Human surfactant protein – A gene locus for genetic studies in the Finnish population

Mika Rämet\textsuperscript{a,b}, Ritva Haataja\textsuperscript{a}, Riitta Marttila\textsuperscript{a,b}, Anu-Maaria Hämaläinen\textsuperscript{c}, Mikael Knip\textsuperscript{c} and Mikko Hallman\textsuperscript{a,b,*}

\textsuperscript{a}Biocenter Oulu, Oulu, Finland
\textsuperscript{b}Department of Paediatrics, University of Oulu, Oulu, Finland
\textsuperscript{c}Department of Paediatrics and Medical School, University of Tampere, Tampere, Finland

Received 18 May 2000

Lung surfactant lowers the surface tension but surfactant proteins also have other functions. Surfactant protein A (SP-A) has a well-defined role in innate immunity. The gene locus for human SP-A genes is in chromosome 10q21 through q24 and consists of two highly homologous functional SP-A genes (SP-A1 and SP-A2) and a pseudogene. Several alleles that differ by a single amino acid have been identified for both SP-A genes. The SP-A gene locus has been shown to be sufficiently polymorphic for genetic studies in the American population. In this study, we analysed the SP-A allele frequencies in a Finnish population (\(n = 790\)) and found them to differ from the frequencies observed in US. Furthermore, we describe several new alleles for both SP-A genes. The heterozygosity indices and polymorphism information content values ranged between 0.50–0.62 indicating that SP-A gene locus is polymorphic enough for studies associating the locus with pulmonary diseases.

Keywords: Surfactant, protein-A, genetics

1. Introduction

Surfactant is a complex mixture of lipids and proteins. Its best-characterised function is its ability to lower surface tension at the air-liquid interface, thus preventing lung collapse [1]. Surfactant also has other functions: It participates in innate host defence and takes part in the inflammatory processes of the lung [2]. The products of the surfactant protein (SP-A, SP-B, SP-C and SP-D) genes are crucial for many aspects of surfactant biology and physiology [2].

SP-A is the most abundant surfactant protein. It is a member of a family of collagenous carbohydrate-binding proteins called collectins [2]. The gene locus for human SP-A genes is located on chromosome 10q21 through q24 [5] and consists of two functional SP-A genes (SP-A1 and SP-A2) and a pseudogene. Several alleles that differ by a single amino acid have been identified for both SP-A genes. The SP-A gene locus has been shown to be sufficiently polymorphic for genetic studies in the American population. In this study, we analysed the SP-A allele frequencies in a Finnish population (\(n = 790\)) and found them to differ from the frequencies observed in US. Furthermore, we describe several new alleles for both SP-A genes. The heterozygosity indices and polymorphism information content values ranged between 0.50–0.62 indicating that SP-A gene locus is polymorphic enough for studies associating the locus with pulmonary diseases.

2. Materials and methods

2.1. Sample collection and study population

The sample collection for this study was carried out in the Department of Paediatrics, University of Oulu, Finland. The parents of the neonates gave written informed consent for their infants’ blood samples to be used in this study. The ethical committee of the Uni-
University Central Hospital approved the study protocol. Samples were collected from infants born in 1997–99.

Two hundred fifty five cord blood specimens were obtained from all infants born at term (i.e. gestation $\geq 37.0$ weeks) during one month in the Oulu University Hospital. In addition, 354 blood samples from premature infants and 211 from infants born at term were analysed for SP-A genotypes. Altogether 790 blood samples were genotyped for the SP-A genes.

### 2.2. DNA samples

Whole blood samples (0.5–3 ml) were collected from the cord into plastic EDTA tubes and stored at $-70^\circ$C until analysis. Genomic DNA was isolated from the whole blood specimens using the Puragene DNA isolation Kit (Gentra Systems). An aliquot of the DNA solution was diluted to 50 ng/l to be used for PCR amplification. When whole blood samples were not available (66 subjects), genotypes were determined using a blood spot dried on a filter paper. A 3 mm disk (corresponding to about 12,000 white blood cells) was punched from the blood spot using a hand-held paper punch (Wallac, Finland). The punch was decontaminated between samples by multiple punching of clean filter paper. DNA was bound to the disk and cellular contaminants were released by three successive 15-minute incubations with 50% ethanol. After drying at 55°C or room temperature, the purified paper disk was directly used as a template for PCR amplification. A blank paper disk treated in a similar manner was included in each series of PCR reactions as a control for DNA cross-contamination.

### 2.3. Sequencing human SP-A1 and SP-A2 genes

Gene-specific amplifications of SP-A1 and SP-A2 genes were performed using two sets of gene-specific primers, with the second set nested with the first. A gene-specific forward primer 5'−ACTCCATGACTGAC CACCTT-3' at nucleotide position 469–488 of the SP-A1 gene (Katyal et al. [13]) and 5'−ATCAGCTGATG AGAGGT-3' at position 472–491 of the SP-A2 gene were used together with a common reverse primer 5'−TGCCACAGACCTCAGAT-3' at position 3364–3387 (SP-A1) or 5'−GAACTGAGGCC AGACAGGA-3' at position 3374–3397 (SP-A2). The 50 µl reaction mixture contained 5 µl of 10x PCR buffer, 0.5 µl of AmpliTaq Gold Polymerase, 0.1 mM of dNTPs and 0.2 M of each primer. The cycling conditions were as follows: initial denaturation at 95°C for 5 min, followed by 26 cycles at 95°C for 30 s, 58°C for 30 s and 72°C for 2 min 30 s, followed by final extension at 72°C for 5 min. For the nested PCR reaction, a common forward primer 5'−GATGGGCTCACGGCCATCCC-3' at position 1023–1042 was used together with a gene-specific reverse primer 5'−GAGGCCGAAGGCCAGAGCAG-3' at position 3364–3387 (SP-A1) or 5'−GAACTGAGGCC AGACAGGA-3' at position 3374–3397 (SP-A2). The 50 µl reaction mixture contained 5 µl of 10x PCR buffer, 0.5 µl of AmpliTaq Gold Polymerase, 0.1 mM of dNTPs and 0.2 M of each primer. The cycling conditions were as follows: initial denaturation at 95°C for 5 min, followed by 26 cycles at 95°C for 30 s, 58°C for 30 s and 72°C for 2 min 30 s, followed by final extension at 72°C for 5 min. For the nested PCR reaction, a common forward primer 5'−GATGGGCTCACGGCCATCCC-3', exon 3 for both SP-A genes using a forward primer 5'−GAGGCCGAAGGCCAGAGCAG-3' and a reverse primer 5'−GAACTGAGGCC AGACAGGA-3' and exon 5 for both SP-A genes using a forward primer 5'−GAGGCCGAAGGCCAGAGCAG-3' and a reverse primer 5'−GAACTGAGGCC AGACAGGA-3' and for SP-A2 using a reverse primer 5'−GAACTGAGGCC AGACAGGA-3'.

### 2.4. Genotyping of SP-A genes

Genotyping was carried out as described earlier [3]. In brief, SP-A genes were amplified using gene-specific primers and conditions as described previously [7]. Genomic clones containing the SP-A1 gene and the SP-A2 gene were included as controls of the gene specificity of amplification in all PCR reactions. A PCR-cRFLP-based method was used to detect single nucleotide polymorphisms at codons 19, 50, 62, 133 and 219 for the SP-A1 gene and at codons 9, 91, 140 and 233 for the SP-A2 gene. Codon 85 was analysed for both genes to further ensure the gene specificity of the PCR amplification. Different combinations of polymorphisms at these sites distinguish between different alleles. At present, 19 different alleles have been described for the SP-A1 gene (denoted as 6A<sub>n</sub>) and 15 for SP-A2 (denoted as 1A<sub>n</sub>) [3]. The two SP-A genes are shown to be in marked linkage disequilibrium [7] and are thus suitable for haplotype analysis. The haplotypes for the SP-A1 and SP-A2 genes are denoted as 6A<sub>n</sub>/1A<sub>n</sub>.
2.5. Statistical analyses

Allele frequency comparisons were performed using chi-squared analysis. The allele distributions in the American and Finnish populations were compared using 2×k tables. The observed genotype frequencies were compared to the expected Hardy-Weinberg distribution using chi-square analysis. To ascertain the validity of the chi-square analysis, rare genotypes with expected cell counts of less than five were pooled together. The observed heterozygosity was a fraction of the heterozygotes in the study population. The expected heterozygosity was determined on the basis of the non-coding exons and introns of SP-A1 and SP-A2 genes in the American and Finnish populations were compared using chi-squared analysis. The allele distributions in the American and Finnish populations were compared using chi-square analysis. To ascertain the validity of the chi-square analysis, rare genotypes with expected cell counts of less than five were pooled together. The observed heterozygosity was a fraction of the heterozygotes in the study population. The expected heterozygosity was determined on the basis of the expected distribution of different alleles in a given population, assuming random mating. Polymorphism information content (PIC) was calculated according to the formula

\[
PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2p_j^2
\]

where \(p_i\) is the frequency of the \(i\)th allele, \(n\) is the total number of alleles and \(i \neq j\).

Haplotype frequencies were determined on the basis of homozygous genotypes. Utilising this information, haplotypes from heterozygous genotypes were determined according to the highest likelihood. The statistical significance of the linkage disequilibrium was calculated by comparing the expected haplotype frequencies (based on allele frequencies) to the observed ones using chi-square analysis.

3. Results

3.1. New SP-A alleles and differences compared to previously published SP-A sequences

To ensure that the alleles found in the American population were also present in the Finnish population, the coding exons of six individuals were sequenced. Variation was observed at seven codons within the coding exons of the SP-A1 gene (codons 19 GC/AG, 39 CAC/CT, 50 C/GTC, 62 CCAG/CT, 133 ACA/AC, 184 TAC/T and 219 C/TG) and at five codons within the coding exons of the SP-A2 gene (codons 9 AA/CC, 50 C/GTC, 91 C/GCT, 140 TCC/T and 223 A/ACG). The allelic variation seen at codons 39 and 184 for SP-A1 and codon 50 for SP-A2 has not been reported previously. Variations at codons 39 and 184 did not change the encoded amino acid, whereas variation at codon 50 caused a change of Leucine to Valine. Furthermore, in the SP-A2 gene codon 246 was constantly GAG, whereas it has been reported to be GAT [13]. The observed variation within the non-coding exons and introns is shown in Table 1.

### Table 1

| Location | Sequenced region | Variation | Sequenced region | Variation |
|----------|-----------------|-----------|-----------------|-----------|
| Exon 2   | Non-coding 5' end | No variation | Non-coding 5' end | No variation |
| Intron 2 | 1219–1298       | 1249 G (ins) | 1246–1325       | No variation |
| Intron 2 | 1502–1530       | No variation | 1522–1550       | No variation |
| Intron 3 | 1651–1704       | 1661 C/T   | 1671–1719       | 1682 G (ins) |
| Intron 3 | 2333–2403       | 2341 G (ins) | 2355–2424       | 2418 G/T |
| Intron 4 | 2940–2956       | No variation | 2962–2969       | No variation |
| Intron 4 | 2482–2540       | 2501C (ins) | 2530 C/T       | 1691 A/G (del) |
| Intron 4 | 2940–2956       | No variation | 2962–2969       | 2966 A/C |

The observed variability within non-coding exons and introns of SP-A1 and SP-A2 genes is the frequency of the \(i\)th allele, \(n\) is the total number of alleles and \(i \neq j\).
Table 2

| Allele frequency | aa 9* | aa 19b | aa 50c | aa 62d | aa 91e | aa 133f | aa 140g | aa 219h | aa 223i |
|------------------|-------|--------|--------|--------|--------|--------|--------|--------|--------|
| SP-A1 allele:    |       |        |        |        |        |        |        |        |        |
| 6A               | A     | C      | C      | G      | C      | G      | C      | C      | 0.04   |
| 6A²              | A     | T      | G      | A      | C      | A      | C      | C      | 0.60   |
| 6A³              | A     | T      | C      | G      | C      | A      | C      | T      | 0.28   |
| 6A⁴              | A     | T      | C      | G      | C      | A      | C      | T      | 0.09   |
| 6A⁵              | A     | T      | C      | A      | C      | A      | T      | C      | < 0.01 |
| SP-A2 allele:    |       |        |        |        |        |        |        |        |        |
| 1A               | C     | C      | G      | G      | C      | A      | C      | C      | 0.04   |
| 1A⁰              | A     | C      | G      | G      | G      | A      | C      | C      | 0.57   |
| 1A¹              | C     | C      | G      | G      | G      | A      | T      | C      | 0.16   |
| 1A²              | C     | C      | G      | G      | G      | A      | C      | C      | 0.13   |
| 1A³              | A     | C      | G      | G      | G      | A      | T      | C      | 0.03   |
| 1A⁴              | C     | C      | G      | G      | A      | T      | C      | C      | 0.07   |
| 1A⁵              | A     | C      | G      | G      | G      | A      | T      | C      | < 0.01 |
| 1A⁶              | C     | C      | G      | G      | C      | A      | T      | C      | < 0.01 |

*a* Asn or Thr; *b* Ala or Val; *c* Leu or Val; *d* no amino acid change; *e* Pro or Ala; *f* Arg or Trp; *g* Lys or Gln.

The two human SP-A genes are in close physical proximity to each other [10] and thus in strong linkage disequilibrium [7]. Therefore, the alleles at these two loci cosegregate as one unit (haplotype). Accordingly the SP-A1 and SP-A2 genes are suitable for haplotype analysis. We identified 13 different haplotypes in the studied Finnish population, namely 6A/1A, 6A²/1A⁰, 6A²/1A¹, 6A²/1A², 6A²/1A³, 6A³/1A⁰, 6A³/1A¹, 6A³/1A², 6A³/1A³, 6A³/1A⁵, 6A⁴/1A², 6A⁴/1A⁰, 6A⁴/1A⁵. The SP-A1 and SP-A2 genes were in strong linkage disequilibrium. This was evident as a deviation from random association of the SP-A1 and SP-A2 alleles: the observed frequencies of eight SP-A1/SP-A2 haplotypes were strikingly different (*p* < 0.0001) from the expected haplotype frequencies. The haplotype distributions of SP-A1 and SP-A2 genes are shown in Table 3. Identification of the SP-A1/SP-A2 haplotypes present in the population increases the informativeness of the SP-A locus as a genetic marker when studying the possible associations between the SP-A genotype and disease.

### 3.3. Evaluation of the genetic informativeness of the SP-A gene locus

The usefulness of a genetic marker varies according to its degree of informativeness. This is shown by the polymorphism information content (PIC) value of the marker and by its heterozygosity index. The expected heterozygosity index was 0.56 for SP-A1 and 0.62 for SP-A2, while the observed indices were 0.54 and 0.61, respectively. The PIC values were 0.50 for SP-A1 and 0.59 for SP-A2. These results indicate that both of these SP-A gene loci are highly polymorphic and can therefore be used in studies evaluating the importance of the SP-A gene locus in pulmonary and infectious diseases.

### 4. Discussion

Surfactant proteins are crucial for many aspects of surfactant biology and physiology [2]. Apart from being essential for the normal surfactant structure [15], the human SP-A protein has a clearly defined role in innate host defence [16,17]. Besides being expressed in type II alveolar cells, the SP-A gene is also expressed not only in peripheral airways but also in central airways, namely in the Eustachian tube [18]. Individual SP-A alleles have been shown to have different properties. One SP-A1 allele (6A²) has been associated with low mRNA levels [6], whereas 6A³ allele-derived mRNA was more prone to the destabilisation effect of glucocorticoid treatment than 6A² or 1A⁰ derived mRNA in a cell culture model [11]. Furthermore, specific SP-A1 and SP-A2 alleles (6A² and 1A⁰) have been suggested to be associated with respiratory distress syndrome in premature infants (RDS) [12, 19]. Deficiency in SP-A expression in mice is associated with enhanced susceptibility to specific pulmonary infections [2]. Accordingly, one can speculate that allelic variability in the SP-A gene locus plays a potential role in the genetic predisposition to a variety of other pulmonary and infectious diseases. In addi-
Table 3
The observed and expected SP-A1/SP-A2 haplotype distribution

| Haplotype       | N obs | Freq obs | N exp | Freq exp | P value | OR     | 99.9% CI         |
|-----------------|-------|----------|-------|----------|---------|--------|-----------------|
| 6A/1A           | 64    | 0.04     | 3     | 0.00     | < 0.0001| 22.2   | 3.2–155.6       |
| 6A/Xa           | 0     | 0.00     | 61    | 0.04     | < 0.0001|        |                 |
| 6A2/1A0         | 878   | 0.56     | 528   | 0.33     | < 0.0001| 2.5    | 2.0–3.2         |
| 6A2/1A1         | 2     | 0.00     | 151   | 0.10     | < 0.0001| 0.0    | 0.0–0.1         |
| 6A2/1A2         | 3     | 0.00     | 116   | 0.07     | < 0.0001| 0.0    | 0.0–0.2         |
| 6A2/1A3         | 41    | 0.03     | 24    | 0.02     | 0.033   | 1.7    | 0.7–4.1         |
| 6A2/Xa          | 0     | 0.00     | 105   | 0.07     | < 0.0001|        |                 |
| 6A3/1A0         | 25    | 0.02     | 259   | 0.16     | < 0.0001| 0.1    | 0.0–0.2         |
| 6A3/1A1         | 257   | 0.16     | 74    | 0.05     | < 0.0001| 4.0    | 2.5–6.2         |
| 6A3/1A2         | 162   | 0.10     | 57    | 0.04     | < 0.0001| 3.1    | 1.8–5.1         |
| 6A3/1A6         | 1     | 0.00     | 0     | 0.00     |         |        |                 |
| 6A3/1A10        | 8     | 0.01     | 2     | 0.00     | 0.057   | 4.0    | 0.3–54.3        |
| 6A3/Xa          | 0     | 0.00     | 60    | 0.04     | < 0.0001|        |                 |
| 6A4/1A2         | 33    | 0.02     | 17    | 0.01     | 0.023   | 2.0    | 0.7–5.3         |
| 6A4/1A5         | 103   | 0.07     | 9     | 0.01     | < 0.0001| 12.2   | 3.9–38.4        |
| 6A4/Xc          | 0     | 0.00     | 110   | 0.07     | < 0.0001|        |                 |
| 6A5/1A5         | 3     | 0.00     | 0     | 0.00     |         |        |                 |
| 6A5/Xc          | 0     | 0.00     | 3     | 0.00     |         |        |                 |
| Total           | 1580  | 1.00     | 1579  | 1.00     | < 0.0001|        |                 |

a Any other SP-A2 allele.

b Chi-squared analysis of the whole haplotype distribution: very low frequency haplotypes have been combined (14 df).

5. Conclusion

Based on our results, the SP-A gene locus is sufficiently polymorphic to be used as a marker in studies of the genetic component of pulmonary and infectious diseases in the Finnish population. The new single nucleotide polymorphism seen in the present study potentially increases the information value of the SP-A gene locus as a genetic marker.

Acknowledgments

The authors are grateful to Ms. Mirkka Parviainen and Ms. Maarat Hännikäinen for their excellent technical assistance. We are grateful to Dr Marja-Leena Pokela for the collection of blood specimens. The work was supported by the Finnish Academy (MH) and the Foundation for Paediatric Research in Finland (MR and MH).

References

[1] M.E. Avery and J. Mead, Surface properties in relation to atelectasis and hyaline membrane disease, *Am J Dis Child* 97 (1959), 517–523.
[2] E.C. Crouch, Collectins and pulmonary host defence, *Am J Respir Cell Mol Biol* 19 (1998), 177–210.
[3] S. DeAngelo, Z. Lin, G. Wang, S. Phillips, M. Ramet, J. Luo and J. Floros, Novel, non-radioactive, simple and multiplex PCR-RFLP methods for genotyping human SP-A and SP-D marker alleles, *Disease Markers* 15 (1999), 269–281.
[4] P. Eggleton and K.B.M. Reid, Lung surfactant proteins involved in innate immunity, *Curr Opin Immunol* 11 (1999), 28–33.
[5] J.H. Fisher, F.T. Kao, C. Jones, R.T. White, B.J. Benson and R.J. Mason, The coding sequence for the 32,000-dalton pulmonary surfactant-associated protein A is located on chromosome 10 and identifies two separate restriction-fragment-length polymorphisms, *Am J Hum Genet* 40 (1987), 503–511.
[6] J. Floros, D.S. Phelps, D. deMello, J. Longmate, H. Harding, B. Benson and T. White, The utility of postmortem lung for RNA studies: Variability and correlation of the expression of surfactant proteins in human lung, *Exp Lung Res* 17 (1991), 91–104.
[7] J. Floros, S. DeAngelo, M. Koptides, A.M. Karinch, P.K. Rogan, H. Nielsen, R.G. Spragg, K. Wattenberg and G. Deiter, 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.
[8] J. Floros and P. Kala, Surfactant proteins: molecular genetics of neonatal respiratory diseases, *Ann Rev Physiol* 60 (1998), 365–384.
[9] H.P. Haagsman, R.T. White, J. Schilling, K. Lau, B.J. Benson and J. Golden et al., Studies of the structure of lung surfactant protein SP-A, *Am J Physiol* 257 (1989), L421–429.
[10] R.R. Hoover and J. Floros, Organization of the human SP-A and SP-D loci at 10q22-q23: physical and radiation hybrid
Mapping reveals gene order and orientation, *Am J Respir Cell Mol Biol* **18** (1998), 353–362.

[11] R.R. Hoover and J. Floros, SP-A 3′-UTR is involved in the glucocorticoid inhibition of human SP-A gene expression, *Am J Physiol* **276** (1999), L917–924.

[12] P. Kala, T. Have, H. Nielsen, M. Dunn and J. Floros, Association of pulmonary surfactant protein A (SP-A) gene and respiratory distress syndrome: Interaction with SP-B, *Pediatric Res* **43** (1998), 169–177.

[13] S.L. Katyal, G. Singh and J. Locker, Characterization of a second human pulmonary surfactant-associated protein SP-A gene, *Am J Respir Cell Mol Biol* **6** (1992), 446–452.

[14] T.R. Korfhagen, R.W. Glasser, M.D. Bruno, M.J. McMahan and J.A. Whitsett, A portion of the human surfactant protein A (SP-A) gene locus consists of a pseudogene, *Am J Respir Cell Mol Biol* **4** (1991), 463–469.

[15] T.R. Korfhagen, M.D. Bruno, G.F. Ross, K.M. Huelsman, M. Ikegami, A.H. Jobe, S.E. Wert, B.R. Wert, B.R. Stripp, R.E. Mums, S.W. Glasser, C.J. Bachurski, H.S. Iwamoto and J.A. Whitsett, Altered surfactant function and structure in SP-A gene targeted mice, *Proc Natl Acad Sci USA* **93** (1996), 9594–9599.

[16] A.M. LeVine, M.D. Bruno, K.M. Huelsman, G.F. Ross, J.A. Whitsett and T.R. Korfhagen, Surfactant protein A-deficient mice are susceptible to group B streptococcal infection, *J Immunol* **158** (1997), 4336–4340.

[17] A.M. LeVine, J. Gwozdz, J. Stark, M.D. Bruno, J.A. Whitsett and T.R. Korfhagen, Surfactant protein-A enhances respiratory syncytial virus clearance in vivo, *J Clin Invest* **103** (1999), 1015–1021.

[18] R. Paananen, V. Glumoff and M. Hallman, Surfactant protein A and D expression in the porcine Eustachian tube, *FEBS Lett* **452** (1999), 141–144.

[19] M. Rämét, R. Haataja, R. Marttila, J. Floros and M. Hallman, Surfactant protein-A (SP-A) gene locus associated with respiratory distress syndrome in Finnish population, *Am J Hum Gen* **66** (2000), 1569–1579.

[20] R.T. White, D. Damm, J. Miller, K. Spratt, J. Schilling, S. Hawgood, B. Benson and B. Cordell, Isolation and characterization of the human pulmonary surfactant apoprotein gene, *Nature* **317** (1985), 361–363.