

ARTICLE

Surfactant protein C gene variation in the Finnish population – association with perinatal respiratory disease

Meri Lahti¹, Riitta Marttila² and Mikko Hallman^{*1}

¹Department of Pediatrics and Biocenter Oulu, University of Oulu, Oulu, Finland; ²Seinäjäki Central Hospital, Seinäjoki, Finland

Surfactant protein C (SP-C) is a small hydrophobic protein component of alveolar surfactant, a lipid–protein complex lining the alveolar surface of the lung. Surfactant deficiency is the main cause of respiratory distress syndrome (RDS) in premature infants. RDS is a major risk factor of a chronic lung disease called bronchopulmonary dysplasia (BPD). The dominant mutations of the SP-C gene have recently been associated with interstitial lung diseases. However, the common genetic variation in the surfactant protein C gene has not been studied in detail. In the present study, the exonic variation of the SP-C gene in the Finnish population ($n = 472$) was defined, and the association of the allelic variants with the susceptibility to RDS and BPD was examined. Conformation-sensitive gel electrophoresis (CSGE) was used to determine the extent of exonic variation in the SP-C gene. Methods of genotyping were generated for three biallelic polymorphisms of the SP-C gene's exons 1, 4 and 5, which encode proSP-C. The frequencies of these polymorphisms were evaluated in a study population consisting of 158 DNA samples from full-term infants. In addition, the linkage disequilibrium between the SP-C alleles was evaluated by haplotype analysis of parent–infant triplets. The role of SP-C gene variation in RDS and in BPD was evaluated in a high-risk population of 245 premature infants. According to the present results, the SP-C polymorphisms were associated with RDS and with very premature birth. The strength of allelic associations differed according to the gender of the premature infants.

European Journal of Human Genetics (2004) 12, 312–320. doi:10.1038/sj.ejhg.5201137

Published online 21 January 2004

Keywords: lung surfactant; genetic polymorphism; respiratory distress syndrome (RDS); bronchopulmonary dysplasia (BPD); allelic association; conformation-sensitive gel electrophoresis (CSGE)

Introduction

Surfactant protein C (SP-C) is a component of pulmonary alveolar surfactant, a complex lipid–protein mixture lining the alveolar surface of the lung. Deficiency of surfactant due to a lack of differentiation of the complex leads to respiratory distress syndrome (RDS), which is the major

cause of mortality and morbidity among premature infants.¹ Many premature infants with RDS develop a chronic lung disease called bronchopulmonary dysplasia (BPD). Some very premature infants, who do not initially have RDS, may develop BPD, suggesting that factors other than the differentiation of alveolar epithelial cells play a role in the pathogenesis.²

SP-C is expressed as a 179 amino-acid proprotein. Mature SP-C is an extremely nonpolar peptide that contains two thioester-linked palmitoyl groups. The amino acids 9–34 of mature SP-C, which are mostly valines and leucines, form a stable α -helix capable of spanning the membrane

*Correspondence: Dr M Hallman, Department of Pediatrics and Biocenter Oulu, PO Box 5000, Oulu FIN-90014, Finland. Tel: +358 8 3155100; Fax: +358 8 3155559; E-mail: mikko.hallman@oulu.fi
Received 21 January 2003; revised 9 October 2003; accepted 30 October 2003

bilayer.³ Unlike other surfactant proteins, SP-C has been considered to be lung-specific. The cell type-specific expression of SP-C and the conservation of its primary structure between species implies specific functions of importance.⁴ However, the functional roles of SP-C are still incompletely understood. Mice lacking SP-C have apparently no serious lung disease, but show abnormalities in surfactant stability and decreased viscoelasticity of the lung. It has been suggested that SP-C is important in stabilizing the phospholipid film, especially at reduced lung volumes.⁵

In addition to SP-C, surfactant protein B (SP-B) is a hydrophobic protein involved in reducing surface tension at the air-liquid interface of alveoli. SP-B has additional roles in maintaining the integrity of alveolar epithelial cells. Absence of SP-B expression results in defective intracellular processing of the surfactant and secretion of a complex that contains an excess of proSP-C and no mature peptide.^{6,7} On the other hand, surfactant protein A (SP-A) and surfactant protein D (SP-D) are hydrophilic C-type lectins involved in the surfactant metabolism and innate immunity of the lung.⁸

The genes encoding surfactant proteins are polymorphic, as they contain several variable sites in their coding and noncoding DNA sequences. The SP-A, -B and -D genes have been studied as candidates for multifactorial lung diseases. Certain SP-A alleles and an SNP in the SP-B gene have been shown to associate interactively with RDS in premature infants.⁹

The SP-C gene is short, spanning only 3500 base pairs on the short arm of chromosome 8 (8p23.1) near the gene coding for bone morphogenetic protein-1 (BMP-1).¹⁰⁻¹³ The extent of allelic variation and the distribution of different SP-C alleles in the general population have not been examined in detail. The first report of an association between a mutation in the SP-C gene and a disease was published by Noguee *et al.*¹⁴ They demonstrated a single base substitution in an interstitial lung disease patient's SP-C gene that resulted in a preterm stop codon causing a large deletion in the carboxyl-terminal domain of the patient's SP-C precursor protein. Recently, Noguee¹⁵ reported several new SP-C gene mutations in 11 patients with chronic lung disease of unknown etiology.

Detection of single-nucleotide polymorphisms (SNPs) in the candidate genes of complex multifactorial diseases is a useful tool in molecular genetics of complex traits. SNPs occur frequently in the genome, and they can be used as markers of disease-causing mutations or other alterations in DNA. SNPs often have functional consequences, especially when they are located in the coding or regulatory regions of a gene.¹⁶ Among the most widely used techniques of screening for single-base changes are single-strand conformation polymorphism and denaturing gradient gel electrophoresis, which are based on the screening of a PCR-amplified DNA fragment for the

presence of mutations or nucleotide mismatches. Conformation-sensitive gel electrophoresis (CSGE) has proved to be a simple and accurate method for detecting single-nucleotide changes even in repetitive and GC-rich templates.¹⁷ Under optimal conditions, the method is highly sensitive and specific.¹⁸

The aim of the present study was to examine the extent of common genetic variation in the SP-C gene in a homogenous Caucasian population of Finnish origin, and to investigate the significance of SP-C gene variation in a high-risk population consisting of very premature infants. The exonic regions of the SP-C gene were analyzed for polymorphisms, and methods for genotyping three biallelic polymorphisms in the coding regions of the SP-C gene were generated. SP-C allele and genotype frequencies were determined from a population of consecutively born healthy newborn infants, and the linkage disequilibrium between the SP-C alleles was evaluated by haplotype analysis. Here, we report for the first time evidence linking the SP-C gene with perinatal disease.

Subjects and methods

Study population

The study population of 245 high-risk newborns consisted of infants born before 34 weeks of gestation (term 37-42 weeks) in Oulu University Hospital, Seinäjoki Central Hospital and Tampere University Hospital. Umbilical cord blood samples ($n=116$) were collected prospectively during the years 1997-2001, and buccal smear samples ($n=129$) were obtained from the retrospectively recruited premature infants born during 1987-1996. Umbilical cord blood specimens were additionally obtained from 158 healthy term infants born consecutively in Oulu University Hospital during 2 months in 1998. All infants were of Finnish origin. In addition, buccal smear samples available from the parents of the premature infants were collected retrospectively ($n=69$). Informed consent was obtained from the parents of the infants, and the study protocol was approved by the ethical committees of the participating centres.

The maternal and neonatal medical histories were evaluated for clinical data concerning gestational age, antenatal and postnatal glucocorticoid treatment, gender, birth order, etiology of preterm birth, mode of delivery and severity of the respiratory disorders of the premature infant. The diagnosis of RDS was made on the basis of the published clinical, radiographic and/or pathologic criteria.¹⁹ The diagnostic criterion of BPD used in the study was the requirement for supplemental oxygen at the postmenstrual age of 36.0 weeks.²

Table 1 presents the full-term infants ($n=158$) and the characteristics of the premature infants ($n=245$). There were altogether 183 cases with RDS (106 males, 77 females). The incidence of RDS among the premature

Table 1 Clinical characteristic of the premature and the full-term infants

	Full-term infants, gestation 37–42 weeks	Premature infants
Gestational age, days (mean±SD)	276±9	200±15
Birth weight, g (mean±SD)	3480±470	1160±350
Male sex, <i>n</i> (%)	75 (47%)	136 (55%)
Female sex, <i>n</i> (%)	83 (53%)	109 (45%)
Bronchopulmonary dysplasia, <i>n</i> (%)	0	75 (31%)
Respiratory distress syndrome, <i>n</i> (%)	0	183 (75%)
Total <i>n</i>	158	245

Table 2 Primers used in CSGE analysis of the SP-C gene

Amplified SP-C exon	Primer orientation	Primer sequence 5'–3'	Primer location	Product length (bp)	Annealing temperature (°C)
1	Forward	TTATCTGGGCTTCGGTTCTG	5'Flanking region	295	58
	Reverse	AAACAGGCCAGGGAGGATAG	Intron 1		
2	Forward	GGGAGGGTGTTCAGCTTGT	Intron 1	334	58
	Reverse	CTCCCTTGGACAGTTTCCT	Intron 2		
3	Forward	AGCCAGAAACACACGGAGATG	Exon 2	732	58
	Reverse	GGCTTGATAGCGATCAGCAG	Exon 4		
4	Forward	AGAGGGAGGTGGCTAAGGAC	Intron 3	264	58
	Reverse	AATAGCCACCATTCCGGACAC	Intron 4		
5	Forward	TCTTTACTGATGAGAAAAGTACG	Intron 4	441	62
	Reverse	TCAGGGGAGTGGGAAGTACC	Intron 5		
6 (3'UT)	Forward	TCATCCAGGCAACTCGGGGGA	Intron 5	430	65
	Reverse	TGCCAGCGCCCTGGGACAGA	Intron 6		

infants was 74.7%. Altogether, 75 of the premature infants developed BPD (45 males and 30 females). Of them, seven had no RDS at birth.

DNA sample preparation

Genomic DNA was extracted from frozen EDTA-anticoagulated blood samples, using the Puregene DNA Isolation kit (Gentra Systems, Minneapolis, MN, USA). DNA was then diluted to 50 ng/μl. Genomic DNA was extracted from buccal smears using Chelex 100 medium.

CSGE analysis

A total of 90 DNA samples from full-term infants were analyzed with CSGE. Six pairs of primers were designed to specifically amplify the six exons of the SP-C gene (Table 2), using the sequence J03890 (NCBI nucleotide database)¹¹ as a reference sequence. The PCR for each CSGE reaction was performed in a 20 μl reaction volume containing 100 ng of genomic DNA, 0.06 mM dNTPs (Pharmacia, Peapack, NJ, USA), 0.2 μM of each primer, 1 × HotPrime incubation buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl), 1.5 mM MgCl₂ and 1.75 U of Hot Prime TM 'Hot start' Polymerase (Qbiogene, France). The amplifying conditions were as follows: an initial denaturation at 95°C for 15 min, followed by 37 cycles of 30 s at 95°C, 1 min at a primer-specific temperature (58–65°C), 1 min at 72°C and final

extension at 72°C for 7 min, and a 10°C dwell cycle (Thermal cycler PTC-200; MJ Research, Waltham, MA, USA). For amplification of the fragment containing exon 3, the denaturation step of the PCR program was 1 min, and the elongation step 2 min, using 0.5 mM dNTPs and 7.5 mM MgCl₂.

Heteroduplexes were generated by incubating the PCR products at 98°C for 5 min, followed by 30 min at 68°C and a 10°C dwell cycle. A suitable amount of the PCR product, usually 10 μl, was taken for CSGE analysis. CSGE was essentially performed as described.²⁰

Nucleotide sequencing

The observed heteroduplexes were identified by direct sequencing (ABI PRISM™ 377 Sequencer and DYEnamic ET Terminator Cycle Sequencing Kit, Amersham Pharmacia Biotech), or cloning (pGEM[®]-T Easy Vector System, Promega Corporation, Madison, USA) and sequencing. If several different heteroduplexes were observed for any analyzed fragment, at least two of each differently migrating heteroduplexes were sequenced. Also, several homoduplexes were sequenced from each CSGE analysis. The sequences were then compared to the published genomic SP-C gene sequences J03890¹¹ and U02948²¹ (NCBI nucleotide database).

Table 3 Genotyping of SP-C polymorphisms in exons 1, 4 and 5

Polymorphism	Location	Forward primer (5'–3') and reverse primer (5'–3')	Restriction enzyme	Resulting fragments
P14P	Exon 1	TTATCTGGGCTTCGGTTCTG AAACAGGCCAGGGAGGATAG	<i>MspI</i>	CCG (Pro) 209 bp+86 bp or CCA (Pro) 295 bp
N138T	Exon 4	CTAGCTGCTGATCGCTAC AATAGCCACCATTCGGACAC	<i>BfaI</i>	AAT (Asn) 189 bp or ACT (Thr) 97 bp+92 bp
N186S	Exon 5	TCTTTACTGATGAGAAAAGTAC TCAGGGGAGTGGGAAAGTACC	<i>BspI286I</i>	AAC (Asn) 297 bp+144 bp or AGC (Ser) 228 bp+144 bp+69 bp

Genotyping of SP-C alleles

For the genotyping of polymorphisms in exons 1 (P14P), 4 (N138T) and 5 (N186S), genotyping methods based on PCR and RFLP were generated. The PCR for each genotyping was performed in a 10 μ l reaction volume containing 50 ng of genomic DNA, 0.06 mM dNTPs (Pharmacia, Peapack, NJ, USA), 0.2 μ M each primer, 1 \times HotPrime incubation buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl), 1.5 mM MgCl₂ and 0.9 U of Hot Prime TM 'Hot start' Polymerase (Qbiogene, France). The PCR cycles consisted of initial denaturation at 95°C for 15 min, followed by 37 cycles of 95°C for 30 s, 1 min at a primer-specific temperature, 72°C for 1 min and a final extension at 72°C for 7 min (Thermal cycler PTC-200; MJ Research, Waltham, MA, USA). An aliquot of the PCR product was digested with a specific enzyme, and the resulting fragments were visualized on 2.5% agarose gel. A summary of the genotyping procedure is presented in Table 3. Altogether, 472 DNA samples were genotyped, including 158 samples from full-term infants, 245 from premature infants and 69 samples from the parents of the premature infants.

Statistical analysis and haplotype analysis between N138T and N186S polymorphisms

The allele and genotype frequencies were calculated using SPSS for Windows versions 9 or 11 (SPSS Inc., Chicago, IL, USA). Comparisons of the allele and genotype frequencies were performed by the χ^2 test (two-sided) using 2 \times 2 or 2 \times 3 contingency tables, respectively. The observed genotype frequencies were tested for Hardy–Weinberg equilibrium by χ^2 analysis. Linkage disequilibrium between the N138T and N186S polymorphisms was determined by haplotype analysis, by genotyping the 69 DNA samples available from the parents of the premature infants. The significance of linkage disequilibrium was determined using χ^2 analysis by comparing the observed and expected haplotype distributions. The expected haplotype distribution was calculated on the basis of the observed allele frequencies. The strength of linkage disequilibrium was determined by calculating the Lewontin's D' .²²

Logistic regression analysis (SPSS for Windows version 11) was used to investigate whether SP-C alleles explained the risk of RDS in the presence of gender as a confounding environmental factor. Variables in the analyses were

included using forced entry, and the presence or absence of RDS was set as the dependent variable.

Results

Polymorphism of the SP-C gene

The primers for CSGE were designed to amplify the six exons of the SP-C gene, each PCR fragment containing one exon and at least 70 base pairs of the 5' and 3' flanking sequences of the exon. In the analysis with CSGE, heteroduplexes were observed frequently in all amplified fragments, except in the fragment containing exon 1, where only two homologous heteroduplexes were detected among the 90 samples analyzed. To define the nucleotide changes causing the heteroduplexes, several samples of each fragment were cloned and sequenced after analysis with CSGE. The sequences were then compared to the published genomic human SP-C sequences. Here, the numbering of the nucleotides refers to the sequence J03890.¹¹

A new variation, a heterozygous substitution of A for G (g.657G>A), was detected in the last nucleotide of exon 1 in two of the 90 samples analyzed with CSGE. This substitution did not change the amino-acid composition, since both codons CCG and CCA encode for Proline. The frequency of the CCA variant was found to be very low among the Finnish population (Tables 5–7). No variation was found in exon 2, which encodes the mature SP-C protein. Sequencing revealed only one polymorphic site in intron 1, 21 base pairs upstream from the exon 2 start site (g.1335C/T), which caused the heteroduplexes observed in CSGE analysis. The codon for amino acid 45 (22 in mature SP-C) was CTC (Leu) in all sequenced samples, which differs from the sequence U02948.²¹ Exon 3 was also found to be conserved, since no polymorphisms were detected in the coding sequence. However, several homozygous changes were detected in intron 3 (g.1993–1994insC, g.2006–2007insC, g.2009–2010insG). Since the fragment of exon 3 was rather large and the sensitivity of CSGE could therefore be decreased, multiple samples were cloned and sequenced to ensure full analysis of this fragment.

An AAT/ACT (g.2294A/C) polymorphism that resulted in p.N138T variation was detected in exon 4. We found both of these alleles to occur frequently in the Finnish



Figure 1 Genetic polymorphisms in the exonic regions of the SP-C gene. The black boxes indicate exons and the gray boxes indicate untranslated exons.

population (Tables 5–7). Also, a TG-deletion polymorphism was observed in intron 4 (g.2328–2329delTG). An AAC/AGC (g.2772A/G) polymorphism, which results in p.N186S variation, was detected in exon 5. Both of these alleles occur frequently in the Finnish population (Tables 5–7). In addition, a G/C polymorphism was detected in intron 4, eight base pairs upstream from the first nucleotide of exon 5 (g.2643G/C). A single base deletion polymorphism leading to *EcoRI*-RFLP²³ was also present in intron 5 (g.2987delA). Exon 6, where several variable sites have been detected, is entirely untranslated. We found polymorphisms in the nucleotides g.3130A/G, g.3180T/C and g.3181A/G. Figure 1 presents the exonic polymorphisms that were detected in the SP-C gene.

Linkage disequilibrium between N138T and N186S polymorphisms

The allele coding for 138 Thr was found to exist more often than expected with the allele 186 Ser, and the allele coding for 138 Asn with the allele 186 Asn, suggesting linkage disequilibrium between these loci. The strength of this linkage disequilibrium was estimated by haplotype analysis, by genotyping the DNA samples available from the parents of 36 premature infants. Altogether, 69 DNA samples of the parents were genotyped, and 70 haplotypes could be confirmed. The results are presented in Table 4. The linkage disequilibrium was very strong (Lewontin's $D' = 0.973$) between these loci.

Association analysis of SP-C alleles with RDS and BPD among premature infants

The frequencies of polymorphisms in the exons 1 (P14P), 4 (N138T) and 5 (N186S) were determined in the population of premature infants with RDS ($n = 183$) and their premature control infants ($n = 62$) (Table 5). The frequencies of the alleles coding for 138 Asn and 186 Asn among the infants with RDS were 0.32 and 0.35, while the frequencies among the infants without RDS were 0.23 and 0.25 ($P = 0.071$ and 0.040, respectively). Among the infants without RDS and with BPD ($n = 7$), the 138 Asn and 186 Asn frequencies were 0.14 and 0.21, respectively.

Among the very premature infants with gestational age of less than 30 weeks, the frequencies of 138 Asn and 186 Asn were 0.33 and 0.35 for those with RDS ($n = 144$) and 0.20 and 0.24 for the control infants ($n = 27$) ($P = 0.059$ and 0.095, respectively).

Table 4 Linkage disequilibrium (LD) between the N138T and N186S polymorphisms

SP-C haplotype	Observed frequency %	Expected frequency %	D'
138 Asn–186 Asn	0.371	0.148	0.973
138 Asn–186 Ser	0.000	0.222	
138 Thr–186 Ser	0.600	0.378	
138 Thr–186 Asn	0.029	0.252	

χ^2 -test (3 d.f.): $P < 0.0001$.

Table 5 SP-C allele and genotype frequencies (%) among the premature infants with RDS and their premature controls

	RDS ($n = 183$)		Controls ($n = 62$)
SP-C allele	14 CCG/Pro	98.9	100.0
	14 CCA/Pro	1.1	0
	138 AAT/Asn	32.0	23.4
	138 ACT/Thr	68.0	76.6
	186 AAC/Asn ^a	35.0	25.0
	186 AGC/Ser	65.0	75.0
SP-C genotype	14 CCG/CCG	98.4	100
	14 CCG/CCA	1.1	0
	14 CCA/CCA	0.5	0
	138 AAT/AAT	9.8	8.1
	138 AAT/ACT	44.3	30.6
	138 ACT/ACT	45.9	61.3
	186 AAC/AAC	12.0	6.5
	186 AAC/AGC	45.9	37.0
	186 AGC/AGC	42.1	56.5

^a $P = 0.04$.

When the infants with RDS and the premature control infants were further divided according to the gender, the association of alleles 138 Asn and 186 Asn with RDS was observed to be stronger among male infants ($P = 0.018$ and 0.045, respectively). Logistic regression analysis was used to test whether alleles coding for 138 Asn or 186 Asn explained the risk of RDS when gender was included in the analysis as a confounding factor. According to this analysis, both alleles were independent risk factors for RDS (for allele 138 Asn: $P = 0.022$, OR 2.01, 95% CI 1.10–3.66 and for allele 186 Asn: $P = 0.036$, OR 1.87, 95% CI 1.04–3.38), as well as haplotype 138 Asn–186 Asn ($P = 0.020$, OR 2.05, 95% CI 1.12–3.74).

The association of SP-C polymorphism with BPD was tested. The study population included 75 infants with BPD and 170 controls. The control group was categorized based on the diagnosis of RDS, and the two groups, controls with RDS and controls without RDS, were independently compared with the group of infants with BPD (Table 6). There were no significant differences in the SP-C allele or genotype frequencies between the infants with BPD and the controls with RDS. When the controls without RDS ($n = 55$) were compared with the infants with BPD, moderate differences between the allele frequencies

Table 6 SP-C allele and genotype frequencies (%) among the premature infants with BPD and the controls with or without RDS

		BPD (n = 75)	Controls with RDS (n = 115)	Controls without RDS (n = 55)
SP-C allele	14 CCG/Pro	100.0	98.3	100.0
	14 CCA/Pro	0	1.7	0
	138 AAT/Asn	32.7	30.4	24.5
	138 ACT/Thr	67.3	69.6	75.5
	186 AAC/Asn	34.7	34.3	25.5
	186 AGC/Ser	65.3	65.7	74.5
SP-C genotype	14 CCG/CCG	100.0	97.4	100.0
	14 CCG/CCA	0	1.7	0
	14 CCA/CCA	0	0.9	0
	138 AAT/AAT	10.7	8.7	9.1
	138 AAT/ACT	44.0	43.5	30.9
	138 ACT/ACT	45.3	47.8	60.0
	186 AAC/AAC	10.7	12.2	7.3
	186 AAC/AGC	48.0	44.3	36.4
	186 AGC/AGC	41.3	43.5	56.4

emerged. The frequency of 138 Asn was 0.25 and the frequency of 186 Asn 0.26 among the controls without RDS, compared to 0.33 and 0.35 among the infants with BPD ($P=0.154$ and 0.112 , respectively). There were no differences in associations between SP-C alleles and BPD among male or female infants.

Low birth weight and very preterm birth are important risk factors for BPD and RDS. We separately compared the SP-C allele frequencies between the BPD infants and their controls among the extremely low-weight infants (birth weight <1000 g, 43 BPD and 44 controls) and the infants with birth weight ≥ 1000 g (32 BPD and 125 controls). There were no significant differences in the SP-C allele frequencies between these groups. The frequencies of the alleles coding for 14 Pro (CCG), 138 Asn and 186 Asn among the extremely low-weight infants were (BPD vs controls) 1.0 vs 0.96, 0.28 vs 0.31 and 0.32 vs 0.32, respectively. Among the infants with birth weight ≥ 1000 g, the frequencies were 1.0 vs 1.0, 0.27 vs 0.28 and 0.30 vs 0.31. We also compared the SP-C allele frequencies between the BPD infants and their controls separately in terms of the degree of prematurity. The SP-C allele frequencies in the infants with BPD and the controls were determined separately for the infants born at gestation <28 weeks (41 BPD and 46 controls) and those born at gestation ≥ 28 weeks (34 BPD and 124 controls). There were no significant differences in the SP-C allele frequencies between these groups. The frequencies of the alleles coding for 14 Pro (CCG), 138 Asn and 186 Asn among the extremely premature (<28 weeks) infants were (BPD vs controls) 1.0 vs 0.97, 0.38 vs 0.32 and 0.39 vs 0.33, respectively. Among the premature infants with gestation ≥ 28 weeks at birth, the frequencies were 1.0 vs 1.0, 0.27 vs 0.27 and 0.29 vs 0.31.

SP-C allele and genotype frequencies according to the degree of prematurity

The SP-C allele and genotype frequencies were determined in relation to the degree of prematurity (Table 7). When the SP-C allele and genotype frequencies were compared between the group of extremely preterm infants (gestational age <28 weeks, $n=87$) and the group of healthy full-term infants (gestational age 37–42 weeks, $n=158$), significant differences were observed in the frequencies of alleles coding for amino acid 138 ($P=0.046$). Further analysis revealed an association only among the female infants. When the subgroups of full-term female infants ($n=83$) and extremely preterm female infants with gestational age less than 28 weeks ($n=42$) were compared, the allele frequencies for 138 Asn and 186 Asn emerged higher in the group of preterm female infants ($P=0.012$ and 0.042 , respectively). Also, the frequency of the haplotype 138 Asn–186 Asn was increased in this group ($P=0.012$).

All the populations studied were in Hardy–Weinberg equilibrium, as tested by χ^2 analysis.

Discussion

In the present study, the common genetic variation in the SP-C gene was defined using conformation-sensitive gel electrophoresis, a highly specific method for screening single-base changes in double-stranded DNA fragments. A comparison of the CSGE and genotyping results of 90 DNA samples analyzed with both of these methods showed the sensitivity of CSGE to be 90% or more. A new variation was described, and a previously established variation was found to be associated with RDS, a serious neonatal disease caused by lung immaturity. Unlike the other surfactant proteins

Table 7 SP-C allele and genotype frequencies (%) according to the degree of prematurity

		Term birth, 37–42 weeks of gestation, n = 158	Premature birth, 28–34 weeks of gestation, n = 158	Premature birth, <28 weeks of gestation, n = 87
SP-C allele	14 CCG/Pro	99.4	99.7	98.3
	14 CCA/Pro	0.6	0.3	1.7
	138 AAT/Asn	25.9	27.2	34.5 ^a
	138 ACT/Thr	74.1	72.8	65.5
	186 AAC/Asn	30.7	30.7	35.6
	186 AGC/Ser	69.3	69.3	64.4
SP-C genotype	14 CCG/CCG	98.7	99.4	97.7
	14 CCG/CCA	1.3	0.6	1.1
	14 CCA/CCA	0	0	1.1
	138 AAT/AAT	6.3	8.2	11.5
	138 AAT/ACT	39.9	38.0	46.0
	138 ACT/ACT	53.8	53.8	42.5
	186 AAC/AAC	9.5	10.1	11.5
	186 AAC/AGC	44.3	41.1	48.3
	186 AGC/AGC	46.2	48.7	40.2

^a $P = 0.05$, when compared to full-term infants.

SP-A and SP-B, an SP-C variant was not only associated with RDS, but also with premature birth.

A new variation, heterozygous substitution of A for G (g.657G>A), was detected in the last nucleotide of exon 1. Since this substitution occurred at the exon–intron boundary, where conserved nucleotides are presumed to play an important role, it may thus have an effect on RNA splicing. However, the frequency of the variant CCA was found to be low, and therefore should be considered rather a rare variant than a common polymorphism. N138T and N186S variations were detected in exons 4 and 5, respectively.²⁴ These two biallelic polymorphisms were frequent in the Finnish population, and are potentially useful tools in association studies. Excluding the differences listed above, the intronic regions seem to be in accordance with the sequence U02948 rather than the sequence J03890. We did not detect any of the mutations recently described by Noguee,¹⁵ which confirms that they are not common polymorphisms, but mutations associated with interstitial lung diseases of varying etiologies.

Both of the polymorphic amino acids 138 and 186 are located in the C-terminal part of the proSP-C molecule. The C-terminal region has been shown to be crucial for the targeting and processing of proSP-C that is eventually secreted as SP-C.^{25,26} Furthermore, most of the mutations described in the SP-C gene have been located in this region. Interestingly, two different single-base substitutions associated with lung disease have been detected in the codon for amino acid 188,^{15,27} only two residues apart from amino acid 186, where the Asn/Ser polymorphism is present.

Antenatal glucocorticoid and surfactant therapy at birth diminish the incidence and severity of RDS considerably, and hence the number of neonatal deaths. Both therapies have a clear-cut influence on the endogenous surfactant system. Interestingly, neither of these synergistic therapies

diminishes the incidence of BPD, which is a heterogeneous disease associated with prematurity and RDS. There are a number of pathogenic factors that potentially influence the susceptibility of BPD. Very low birth-weight infants, despite minimal or no respiratory distress shortly after birth, may develop BPD.^{28,29} Dominant SP-C gene alterations may play a role in the etiology of BPD. The mutations in the SP-C gene encoding proSP-C were associated with interstitial lung disease often manifesting in infancy.^{14,15} In addition, Thomas *et al*²⁷ reported a single-base SP-C gene mutation (188L>Q) that segregated with the pulmonary fibrosis phenotype in one kindred. However, we did not detect any direct association between the common SP-C alleles and BPD, as studied in 75 cases and 170 premature controls. In the present study, seven of the 75 BPD cases did not have RDS at birth. In these seven cases, the frequencies of 138 Asn and 186 Asn (0.14 and 0.21, respectively) tended to be lower than in RDS.

Considering the role of SP-C in surfactant function, we propose that the SP-C alleles influence the susceptibility to RDS in premature infants. A significant difference was observed in the allele frequency of 186 Asn ($P = 0.040$) between the infants with RDS ($n = 183$) and the gestation controls ($n = 62$), and a trend was observed in the allele frequency of 138 Asn ($P = 0.071$). When the groups of very premature infants (gestational age less than 30 weeks) were compared, the differences in the 138 Asn and 186 Asn allele frequencies between the RDS cases and the controls appeared distinct, although the sample size of the controls was small ($n = 27$) and the results were not statistically significant (P -values 0.059 and 0.095).

Apart from genetic factors, RDS is strongly influenced by environmental factors, most notably the length of gestation. In the present study, a trend towards gestational dependence of SP-C alleles was found. Among the group of extremely premature infants with gestational age of less

than 28 weeks ($n=87$), the frequency of 138 Asn was increased compared to the infants with gestational age of 28–32 weeks ($n=150$) or the full-term infants ($n=158$, $P=0.046$).

Since RDS and BPD are highly complex diseases with multiple environmental and genetic factors affecting the susceptibility, the impact of one candidate gene on the outcome is not likely to be dominating. Association studies with multiple comparisons may in some occasions result in Type I error. However, statistical correction methods can easily decrease the power of the study, especially when the effect of the gene examined is known to be limited. Therefore, we want to point out that the P -values from the significance tests in this study were not corrected for multiple comparisons, and therefore, the possibility for Type I error can not be ruled out.

Almost all extremely premature infants have RDS (90% in the present series) that is a consequence rather than a cause of premature birth. The association between the SP-C alleles, premature birth and RDS could have arisen by change. However, these observations raise the possibility that the differentiation of the surfactant system required for the neonatal respiratory adaptation and the labor process share some regulatory pathways. SP-C and the other surfactant components are secreted into the amniotic fluid. Their concentration in the amniotic cavity increases as a function of the length of gestation, reducing the risk of RDS in an unborn fetus. The hydrophobic surfactant proteins (SP-C and SP-B), when added to cultured amnion cells, stimulate the production of prostaglandins required in the labor process.³⁰ The present finding associating very premature birth with the SP-C allele 138 Asn, that potentially affects the secretion of SP-C from immature lung, is consistent with foetal involvement in the labor process.³¹ The present observation that the allelic association with RDS and very premature births was influenced by the foetal sex is appealing, since the risk of premature birth and the incidence of RDS among premature infants are both influenced by the foetal gender. The expression levels of SP-C in the foetal lung reveal the differences between males and females.³² The possibility that a foetal factor required for neonatal adaptation is additionally involved in the onset of the labor process remains to be substantiated by further genetic and experimental studies.

Acknowledgements

We thank Ms Maarit Hännikäinen for excellent technical assistance. The study was supported by grants from the Biocenter Oulu, the Finnish Academy and Sigrid Juselius Foundation.

References

- 1 Jobe AH: Pulmonary surfactant therapy. *N Engl J Med* 1993; **328**: 861–868.
- 2 Jobe AH, Bancalari E: Bronchopulmonary dysplasia. *Am J Respir Crit Care Med* 2001; **163**: 1723–1729.
- 3 Johansson J, Curstedt T, Robertson B: The proteins of the surfactant system. *Eur Respir J* 1994; **7**: 372–391.
- 4 Johansson J: Structure and properties of surfactant protein C. *Biochim Biophys Acta* 1998; **1408**: 161–172.
- 5 Glasser SW, Burhans MS, Korfhagen TR *et al*: Altered stability of pulmonary surfactant in SP-C-deficient mice. *Proc Natl Acad Sci USA* 2001; **98**: 6366–6371.
- 6 Akinbi HT, Breslin JS, Ikegami M *et al*: Rescue of SP-B knockout mice with a truncated SP-B proprotein. Function of the C-terminal propeptide. *J Biol Chem* 272, 1997; 9640–9647.
- 7 Pryhuber GS: Regulation and function of pulmonary surfactant protein B. *Mol Genet Metab* 1998; **64**: 217–228.
- 8 Crouch E, Hartshorn K, Ofek I: Collectins and pulmonary innate immunity. *Immunol Rev* 2000; **173**: 52–65.
- 9 Haataja R, Ramet M, Marttila R, Hallman M: Surfactant proteins A and B as interactive genetic determinants of neonatal respiratory distress syndrome. *Hum Mol Genet* 2000; **9**: 2751–2760.
- 10 Glasser SW, Korfhagen TR, Weaver TE *et al*: cDNA, deduced polypeptide structure and chromosomal assignment of human pulmonary surfactant proteolipid, SPL(pVal). *J Biol Chem* 1988; **263**: 9–12.
- 11 Glasser SW, Korfhagen TR, Perme CM, Pilot-Matias TJ, Kister SE, Whitsett JA: Two SP-C genes encoding human pulmonary surfactant proteolipid. *J Biol Chem* 1988; **263**: 10326–10331.
- 12 Fisher JH, Emrie PA, Drabkin HA *et al*: The gene encoding the hydrophobic surfactant protein SP-C is located on 8p and identifies an *EcoRI* RFLP. *Am J Hum Genet* 1988; **43**: 436–441.
- 13 Wood S, Yaremko ML, Schertzer M, Kelemen PR, Minna J, Westbrook CA: Mapping of the pulmonary surfactant SP5 (SFTP2) locus to 8p21 and characterization of a microsatellite repeat marker that shows frequent loss of heterozygosity in human carcinomas. *Genomics* 1994; **24**: 597–600.
- 14 Nogue LM, Dunbar III AE, Wert SE, Askin F, Hamvas A, Whitsett JA: A mutation in the surfactant protein C gene associated with familial interstitial lung disease. *N Engl J Med* 2001; **344**: 573–579.
- 15 Nogue LM, Dunbar III AE, Wert S, Askin F, Hamvas A, Whitsett JA: Mutations in the surfactant protein C gene associated with interstitial lung disease. *Chest* 2002; **121**: 20S–21S.
- 16 Gray IC, Campbell DA, Spurr NK: Single nucleotide polymorphisms as tools in human genetics. *Hum Mol Genet* 2000; **9**: 2403–2408.
- 17 Ganguly A, Rock MJ, Prockop DJ: Conformation-sensitive gel electrophoresis for rapid detection of single-base differences in double-stranded PCR products and DNA fragments: evidence for solvent-induced bends in DNA heteroduplexes. *Proc Natl Acad Sci USA* 1993; **90**: 10325–10329.
- 18 Korkko J, Kaitila I, Lonnqvist L, Peltonen L, Ala-Kokko L: Sensitivity of conformation sensitive gel electrophoresis in detecting mutations in Marfan syndrome and related conditions. *J Med Genet* 2002; **39**: 34–41.
- 19 Ramet M, Haataja R, Marttila R, Floros J, Hallman M: Association between the surfactant protein A (SP-A) gene locus and respiratory-distress syndrome in the Finnish population. *Am J Hum Genet* 2000; **66**: 1569–1579.
- 20 Korkko J, Annunen S, Pihlajamaa T, Prockop DJ, Ala-Kokko L: Conformation sensitive gel electrophoresis for simple and accurate detection of mutations: comparison with denaturing gradient gel electrophoresis and nucleotide sequencing. *Proc Natl Acad Sci USA* 1998; **95**: 1681–1685.
- 21 Hatzis D, Deiter G, deMello DE, Floros J: Human surfactant protein-C: genetic homogeneity and expression in RDS; comparison with other species. *Exp Lung Res* 1994; **20**: 57–72.
- 22 Lewontin RC: The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 1964; **49**: 49–67.
- 23 Takahara K, Lee S, Wood S, Greenspan DS: Structural organization and genetic localization of the human bone morphogenetic protein 1/mammalian tolloid gene. *Genomics* 1995; **29**: 9–15.

- 24 Warr RG, Hawgood S, Buckley DI *et al*: Low molecular weight human pulmonary surfactant protein (SP5): isolation, characterization, and cDNA and amino acid sequences. *Proc Natl Acad Sci USA* 1987; **84**: 7915–7919.
- 25 Keller A, Steinhilber W, Schafer KP, Voss T: The C-terminal domain of the pulmonary surfactant protein C precursor contains signals for intracellular targeting. *Am J Respir Cell Mol Biol* 1992; **6**: 601–608.
- 26 Beers MF, Lomax CA, Russo SJ: Synthetic processing of surfactant protein C by alveolar epithelial cells. The COOH terminus of proSP-C is required for post-translational targeting and proteolysis. *J Biol Chem* 1998; **273**: 15287–15293.
- 27 Thomas AQ, Lane K, Phillips III J *et al*: Heterozygosity for a surfactant protein C gene mutation associated with usual interstitial pneumonitis and cellular nonspecific interstitial pneumonitis in one kindred. *Am J Respir Crit Care Med* 2002; **165**: 1322–1328.
- 28 Watterberg KL, Demers LM, Scott SM, Murphy S: Chorioamnionitis and early lung inflammation in infants in whom bronchopulmonary dysplasia develops. *Pediatrics* 1996; **97**: 210–215.
- 29 Charafeddine L, D'Angio CT, Phelps DL: Atypical chronic lung disease patterns in neonates. *Pediatrics* 1999; **103**: 759–765.
- 30 Lopez BA, Newman GE, Phizackerley PJ, Turnbull AC: Effect of lipid and protein fractions from fetal pulmonary surfactant on prostaglandin E production by a human amnion cell line. *Eicosanoids* 1989; **2**: 29–32.
- 31 Challis JR, Smith SK: Fetal endocrine signals and preterm labor. *Biol Neonate* 2001; **79**: 163–167.
- 32 Dammann CE, Ramadurai SM, McCants DD, Pham LD, Nielsen HC: Androgen regulation of signaling pathways in late fetal mouse lung development. *Endocrinology* 2000; **141**: 2923–2929.