method for genotyping both SP-A and SP-D; 2) describe two previously unpublished biallelic polymorphisms within the SP-D gene; 3) present the partial sequence of one new SP-A1 allele (6A<sup>14</sup>) and describe other new SP-A1 and SP-A2 alleles; and 4) describe additional methodologies for SP-A genotype assessment. The ability to more accurately and efficiently genotype samples from individuals with various pulmonary diseases will facilitate population and family based association studies. Genetic polymorphisms may be identified that partially explain individual disease susceptibility and/or treatment effectiveness.

# Correspondence: Joanna Floros, Ph.D., Department of Cellular and Molecular Physiology, Pennsylvania State University College of Medicine, 500 University Drive, H166, Hershey, PA 17033, USA, Tel.: +1 717 531 6972, Fax: +1 717 531 7667, E-mail: jxf19@psu.edu

**KEYWORDS:** Surfactant SP-A, SP-D, SNPs, genotype analysis

## INTRODUCTION

Pulmonary surfactant is essential for normal lung function. It is a lipoprotein complex and consists of phospholipids and surfactant proteins (SP)-A, B, C, and D. Deficiency of surfactant or impaired surfactant activity could lead to a variety of pulmonary diseases [6,8]. Although compromised or altered surfactant states in a normal healthy individual may not be problematic, in the face of a variety of stresses these may manifest themselves into a disease. The surfactant proteins have been shown in the last 10–15 years to play important roles in surfactant physiology and biology [5,6,8]. In addition, SP-A and SP-D are involved in the innate local host defense and the regulation of inflammatory processes in the lung [23].

Alterations in the levels of the surfactant proteins have been associated with respiratory distress syndrome (RDS) in the prematurely born infant [2,21], with acute respiratory distress syndrome (ARDS) [9], and with other pulmonary diseases [6]. Genetic alterations in surfactant protein genes have been associated with disease as well. Mutations in the SP-B gene that lead to SP-B deficiency have been described in congenital alveolar proteinosis [7,20,22] and certain SP-A and SP-B alleles have been observed with higher frequency in infants with RDS [12]. Genetic alterations in the genes responsible for the lipid component of surfactant have not been described, but these are likely to affect other organs, and thus may be lethal and not observed.

Therefore, the available information indicates that the surfactant protein genes may be good candidate genes in the study of genetic contribution to the pathogenesis of pulmonary disease.

The SP-A and SP-B loci are shown to be polymorphic and methods to determine SP-A [4] and SP-B [19,20] genotypes have been published. The human SP-A locus consists of two (SP-A1 and SP-A2) functional genes in opposite transcriptional orientation and one pseudogene [10]. A number of alleles have been characterized for each human SP-A gene [5]. Because both SP-A genes and their respective alleles share a high degree of sequence similarity (> 95%), SP-A genotyping has been challenging.

We have previously described a method to distinguish the two human SP-A genes, SP-A1 and SP-A2, and their corresponding alleles [4]. In this method we took advantage of the biallelic nucleotide variants at specific positions within the coding region as well as the ordered sequence of these variants within the various SP-A sequences in order to make gene specific and allelic designations possible. Biallelic nucleotide variants that result in a change of the encoded amino acid and distinguish the SP-A1 gene and its corresponding alleles from the SP-A2 gene and its corresponding alleles occur at residues 66, 73, 81 and 85. Biallelic SP-A1 nucleotide variants that change the encoded amino acid and distinguish SP-A1 alleles from one another occur at amino acids 19, 50 and 219, and for SP-A2 at amino acids 9, 91 and 223. In addition nucleotide differences that do not change the encoded amino acid of the SP-A alleles have also been characterized, and used for the designation of the individual alleles. Therefore, the pattern of various combinations of polymorphisms (at specified amino acids) serves as a key to distinguish one allele from another.

The genotyping method we described previously [4] is based on gene specific PCR amplification and allele specific hybridization. Using this hybridization and direct sequencing we have reported 5 alleles (6A, 6A<sup>2</sup>, 6A<sup>3</sup>, 6A<sup>4</sup> and 6A<sup>5</sup>) for the SP-A1 gene and 6 alleles (1A, 1A<sup>0</sup>, 1A<sup>1</sup>, 1A<sup>2</sup>, 1A<sup>3</sup>, 1A<sup>4</sup>) for the SP-A2 gene. Families and twins were genotyped and showed

consistent patterns of allelic inheritance. An association between certain genotypes and low levels of mRNA [14] was possible through this genotyping method. Moreover, using the genotype data from >200 unrelated individuals and the EH linkage program [25] we were able to identify SP-A haplotypes [4]. Although all of these findings confirmed that this method is reliable, genotyping human SP-A variants by hybridization remained cumbersome and problematic.

In the present report, we describe a new method for SP-A and SP-D genotyping. These methods are based on a PCR-cRFLP analysis described previously in detail for SP-B [20]. PCR primers are designed to contain nucleotide(s) mismatched to the sequence flanking a single base polymorphism in order to create a restriction enzyme recognition site including the SNP. These are then used to amplify gene specific products that contain sequences from both alleles. The converted PCR products are digested with the appropriate enzyme, separated on PAGE and visualized with ethidium bromide staining, in order to assess whether the given sample is homozygous for either allele or heterozygous. The advantages of the new method include the lack of use of radioactivity, clearer results, time efficiency (no film exposures, filter stripping), and better reproducibility. The ability to identify and characterize surfactant protein genotypes in a timely, efficient, and reproducible manner, will facilitate population and family based association studies. Populations with different pulmonary diseases can be genotyped in the hope of identifying genetic polymorphisms that may in part, explain the genetic basis of individual variability to disease susceptibility, and therefore identify disease subgroups and/or response to treatment.

## METHODS AND MATERIALS

### Preparation of DNA for genotyping

A 200  $\mu$ L aliquot is removed from fresh blood samples and then DNA is prepared or the aliquot

is stored along with the remainder of the sample at  $-20^{\circ}\text{C}$ . Samples received frozen are kept frozen until DNA extraction, at which time these are allowed to thaw partially to remove enough for DNA preparation. DNA is prepared from 200  $\mu\text{L}$  of blood using a QIAamp Blood Kit (Qiagen) according to the Blood and Body Fluid Protocol provided by the manufacturer. The DNA is eluted in the final step using 200  $\mu\text{L}$  distilled water preheated to  $70^{\circ}\text{C}$ , which is then reheated to  $70^{\circ}\text{C}$  and passed through the same column a second time to maximize yield. The final elution volume is 200  $\mu\text{L}$ , and the expected DNA yield reported by the manufacturer is 4–12  $\mu\text{g}$  DNA (average 6  $\mu\text{g}$ ). We have found, by comparing the DNA intensity of many samples on agarose gels with known amounts of DNA, that the yield is much higher, approximately 14–18  $\mu\text{g}$  DNA in a total volume of 200  $\mu\text{L}$ .

### SP-A1 and SP-A2 Genotyping

#### PCR Amplification

SP-A1 and SP-A2 gene specific 3.3 kb fragments are amplified from genomic DNA prepared from blood. All PCR is done in an Ericomp Twin Block System thermocycler. Gene specific sense primers 326 (SP-A1) or 327 (SP-A2) are used with common antisense oligonucleotide 68A in a total volume of 30  $\mu\text{L}$  (see below for primer sequences). The reaction

mixture contains 1.5  $\mu\text{L}$  10x Buffer 1 and 1.5  $\mu\text{L}$  10x Buffer 2 (Roche), 2.4  $\mu\text{L}$  dNTPs (1.25 mM each dNTP), 0.3  $\mu\text{L}$  of each oligo (#327-100 ng/ $\lambda$ ; #326 and #68A-50 ng/ $\lambda$ ), and 0.52 units of Expand Long enzyme (Roche). The cycling conditions are an initial denaturation of  $95^{\circ}\text{C}$  for 2 min., followed by 33 cycles of  $95^{\circ}\text{C}$  30 sec.,  $58^{\circ}\text{C}$  30 sec. and  $72^{\circ}\text{C}$  3 min. followed by a final extension at  $72^{\circ}\text{C}$  for 5 min. Genomic clones containing an SP-A1 allele and an SP-A2 allele are routinely included as controls of gene specificity in each set of reactions. PCR products from each gene specific reaction are run on a 1% agarose gel and stained with ethidium bromide to assess the intensity of each product, and then diluted either 1:9 or 1:99 in  $\text{H}_2\text{O}$  (based on ethidium bromide DNA-stained intensity). These dilutions are used as template for further analysis in PCR-based cRFLP genotyping method as described below.

As shown schematically in Table 1, to identify the allele for an SP-A1 gene, one needs to determine nucleotides at five positions, aa 19, 50, 62, 133 and 219, and each codon corresponds to a different amino acid. The nucleotide change may or may not change the encoded amino acid. Similarly, for the designation of the SP-A2 alleles four such nucleotide determinations, aa 9, 91, 140 and 223, are needed. Table 2 shows the converted primers used for each amplification, the mismatch needed to create the recognition site, and the restriction endonuclease enzyme

Table 1  
The biallelic polymorphisms for both SP-A1 and SP-A2

SP-A1	aa 9 <sup>+</sup>	aa 19 <sup>*</sup>	aa 50 <sup>*</sup>	aa 62	aa 85	aa 91 <sup>+</sup>	aa 133	aa 140	aa 219 <sup>*</sup>	aa 223 <sup>+</sup>
	A	C or T Ala or Val	C or G Leu or Val	A or G	T	C	A or G	C	C or T Arg or Trp	C
SP-A2	A or C Asn or Thr	C	G	G	C	C or G Pro or Ala	A	C or T	C	A or C Lys or Gln

Various combinations of these biallelic markers on any given allele result in an allelic designation 1A<sup>x</sup> or 6A<sup>x</sup>. Currently, 15 alleles have been identified for SP-A1 and 15 alleles for SP-A2. Amino acid (aa) numbers followed by a \* indicate bp differences that result in a change of the particular amino acid between two alleles of SP-A1 and a+ indicates bp differences that result in a change of amino acid between two alleles of SP-A2. Amino acid 85 is one of the invariant amino acids that distinguishes SP-A1 from SP-A2 and the codon for this amino acid is used to confirm gene specificity. The corresponding amino acid appears below the nucleotide designation. The nucleotide changes at the other amino acid locations do not change the encoded amino acid.

used. Table 3 shows the other primers used in this study. To determine SP-A1 alleles, converted PCR (cPCR) is used for aa 19 and 50 with primer pair 787/765 and multiplex converted PCR for aa 62, 133 and 219 with primer pairs 788/21 (aa 62) and 767/768 (aa 133 and 219). To determine SP-A2 alleles, for aa 9 the primer pair 726/96 is used; for aa 223 the primer pair 805/494; and for aa 91 and aa 140 multiplex converted PCR is used with primer pairs 727/21 (aa 91) and 799/28A (aa 140).

The 30  $\mu$ L reaction conditions for the cPCR are as follows: 1  $\mu$ L (of 1:9 or 1:99 dilution from above) of the gene specific 3.3 kb PCR product, 1.5  $\mu$ L each of 10x Buffer 1 and 10x Buffer 2 (Roche), 1  $\mu$ L dNTPs (1.25 mM each dNTP), and 0.75 units of Taq DNA polymerase (Roche). The stock concentration of each set of the primers used in the 30  $\mu$ L reaction are: for 726/96 and 805/494, 100 ng/ $\lambda$ ; and for the remaining primer pairs 50 ng/ $\lambda$ . The volumes of the various primer pairs used for each 30  $\mu$ L reaction are: 788/21, 0.75  $\mu$ L; 805/494 and 726/96, 1.0  $\mu$ L; 787/765, 727/21 and 799/28A, 1.5  $\mu$ L; and 767/768, 3.0  $\mu$ L.

The cycling conditions include an initial denaturation at 95 °C 2 min. followed by 5 cycles at 95 °C 30 sec., 50 °C 1 min., 70 °C 1 min. and then at 95 °C 30 sec., 55 °C 1 min., 70 °C 1 min. for 27 cycles for primer pair 787/765. The same cycling conditions are used for primer pairs 805/494 and 726/96 except 28 and 30 cycles are used respectively. For primer pairs 788/21 and 767/768 the conditions are: 95 °C 30 sec., 58 °C 30 sec., and 72 °C 40 sec. for 30 cycles, and for 727/21 and 799/28 95 °C 30 sec., 58 °C 30 sec., 72 °C 30 sec. for 30 cycles. All reactions have a final extension of 72 °C 5 min.

In addition to using genomic clones to confirm the specificity of the initial 3.3 kb gene specific reaction, we also do a converted PCR reaction of the PCR gene specific products using primer pair, 807/18. This checks the nucleotide at aa 85 which is considered to be an invariant nucleotide and is used to distinguish the SP-A1 gene from the SP-A2 and their corresponding alleles [13]. The PCR conditions match those of primer pair 726/96 and the cycling conditions match those of primer pair 799/28 except 35 cycles are performed instead of 30.

To determine the presence of the 11 bp insertion in the SP-A2 3'UT (as discussed in the Results and Discussion Section) primer pair 854/855 is used in the PCR reaction. The reaction conditions are those of primers 726/96, with cycling conditions at 95 °C for 2 min followed by 35 cycles of 95 °C 30 sec., 58 °C 1 min., 72 °C 1 min. and then a final extension of 72 °C for 5 min.

#### *Restriction Endonuclease Analysis*

All reactions are done in a 10  $\mu$ L digestion volume. For double digestions, the following NEB buffers are used: primer pairs 727/21 and 799/28A, 10x buffer 4; primer pairs 788/21 and 767/768, 10x buffer 4 (for aa 62 and aa 133). BSA is added (final 100  $\mu$ g/ $\mu$ L) as instructed by the manufacturer. For each reaction, 3  $\mu$ L of PCR product is used and 7  $\mu$ L mix with the exception of primer pair 765/787 in which 5  $\mu$ L PCR product and 5  $\mu$ L mix is used. The enzymes required for each converted PCR are listed in Table 2 (All enzymes are purchased from New England Biolabs with the exception of Tru 91 from Promega). The following volumes of enzyme are used for digestion of each primer pair amplification. For SP-A1 primer pair 765/787, the two digestions are done separately using enzymes Bbv1, 0.15  $\mu$ L (2 u/ $\mu$ L) (for aa 19) and Dde1, 0.3  $\mu$ L (10 u/ $\mu$ L) (for aa 50); for primer pairs 788/21 and 767/768, a double digestion with Hha1 and Msp1, 0.1  $\mu$ L (20 u/ $\mu$ L) to determine aa 62 and aa 133. A separate digestion to determine aa 219 is performed using PCR product from 767/768 and  $\alpha$ Taq1, 0.2  $\mu$ L (20 u/ $\mu$ L). For SP-A2, primer pair 726/96, 0.15  $\mu$ L of Tru 91 (10 u/ $\mu$ L) is used; for primer pairs 727/21 and 799/28A, a double digestion with Bbv1, 0.15  $\mu$ L (2 u/ $\mu$ L) and Bfa1, 0.2  $\mu$ L (5 u/ $\mu$ L), and for primer pair 805/494,  $\alpha$ Taq1, 0.1  $\mu$ L (20 u/ $\mu$ L). Digestion of the PCR product from the 807/18 reaction to confirm gene specificity is done using 0.1  $\mu$ L Hha1 (20 u/ $\mu$ L). All digestions are run on 8% PAG, stained with ethidium bromide and samples compared to uncut PCR product and known controls.

Table 2  
Converted primers for cRFLP analysis of SP-A and SP-D

Polymorphism and Amino Acid (aa)	Converted PCR		Modified Primer Sequence‡ 5' → 3' lower case letter is mismatch	Location of modified primer in gene	Allelic products	
	Primers + Sense - Antisense	Fragment size			Alleles†	RE
<b>SP-A1</b>						
CT at aa 19	765 <sup>-</sup> /787 <sup>+</sup>	143 bp	787, Sense ACC TCA TCT TGA TGT•CAG CCT CTG GTG CaG	Exon I	<u>GCaGC</u> GCaGT	BbvI 103, 40 143
CG at aa 50	765 <sup>-</sup> /787 <sup>+</sup>	143 bp	765, Antisense AGG GCC AGG GTC TCC TcTGA	Exon I	cTGAG cTGAC	DdeI 123, 20 143
AG at aa 62	788 <sup>+</sup> /21 <sup>-</sup>	222 bp	788, Sense (TTT TCT CTG CAG) GCC CCA TGGGTgC	(Intron)	<u>gCGC</u> gCAC	HhaI 196, 26 222
AG at aa 133	767 <sup>+</sup> /768 <sup>-</sup>	299 bp	767, Sense TCT GCA GGG CTC CAT AAT Gc	Exon II	<u>gCGG</u> cCAG	MspI 279, 20 299
CT at aa 219	767 <sup>+</sup> /768 <sup>-</sup>	299 bp	768, Antisense CAC ACA CTG CTC TTT TCC tc	Exon IV	<u>tCGA</u> tCAA	αTaqI 278, 21 299
<b>SP-A2</b>						
AC at aa 9	726 <sup>+</sup> /96 <sup>-</sup>	151 bp	726, Sense GCT GTG CCC TCT GGC CCT tA	Exon I	<u>TtAA</u> TtAC	Tru 91 133, 18 151
CG at aa 91	727 <sup>+</sup> /21 <sup>-</sup>	132 bp	727, Sense AGA GCG TGG AGA GAA GGG GcA	Exon II	<u>GcAGC</u> GcAGG	BbvI 101, 31 132
CT at aa 140	799 <sup>+</sup> /28A <sup>-</sup>	187 bp	799, Sense CAT AAT GAC AGT AGG AGA GAA GGT CTT CTC	Exon IV	<u>CTAG</u> CCAG	BfaI 157, 30 187
AC at aa 223	805 <sup>+</sup> /494 <sup>-</sup>	106 bp	805, Sense GGA GCC TGC AGG TCG GGG AAA AtcG	Exon IV	<u>tCGA</u> tCGC	αTaqI 83, 23 106
<b>SP-A1 SP-A2</b>						
CT at aa 85	807 <sup>-</sup> /18 <sup>+</sup>	151 bp	807, Antisense AGC CTC CCC CTT CTC TCC gc	Exon II	<u>gCGC</u> gCAC	HhaI 130, 21 151
<b>SP-D</b>						
CT at aa 11	825 <sup>+</sup> /920 <sup>-</sup>	59 bp	920, Antisense ACC AGG GTG CAA GCA CTG cG	Exon I	<u>TGcGCA</u> TGcGCG	FspI 40, 19 59
AG at aa 160	934 <sup>+</sup> /935 <sup>-</sup>	163 bp	934, Sense AGC GTG GAG TCC CTG GAA gC	Exon IV	<u>gCGC</u> gCAC	HhaI 142, 21 163

— Restriction endonuclease recognition site completed by one bp of polymorphic site.

• Mismatch (from G to T) to block natural BbvI site within primer.

† Flanking sequence at the polymorphic site is shown. Bold: polymorphic site. If the modified primer is a sense primer, the sequence is shown as plus strand, if modified primer is an antisense primer the sequence is shown as minus strand.

‡ Sequence according to White et al. [27] (SP-A1), Katyal et al. [15] (SP-A2) and Crouch et al. [1] (SP-D).

## SP-D Genotyping

### PCR Amplification

The cRFLP method is used for two biallelic polymorphisms at the SP-D codons for aa11 and aa160 (our unpublished observations). To improve the final PCR product yield and specificity, a larger fragment is first amplified from genomic DNA, which then serves as template for a nested reaction using one converted primer. Converted primers and the corresponding information are shown in Table 2, all other primers are listed in Table 3. For the polymorphism at the codon for amino acid 11, the first reaction is with primer pair 930/999 and the nested PCR with primers 825/920. For the polymorphism at the codon for amino acid 160, the initial reaction is with primer pair 904/936 and the nested PCR with primers 934/935. The 30  $\mu$ L reaction conditions for all four reactions are as follows: 1.5  $\mu$ L 10x buffer 1 and 1.5  $\mu$ L 10x buffer 2 (Roche), 1  $\mu$ L dNTPs (1.25 mM each dNTP), 0.75 units of Taq DNA polymerase (Roche), and 1  $\mu$ L of each primer. Stock concentration of primer pairs 930/999 and 904/936 is 100 ng/ $\lambda$ , and for 934/935 and 825/920 is 25 ng/ $\lambda$ . Cycling con-

ditions for all four reactions are 95 °C for 2 min. followed by 35 cycles of 95 °C 30 sec., 58 °C 1 min., 72 °C 1 min. and a final extension of 72 °C for 5 min.

### Restriction Endonuclease Analysis

All reactions are done in a 10  $\mu$ L reaction volume with 0.7x final concentration of the recommended buffer and BSA (final 100  $\mu$ g/ $\mu$ L) if needed. For each reaction 8  $\mu$ L of PCR product is used and 2  $\mu$ L of digestion mix. The PCR product generated with primer pair 825/920 is digested with 0.2  $\mu$ L (5 u/ $\mu$ L) of FspI, and for primer pair 934/935, is digested with 0.3  $\mu$ L (20 u/ $\mu$ L) of HhaI. The digested products are run on 10% and 8% PAGE respectively, stained with ethidium bromide, and samples compared to the uncut PCR product.

### Cloning and sequencing analysis of a new SP-A1 allele, the 6A<sup>14</sup>

Based on the genotyping results of > 2500 samples, the frequency of 6A<sup>14</sup> was > 0.01. Therefore, it became important to verify the 6A<sup>14</sup> sequence by cloning and sequencing of the 6A<sup>14</sup>

Table 3  
Non-converted primers for cRFLP analysis of SP-A and SP-D

	Primer	Orientation	Primer Sequence (5' → 3')*	Location in gene
SP-A	18	Sense	ACC AGT TGT GGG TGA CAGAT	Intron I
	21	Antisense	GGGTTTGTCTGATCCCCATC	Intron II
	28A	Antisense	ACCCTCAGTCAGGCCTACAT	Exon IV
	68A	Antisense	TGCCACAGAGACCTCAGAGT	3'UT
	96	Antisense	TCC TTT GAC ACC ATC TC	Exon I
	326	Sense	ACT CCA TGACTGACCACCTT	5'UT
	327	Sense	ATC ACT GAC TGT GAGAGGGT	5'UT
	494	Antisense	TCA GAA CTC ACA GAT GGT CA	Exon IV
	854	Sense	GAT CCT CCA AAG TCA CAC	3'UT
	855	Antisense	TCA GAG TGC CAG AGA ATC	3'UT
	SP-D	825	Sense	CTG GAA GCA GAA ATG AAGAC
904		Antisense	AGG CCC AAA AGG AGA AGC TG	Exon III
930		Sense	TCA CCT CTA GAA GCT GAG CCA AGCC	Intron I
935		Antisense	AGA TTC TCT CCA TGT TCC CAG	Intron IV
936		Antisense	TCT CTG CCC ACC CAT CCCAG	Intron V
999		Antisense	ACA AAG TAC CCA GAG TTG CTG	Intron II

\* Nucleotide sequence according to White et al. [27] (SP-A1), Katyal et al. [15] (SP-A2), and Crouch et al. [1] 1993 (SP-D).

DNA. SP-A1 specific PCR products are generated as described above using as template 2 genomic DNAs homozygous for the 6A<sup>14</sup> allele. Two separate PCR ligation and cloning reactions were performed for each of the two genomic DNAs using standard methodologies. Four positive 6A<sup>14</sup> clones were picked (2 from each cloning) and sequenced to eliminate the possibility that the observed 6A<sup>14</sup> sequence was the result of PCR errors. The sequence of each of the four 6A<sup>14</sup> clones was identical except for aa 202. Two clones from one individual showed a C at aa 202 and two clones from another individual showed a T at aa 202. This results in 6A<sup>14</sup> being 2 different alleles: 6A<sup>14</sup>(C) and 6A<sup>20</sup>(T) when including aa 202 in the genotyping procedure. Following this finding, we genotyped 26 previously scored 6A<sup>14</sup> alleles using converted PCR for aa 202. Of these, 65% contain a “C” at aa 202 and 35% indicate a “T” at aa 202. With this additional resolution, the frequency of the 6A<sup>14</sup> allele is <0.01 (0.7%). The sequence of 6A<sup>14</sup> was compared to SP-A1 alleles. The 6A<sup>14</sup> sequence is identical to that of 6A<sup>2</sup> except in the 2 locations, encoding amino acids 50 and 62, as shown below.

	<u>aa50</u>		<u>aa62</u>
6A <sup>2</sup>	ATG <u>GTG</u> TCAAA	GGT <u>CCA</u> CCTGG	
	Val	Pro	
6A <sup>14</sup>	ATG <u>GTC</u> TCAAA	GGT <u>CCG</u> CCTGG	
	Leu	Pro	

## RESULTS AND DISCUSSION

### Scoring of known SP-A alleles

We have developed a non-radioactive method for genotyping of human SP-A1 and SP-A2 alleles. This method is based on PCR-cRFLP analysis and yields reliable results for all known SP-A alleles in a time efficient manner. The scoring of the SP-A alleles is done after evaluation of the digested products of the converted PCR and comparison to the corresponding undigested PCR product. From this, it can be determined whether a sample is completely

uncut, completely cut or heterozygous at the biallelic polymorphic site. In some heterozygous samples the digested product is sometimes of low intensity upon ethidium bromide staining, compared to the uncut or in some homozygous cut samples a trace of uncut DNA may be visible. By testing many samples, the appropriate appearance of bands for a heterozygous sample can be determined. Thus, the pattern that is consistently reproduced for a given set of primers is used as the standard to determine all heterozygotes for that specific primer pair. Figure 1A (lanes 5, 8) depicts an example of a digested product with low staining intensity compared to the uncut allele, and Figure 1B lane 2 depicts two products of equal intensity following enzyme digestion (see details below).

Figures 1 and 2 show the expected digestion patterns for all primer pairs used in this genotype analysis. Primer pairs and the corresponding amino acid are listed below the lanes of the gels. After digestion, a sample will appear as homozygous uncut, heterozygous, or homozygous cut depending on the nucleotide contributed by the polymorphic site. The resulting fragment sizes reflect the specific nucleotide present at the polymorphic site. Table 2 shows the restriction enzyme, the 2 possible alleles, and the resulting fragment sizes for each allele. The details of alleles indicated by each band are shown in figure legends. Figure 1, Panel A shows the gel with the PCR digestion products for the 5 amino acids necessary to assign a genotype for SP-A1. Lanes 1–3 are for aa 19, lanes 4–6, aa 50, lanes 7–9 are for both aa 62 (lower portion of lane) and aa 133 (upper portion of lane), and lanes 10–12 are for aa 219. For example, in lanes 1–3 amplified product using oligos 765/787 has been digested with Bbv1. If the “T” allele is present, no digestion will occur and a 143 bp band is visible. If a “C” allele is present the 143 bp product will be digested to yield 2 bands of 103 bp and 40 bp (the 40 bp product is not visible on the gel). The 3 possible combinations of the 2 alleles results in the bands seen in lanes 1–3, lane 1, TT, lane 2, TC, and lane 3, CC. The lower band (~ 222 bp), in lanes 10–12 is the amplified product of the primer pair 788/21 from the multiplex PCR.

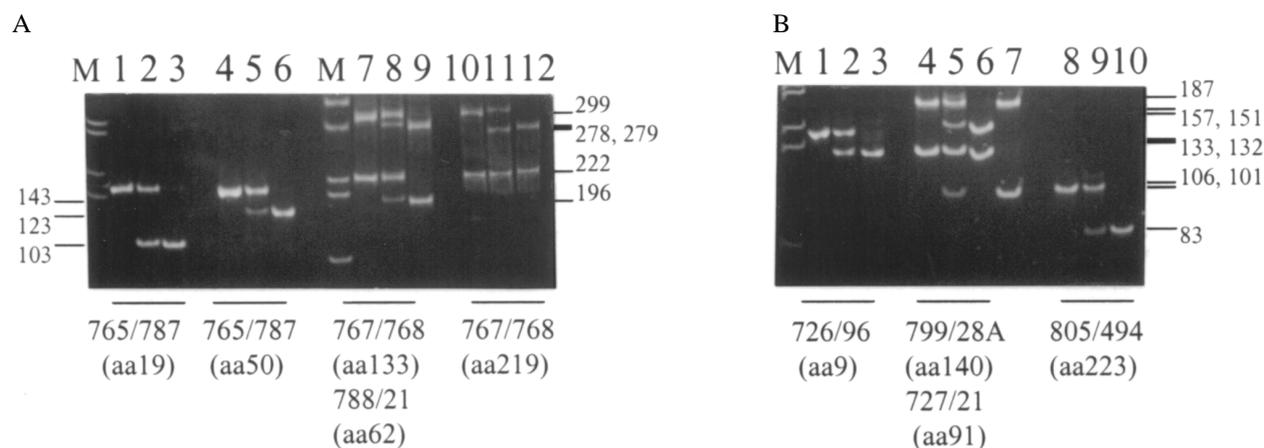


Fig. 1. Allelic patterns for SP-A1 and SP-A2. The fragments containing SNPs were amplified by converted PCR with the primer pairs listed below the gel. The PCR products were digested with appropriate restriction enzymes (Table 2), separated by 8% PAGE, and stained with ethidium bromide.

**Panel A** shows all possible SP-A1 allelic patterns for aa 19, aa 50, aa 62, aa 133, and aa 219 as amplified by the primer pairs listed below the gel. Corresponding fragment sizes can be found in Table 2. Lanes 1–3 show the possible genotypes for aa 19; lane 1, homozygous uncut (TT), lane 2, heterozygous (CT), and lane 3, homozygous cut (CC). Lanes 4–6 for aa 50; lane 4, homozygous uncut (GG), lane 5, heterozygous (CG) and lane 6, homozygous cut (CC). Lanes 7–9 show the combined digestions for aa 62 (lower portion of lane) and aa 133 (upper portion of lane) of the products from the multiplex PCR reaction; lane 7, homozygous uncut (AA, aa 133; AA, aa 62), lane 8, heterozygous (AG, aa 133; AG, aa 62), lane 9, homozygous cut (GG, aa 133; GG, aa62). Lanes 10–12 show the genotypes for aa 219 (upper portion of lane) (lower portion of lane depicts the undigested amplified product of aa 62 from the multiplex PCR reaction); lane 10, homozygous uncut (TT), lane 11, heterozygous (CT), and lane 12, homozygous cut (CC). Marker is 1 kb ladder.

**Panel B** shows all possible SP-A2 allelic patterns for aa 9, aa 91, aa 140 and aa 223 as amplified by the primer pairs listed below the gel. Lanes 1–3 show the 3 possibilities for aa 9; lane 1, homozygous uncut (CC), lane 2, heterozygous (AC), and lane 3, homozygous cut (AA). Lanes 4–7 show the combined digestions for aa 91 (lower portion of lane) and aa 140 (upper portion of lane) of the products from the multiplex PCR reaction; lane 4, homozygous uncut (CC, aa 140; GG, aa91), lane 5, heterozygous (CT, aa 140; GC, aa 91), lane 6 homozygous cut (TT, aa 140), and lane 7, homozygous cut (CC, aa 91). Lanes 8–10 shows 3 genotypes for aa 223; lane 8, homozygous uncut (CC), lane 9, heterozygous (AC), and lane 10, homozygous cut (AA). The marker is 1 kb ladder (Gibco BRL).

Figure 1 panel B shows the gel with the digestion products of the 4 amino acids necessary to assign an SP-A2 genotype. Lanes 1–3 show the digestion pattern for aa 9, lanes 4–7 for multiplex PCR and double digestion for aa 91 (lower portion of lane) and aa 140 (upper portion of lane), and lanes 8–10 for aa 223. Not shown are the digestion products for aa 85 which, if gene specific, result in either an uncut (SP-A1) or a cut (SP-A2) product.

Figure 2 shows the digestion patterns for the two biallelic polymorphisms in SP-D. Lanes 1–3 are for aa 11 and lanes 4–6 for aa 160.

For any given amino acid, genomic controls are determined, sequenced at the polymorphic site, and then used in subsequent reactions as controls for the given digestion.

#### Genotype analysis reveals new SP-A patterns

When the genotyping results of a given genomic DNA sample reveals an SP-A biallelic variant pattern that has not been previously observed then the SP-A coding regions are sequenced to confirm the particular pattern. For example, when alleles 1A<sup>4</sup> and 6A<sup>5</sup> were first

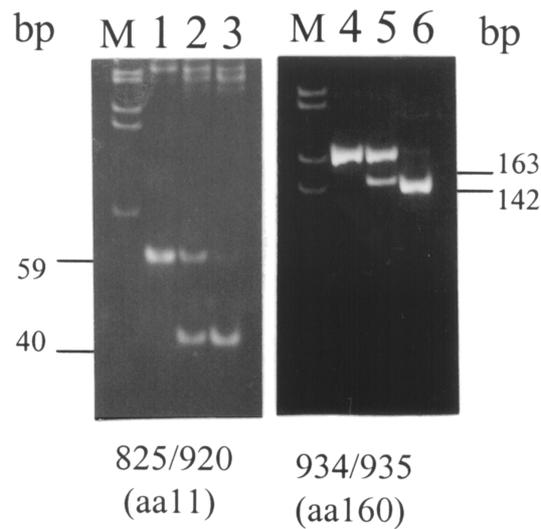


Fig. 2. Allelic patterns for SP-D. The fragments containing SNPs at aa 11 and aa 160 were amplified by converted PCR with the primer pairs listed below the gel. The PCR products were digested with appropriate restriction enzymes (Table 2), separated by 8% or 10% PAGE, and stained with ethidium bromide.

Lanes 1–3 show the 2 alleles at aa 11; lane 1, homozygous uncut (CC), lane 2, heterozygous (CT), lane 3, homozygous cut (TT). Lanes 4–6 show the 2 alleles at aa 160; lane 4, homozygous uncut (AA), lane 5, heterozygous (AG), and lane 6, homozygous cut (GG).

discovered [13] and were completely sequenced these were found to contain nucleotide differences at amino acid 133 (SP-A1), and at amino acids 140 and 202 (SP-A2). Although these 3 polymorphisms had not been previously targeted with the hybridization procedure [4], these polymorphic sites were subsequently incorporated into the current genotyping methodology. After studying an additional 200 samples, amino acid 202, which was useful only for the detection of the rare  $1A^4$  allele [13], was omitted from the methodology, due to its limited value. Because we are now using biallelic markers at aa 133 and aa 140, we are able to distinguish more alleles than was possible with the old hybridization method (Tables 4 and 5).

### Challenges in assigning genotypes

As combinations of nucleotides at specified amino acids are determined, problems can occur in assigning genotypes. A genotype of  $1A^01A^1$  has the same pattern as a genotype of  $1A^21A^3$  (Table 4), and thus it is not possible, without

additional information or prior relevant knowledge, to determine which SP-A2 genotype is correct. This difficulty stems from the fact that both SP-A2 alleles are observed simultaneously and we cannot determine which nucleotides are contributed by each SP-A2 allele. The SP-A1 and SP-A2 genes are linked [10] and association studies of unrelated individuals have previously provided SP-A haplotype information that we use at times to distinguish between possible genotypes [4]. The  $1A^0$  allele has been shown to associate with  $6A^2$ ,  $6A^3$ , or  $6A^4$ , the  $1A^1$  allele with  $6A^2$  or  $6A^3$ , and the  $1A^2$  has been shown to associate only with the  $6A^4$  allele. Therefore, if the SP-A1 genotype is known and a  $6A^4$  allele is not present we assume that the SP-A2 genotype is  $1A^01A^1$  and not  $1A^21A^3$ . Currently, all assigned genotypes of  $1A^01A^1$  have corresponded to  $6A^26A^3$ . Data from CEPH families (our unpublished observations) have confirmed haplotypes  $1A^16A^3$  and  $1A^06A^2$ .

There is no difficulty in scoring all currently known fully sequenced (Table 5) SP-A1 alleles, individually or in combination. However, a

Table 4  
SP-A2 alleles as determined by cRFLP at designated amino acids

Alleles	Amino acid			
	aa 9	aa 91	aa 140	aa 223
1A*	C	C	C	C
1A <sup>0</sup> *	A	G	C	C
1A <sup>1</sup> *	C	G	T	A
1A <sup>2</sup> *	C	G	C	C
1A <sup>3</sup> *	A	G	T	A
1A <sup>4</sup> *	C	G	T	A
1A <sup>5</sup> *	C	C	T	C
1A <sup>6</sup>	C	G	T	C
1A <sup>7</sup>	A	C	T	C
1A <sup>8</sup>	C	G	C	A
1A <sup>9</sup>	A	G	T	C
1A <sup>10</sup>	C	C	T	A
1A <sup>12</sup>	C	C	C	A
1A <sup>12</sup>	A	C	C	C
1A <sup>13</sup>	A	C	T	A

\* Fully sequenced alleles.

Table 5  
SP-A1 alleles as determined by cRFLP at designated amino acids

Alleles	Amino acid				
	aa 19	aa 50	aa 62	aa 133	aa 219
6A*	C	C	G	G	C
6A <sup>2</sup> *♣	T	G	A	A	C
6A <sup>3</sup> *	T	C	A	A	C
6A <sup>4</sup> *	T	C	G	A	T
6A <sup>5</sup> *	C	C	G	A	T
6A <sup>6</sup>	T	G	G	G	C
6A <sup>8</sup>	T	C	A	A	T
6A <sup>9</sup>	T	C	G	G	T
6A <sup>10</sup>	C	C	G	A	C
6A <sup>11</sup>	C	C	A	A	C
6A <sup>12</sup>	T	G	A	A	T
6A <sup>13</sup>	T	G	G	A	C
+6A <sup>14</sup> *	T	C	G	A	C
6A <sup>15</sup> ♣	C	G	G	A	C
6A <sup>16</sup>	T	C	A	G	C
6A <sup>17</sup>	C	C	A	G	C
6A <sup>18</sup>	T	G	G	A	T
6A <sup>19</sup>	C	G	G	G	C
+6A <sup>20</sup>	T	C	G	A	C

\* Fully sequenced alleles; ♣ contain 11 bp insertion in 3'UT; + patterns shown here are identical; inclusion of aa 202 distinguished 6A<sup>14</sup>(C) from 6A<sup>20</sup>(T) (see text).

challenge may arise to correctly assign a 6A<sup>2</sup> or 6A<sup>3</sup> allele when the sequence pattern of either one of these appears in combination with a possibly new, unknown allele. To correctly score the 6A<sup>2</sup> and 6A<sup>3</sup> alleles, there is an additional PCR reaction that takes advantage of an 11bp size difference found in 3'UTR approximately 400 bp past the translation termination codon [18]. All of the more abundant SP-A2 alleles tested (1A, 1A<sup>0</sup>, 1A<sup>1</sup>, 1A<sup>2</sup>, 1A<sup>3</sup>, 1A<sup>5</sup>) and only the

SP-A1, 6A<sup>2</sup> and the rare 6A<sup>15</sup>, alleles contain this 11 bp insertion. The 11 bp determination has become essential for distinguishing 6A<sup>2</sup> and 6A<sup>3</sup> under certain circumstances. In this reaction, we use gene-specific template to amplify the region flanking the 11 bp and then the size difference is visualized by PAGE. A band size of 187 bp indicates that the 11 bp insertion is present and a 6A<sup>2</sup> can be assigned. Alternatively, a 176 bp band indicates that the 11 bp is missing and a 6A<sup>3</sup>

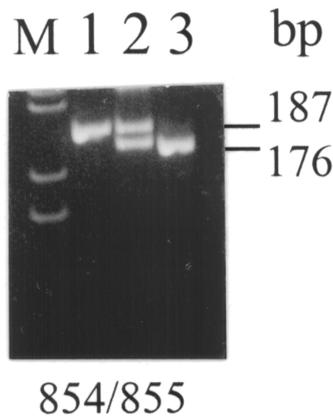


Fig. 3. PCR to determine the presence of 11 bp fragment in 3' UT of SP-A alleles. This figure shows the amplified size difference resulting from presence or absence of 11 bp fragment in 3' UT of SP-A1 alleles. Lane 1 shows a sample homozygous for both alleles containing the 11 bp fragment (6A<sup>2</sup>6A<sup>2</sup>), lane 2 a sample heterozygous for the 11 bp (6A<sup>2</sup>6A<sup>3</sup>) and lane 3, a sample homozygous for absence of 11 bp fragment (6A<sup>3</sup>6A<sup>3</sup>).

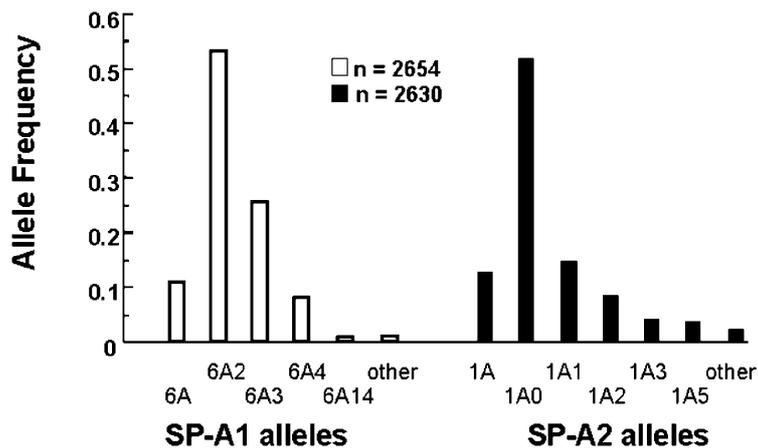


Fig. 4. Human SP-A allele frequency. The allele frequency is based on general population studies of > 2600 genotypes. The frequency only of the alleles that occur at > 1% in the general population is shown.

can be assigned (Figure 3). Once the correct designation of the 6A<sup>2</sup> or 6A<sup>3</sup> allele is made, the remaining pattern is designated as the other (new) allele and can be assessed at the specified nucleotide positions. Using this method we genotyped more than 2500 samples to determine the frequency of the SP-A alleles in the general population. The distribution of the more abundant (> 1%) SP-A1 and SP-A2 alleles is shown in Figure 4. Rare alleles (< 1%) are not shown.

### Comments

Human SP-A exhibits extensive complexity and heterogeneity at the genetic [5,14], mRNA level [13], and protein levels [16,26], and plays multiple roles in two major groups of functions: the surfactant related functions [3,11,17,24] and those of innate host defense and the regulation of inflammatory processes [23]. Altered levels of SP-A have been observed in several diseases, and SP-A has been implicated in the susceptibility to and/or pathogenesis of several pulmonary diseases [6]. We believe that the genotype method described in this report will be essential for SP-A genotype-phenotype correlations in population and family based studies for several disease groups.

### Acknowledgement

This work was supported by NIH R37HL34788-S1.

### References

- [1] Crouch, E., Rust, K., Veile, R., Doms-Keller, H. and Grosso, L. Genomic organization of human surfactant protein D (SP-D). *J. Biol. Chem.* **268**, (1993) 2976–2983.
- [2] deMello, D.E., Hayman, S., Phelps, D. and Floros, J. Immunolocalization of SP-A in lungs of infants dying from respiratory distress syndrome. *Am. J. Pathol.* **142**, (1993) 1631–1640.
- [3] deMello, D.E., Chi, E.Y., Dou, E. and Lagunoff, D. Absence of tubular myelin in lungs of infants dying with hyaline membrane disease. *Am. J. Pathol.* **127**, (1987) 131–139.
- [4] Floros, J., DiAngelo, S., Koptides, M., Karinch, A.M., Rogan, P., Nielsen, H., Spragg, R.G., Watterberg, K. and Deiter, G. Human SP-A locus: Allele frequencies and linkage disequilibrium between the two surfactant protein A genes. *Am. J. Respir. Cell Mol. Biol.* **15**, (1996) 489–498.
- [5] Floros, J. and Hoover, R.R. Genetics of the hydrophilic surfactant proteins A and D. *Biochim. Biophys. Acta* **1408**, (1998) 312–322.
- [6] Floros, J. and Kala, P. Surfactant proteins: Molecular genetics of neonatal pulmonary disease. *Ann. Rev. Physiol.* **60**, (1998) 365–384.
- [7] Floros, J. and Lin, Z. Genetic variability of surfactant protein-B and respiratory distress syndrome: Clinical implications. *Medscape Respiratory Care*.  
[http://www.medscape.com/Medscape/respiratory\\_Care/journal/1999/v03.n02/mrc4664.flor/mrc4664.flor01.html](http://www.medscape.com/Medscape/respiratory_Care/journal/1999/v03.n02/mrc4664.flor/mrc4664.flor01.html).
- [8] Floros, J. and Phelps, D.S. Pulmonary surfactant. In: *Anesthesia: Biologic Foundations*. Biebuyck, J.F., Lynch, C. III, Maze, M., Saidman, L.J., Yaksh, T.L. and Zapol, W.M. (eds.) Raven, New York (1997) 1257–1279.
- [9] Gregory, T.J., Longmore, W.J., Moxley, M.A., Whitsett, J.A., Reed, C.R., Fowler, A.A., et al. Surfactant chemical composition and biophysical activity in acute respiratory distress syndrome. *J. Clin. Invest.* **88**, (1991) 1976–1981.
- [10] Hoover, R.R. and Floros, J. Organization of the human SP-A and SP-D loci at 10q22-23. Physical and radiation hybrid mapping reveals gene order and orientation. *Am. J. Respir. Cell Mol. Biol.* **18**, (1998) 353–362.
- [11] Ikegami, M., Korfhagen, T.R., Bruno, M.D., Whitsett, J.A. and Jobe, A.H. Surfactant metabolism in surfactant protein A-deficient mice. *Am. J. Physiol.* **272**, (1997) L479–L485.
- [12] Kala, P., Ten Have, T., Nielsen, H., Dunn, M. and Floros J. Association of pulmonary surfactant protein A (SP-A) gene and RDS: Interaction with SP-B. *Pediatr. Res.* **43**, (1998) 1–9.
- [13] Karinch, A.M., Deiter, G., Ballard, P.L. and Floros, J. Regulation of expression of human SP-A1 and SP-A2 genes in fetal lung explant culture. *Bioch. Biophys. Acta.* **1398**, (1998) 192–202.

- [14] Karinch, A.M., deMello, D.E. and Floros, J. Effect of genotype on the levels of surfactant protein-A mRNA and on the SP-A2 splice variants in adult humans. *Biochem. J.* **321**, (1997) 39–47.
- [15] Katyal, S.L., Sugh, G. and Locker, J. Characterization of a second pulmonary surfactant-associated protein SP-A gene. *Am. J. Respir. Cell Mol. Biol.* **6**, (1992) 446–452.
- [16] Phelps, D.S., Floros, J. and Taeusch, H.W. Jr. Post-translational modification of the major human surfactant-associated proteins. *Biochem. J.* **237**, (1986) 373–377.
- [17] Korfhagen, T.R., Bruno, M.D., Ross, G.F., Huelsman, K.M., Ikegami, M., Jobe, H.H., Wert, S.E., Stripp, B.R., Mums, R.E., Glasser, S.W., Bachurski, C.J., Iwamoto, H.S. and Whitsett, J.A. Altered surfactant function and structure in SP-A gene targeted mice. *Proc. Natl. Acad. Sci. USA* **93**, (1996) 9594–9599.
- [18] Krizkova, L., Sakthivel, R., Olowe, S.A., Rogan, P. and Floros, J. Human SP-A: Genotype and single strand conformation polymorphism analysis. *Am. J. Physiol.* **266** (*Lung Cell. Mol. Physiol.* **10**) (1994) L519–L527.
- [19] Lin, Z., deMello, D.E., Batanian, J.R., Khammash, H.M., DiAngelo, S., Luo, J. and Floros, J. Aberrant SP-B mRNA in lung tissue of patients with congenital alveolar proteinosis. (1999) Submitted.
- [20] Lin, Z., deMello, D.E., Wallot, M. and Floros, J. An SP-B gene mutation responsible for SP-B deficiency in fatal congenital alveolar proteinosis: Evidence for a mutation hotspot in exon 4. *Mol. Genet. Metab.* **64**, (1998) 25–35.
- [21] Moya, F.R., Montes, H.F., Thomas, V.L., Mouzinho, A.M., Smith, J.F. and Rosenfeld, C.R. Surfactant protein A and saturated phosphatidylcholine in respiratory distress syndrome. *Am. J. Respir. Crit. Care Med.* **150**, (1994) 1672–1677.
- [22] Noguee, L.M., Garnier, G., Dietz, H.C., Singer, L., Murphy, A.M., deMello, D.E. and Colten, H.R. A mutation in the surfactant protein B gene responsible for fatal neonatal respiratory disease in multiple kindreds. *J. Clin. Invest.* **93**, (1994) 1860–1883.
- [23] Phelps, D.S. Pulmonary surfactant modulation of host-defense function. *Appl. Cardiopulm. Pathophysiol.* **5**, (1995) 221–229.
- [24] Suzuki, Y., Fujita, Y. and Kogishi, K. Reconstitution of tubular myelin from synthetic lipids and protein associated with pig pulmonary surfactant. *Am. J. Respir. Dis.* **140**, (1989) 75–81.
- [25] Terwillinger, J.D. and Ott, J. Handbook of human genetic linkage. The Johns Hopkins University Press, Baltimore (1994) 188–210.
- [26] Voss, T., Eistetter, H., Schafer, K.P. and Engel, J. Macromolecular organization of natural and recombinant lung surfactant protein SP 28-36. Structural homology with complement factor Clq. *J. Mol. Biol.* **20**, (1988) 219–227.
- [27] White, M., Damm, D., Miller, J., Spratt, K., Schilling, J., Hawgood, S., Benson, B. and Cordell, B. Isolation and characterization of the human pulmonary surfactant apoprotein gene. *Nature* **317**, (1985) 361–363.



**Hindawi**  
Submit your manuscripts at  
<http://www.hindawi.com>

