Dinucleotide repeats in the human surfactant protein-B gene and respiratory-distress syndrome

Joanna FLOROS,*§ Stavroula V. VELETZA,* Padmasree KOTIKALAPUDI,* Livia KRIZKOVA,* Anne M. KARINCH,* Charles FRIEDMAN,† Susie BUCHTER‡ and Keith MARKS*

*Department of Cellular and Molecular Physiology, Pennsylvania State University College of Medicine, Hershey, PA 17033, †Methodist Medical Center, Jackson, MS, and ‡Emory University School of Medicine, Atlanta, GA, U.S.A.

Pulmonary surfactant, a lipoprotein complex, is essential for normal lung function, and deficiency of surfactant can result in respiratory-distress syndrome (RDS) in the prematurely born infant. Some studies have pointed towards a genetic contribution to the aetiology of RDS. Because the surfactant protein B (SP-B) is important for optimal surfactant function and because it is involved in the pathogenesis of pulmonary disease, we investigated the genetic variability of the SP-B gene in individuals with and without RDS. We identified a 2.5 kb *Bam*HI polymorphism and studied its location, nature and frequency. We localized this polymorphism in the first half of intron 4 and found that it is

derived by gain or loss in the number of copies of a motif that consists of two elements, a 20 bp conserved sequence and a variable number of CA dinucleotides. Variability in the number of motifs resulting from either deletion (in 55.3 % of the cases with the variation) or insertion (44.7 %) of motifs was observed in genomic DNAs from unrelated individuals. Analysis of 219 genomic DNAs from infants with ($n = 82$) and without ($n = 137$) RDS showed that this insertion/deletion appears with significantly higher frequency in the RDS population (29.3 as against 16.8 %, $P < 0.05$).

INTRODUCTION

Pulmonary surfactant, a lipoprotein complex, is essential for normal lung function. Surfactant progresses through several morphological forms before reaching its functionally active form at the air/liquid interface. Within alveolar type II cells, lipid and protein components of surfactant are stored in lamellar bodies. The lamellar bodies are secreted into the hypophase (the liquid covering the alveolar surface) where they are transformed into tubular myelin. Tubular myelin is thought to be an obligatory intermediate of the active form of surfactant, the surface monolayer that is found at the air/liquid interface of the alveolus.

The surfactant proteins (SPs) SP-A, SP-B, and SP-C play important roles in the structure, function and metabolism of surfactant. SP-A and SP-B are essential for *in vitro* formation of tubular myelin [1,2], and the importance of SP-B for normal lung function has been confirmed by animal studies. When a monoclonal antibody to SP-B was instilled in the lungs of neonatal rabbits, inflammation, lesions that included hyaline membranes [a pathological finding in respiratory-distress syndrome (RDS)] and a decrease in compliance were observed [3]. *In vivo* administration of reconstituted surfactant containing SP-B to pre-term rabbits was shown to restore function to levels comparable with that of natural sheep-derived surfactant, with some dependence on the characteristics of the accompanying mechanical ventilation [4]. *In vitro* studies have shown that reconstituted surfactant preparations containing SP-B exhibit better surface-tension-reducing properties [5] and are more resistant to fibrinogen inactivation [6] than preparations that do not contain SP-B. Moreover, the pivotal role of the hydrophobic surfactant proteins SP-B and SP-C in normal lung function is shown in studies where

surfactants containing these proteins produce a faster and greater improvement in lung function than do preparations that do not include these proteins [7–9].

Deficiency of surfactant can result in RDS in the prematurely born infant [10]. Several factors, including gender, race and gestational age, appear to contribute to the development of RDS [11–13]. In addition, a variety of studies have strongly hinted at a genetic component in the development of the disease in at least some cases. Thus the aetiology of RDS is considered multifactorial and/or multigenic. In 1965, Graven and Mesenheimer [14] reported that mothers who delivered low-birth-weight (LBW) babies could be divided into two groups: one group with LBW babies who rarely develop RDS and another group with LBW infants having a high incidence of RDS. Because the groups did not differ in the progress of the pregnancy, labour or delivery, the authors suggested that an unknown genetic factor was involved in the development of the disease. In a more recent study, the risk for RDS was shown to be higher for the babies of women with a previously affected pre-term infant (B. A. Nagourney, personal communication). A study of twins showed a high concordance of RDS among monozygotic twins, but not among dizygotic twins [15]. Other studies [16] also have suggested a genetic contribution to the aetiology of RDS. The well-documented effect of race (prematurely born African-American infants are less likely to develop RDS than their weight-matched Caucasian counterparts) [13,17–19] may also in some cases reflect the impact of genotype [20,21] upon susceptibility to the condition.

Although it has been known for decades that surfactant deficiency can result in RDS in the prematurely born infant, it is only recently that the surfactant proteins have been implicated in

Abbreviations used: SP, surfactant protein; RDS, respiratory-distress syndrome; RFLP, restriction-fragment-length polymorphism; LBW, low birth weight; FIGE, field-inversion gel electrophoresis.

§ To whom correspondence should be addressed.

the aetiology of RDS or other pulmonary diseases. Variation in their expression [22,23] has been associated with RDS, and mutation in the coding sequence of SP-B has been associated with congenital pulmonary alveolar proteinosis [24].

The human SP-B gene has been localized on the short arm of chromosome 2 [25,26], and its genomic and cDNA sequences have been characterized [27,28]. The SP-B gene consists of 11 exons: 10 encode the precursor SP-B molecule, and exon 11 contains sequences for the 3-untranslated region. Repetitive elements such as *Alu* sequences and dinucleotide (CA) repeats are also present within the SP-B gene [28,29]. Its expression is both developmentally and hormonally regulated. Because of the importance of SP-B in the surface-tension-lowering properties of surfactant and because of its potential involvement in the pathogenesis of pulmonary disease [22–24], we investigated the genetic variability of the SP-B gene. In the present study we have identified and characterized a variant locus within intron 4 of the SP-B gene. Analysis of 219 genomic DNAs from infants with ($n = 82$) and without ($n = 137$) RDS showed that insertion/deletion genetic variants within SP-B intron 4 appear with significantly higher frequency in the RDS group than in infants without RDS.

MATERIALS AND METHODS

Restriction-fragment-length-polymorphism (RFLP) analysis

Blood specimens were collected from infants with and without RDS. Diagnosis of RDS was made by clinical criteria (grunting, retraction and flaring) and verified by X-ray analysis (reticulogranular pattern). The work was carried out in accordance with the Declaration of Helsinki of the World Medical Association and was approved by the ethical committee of the Pennsylvania State University College of Medicine. Genomic DNA was prepared from the blood specimens. Genomic DNA (10 µg) was then digested with restriction endonuclease *Bam*HI in a 300 µl reaction mixture for 12–16 h at 37 °C. The reaction mixture contained 1 × *Bam*HI buffer provided by New England Biolabs, Beverly, MA, U.S.A., 10 mM β-mercaptoethanol, 14 units of *Bam*HI enzyme/µg of DNA and 2 mM spermidine. Spermidine was added when the buffer and DNA were at 37 °C. The reaction was stopped with the addition of EDTA (final concn. 10 mM), and the DNA was ethanol-precipitated and resuspended in de-ionized water. The digested DNA products were separated using field-inversion gel electrophoresis (FIGE). The gel cast was 20 cm × 25 cm, and the agarose concentration was 0.9% in 0.5 × TBE buffer (1 × TBE is 90 mM Tris/borate/2 mM EDTA), containing 0.5 mg/ml ethidium bromide. Electrophoresis was performed at room temperature in precooled 20 °C buffer. The buffer circulator maintained uniform temperature during the run. Voltage was set at 200 V, and pulsing was performed using standard programs of the Programmable Power Inverter-200 purchased from PPI-200, MJ Research, Watertown, MA, U.S.A. Simple linear ramps with constant forward-to-reverse ratios were used to improve linearity of migration of DNA molecules over the entire range of FIGE. Following electrophoresis, the DNA was transferred under vacuum in VacuGene XL, supplied by Pharmacia Biotech., Piscataway, NJ, U.S.A., on to a nylon membrane (Genescreen or Oncor, supplied by du Pont-NEN, Boston, MA, U.S.A., and Oncor, Gaithersburg, MD, U.S.A., respectively). The DNA was u.v.-cross-linked to the membrane and the filter was hybridized to ³²P-labelled human SP-B cDNA [27]. Prehybridization (3–4 h) and hybridization (20–24 h) took place at 42 °C in hybridization solution I, supplied by Oncor, containing 5 × Denhardt's solution and 100 ng/ml salmon sperm DNA. The washes were at room temperature, 20 min each, in 1 × SSC (150 mM NaCl/15 mM sodium citrate, pH 7.0) and 0.5% SDS

(repeated three times) and one wash at 45 °C for 30 min in 0.5 × SSC and 0.5% SDS. The filter was then exposed to Kodak XAR film using two intensifying screens at –70 °C for 24 h–7 days (see Figure 1 below).

PCR

PCR was conducted in 100 µl final volume of 1 × *Taq* polymerase buffer supplied by Promega Corp., Madison, WI, U.S.A. [50 mM KCl/10 mM Tris/HCl (pH 9)/1% Triton X-100], 1.5 mM MgCl₂, 0.2 mM of each dNTP and 7.5 pmol of each of the primers. For ³²P-labelled PCR-generated probes (see Figure 2 below) using the SP-B cDNA as a template, we used 10 ng of cDNA and the following PCR conditions. Samples were heated at 95 °C for 10 min and, following addition of 2.5 units of *Taq* polymerase, they went through 30 cycles of 94 °C, 60 s; 55 °C, 60 s; 72 °C, 60 s for primers nos. 94 and 171. For primers nos. 172 and 102 the samples went through 30 cycles of 94 °C, 60 s; 65 °C, 60 s; 72 °C, 60 s. Primers 94 and 171 were used to generate probe B (see Figure 2 below), and primers 172 and 102 were used to generate probe C (see Figure 2 below).

94: GAGGTGCCATGGCTGAGTCA, sense, positions 1046–1065, Genbank HUMSPBAA

102: CCAGCTGAGCTTTCAGCAGA, antisense, positions 10276–10295

171: GTCGTCAAGCACTTGGTTGCA, antisense, positions 2561–2581

172: CTGGTCATCGACTACTTCCA, sense, positions 2591–2610

For the (CA)_n study, 220 ng of genomic DNA was used in each reaction. Samples were heated to 98 °C for 6 min and, following addition of 2.5 units of *Taq* polymerase, they went through 25 cycles of 94 °C, 30 s; 57 °C, 60 s; 72 °C, 90 s, followed by 5 min extension at 72 °C in a Thermal Cycler from Perkin-Elmer, Norwalk, CT, U.S.A. We used primers 172 (see above) and 161 (TGTGTGTGAGAGTGAGGGTGTAAG, antisense, positions 3174–3197) for the (CA)_n study and 172 and 189 (CATACAGATGCCGTTTGAGTC, antisense, positions 3447–3467) for subcloning and sequencing. For amplification of ³²P-labelled products, 3 pmol of unlabelled primer 172 were substituted with an equal amount of the same ³²P-labelled primer (sp. radioactivity 1.25 × 10⁸ c.p.m./µg). Labelling was carried out with T₄ DNA kinase from New England Biolabs. PCR products were first screened on 1.5%-agarose gels composed of two parts of agarose from Gibco BRL, Gaithersburg, MD, U.S.A., and one part of Nusieve agarose from FMC BioProducts, Rockland, ME, U.S.A., and then analysed on acrylamide gels (7 M urea/4% acrylamide/1 × TBE). The gels were vacuum-dried on 3MM Whatman paper and exposed to Kodak XAR5 film.

Subcloning and sequencing

The PCR-generated intron-4 products (primers 172 and 189) were size-fractionated by electrophoresis on a 3% Nusieve agarose gel from FMC BioProducts. Individual DNA fragments, both invariant and variant (similar to those shown in Figure 3 below), were cut out, phenol-extracted, ether-extracted, ethanol-precipitated and subsequently ligated to a pT7 blue vector from Novagen, Madison, WI, U.S.A., carrying T 3' overhangs, and used to transform Novablue competent cells. Double-stranded templates were prepared by alkaline lysis, and both strands were sequenced with the fmol sequencing kit from Promega Corp. and ³⁵S-dATP, and separated on 7 M urea/6% acrylamide gels in

1 × TBE, run at 1600 V. The sequencing gels were soaked in 5 % acetic acid/5 % methanol to remove the urea, vacuum-dried on 3MM Whatman paper and exposed to Kodak XAR5 film.

RESULTS

RFLP and localization of a *Bam*HI polymorphism

To study the genetic variability of the human SP-B locus in RDS and control individuals, we used RFLP analysis. Two classes of variable *Bam*HI fragments, with sizes of about 6.6 and 2.5 kb, were identified consistently under the experimental conditions used (Figure 1). Initial RFLP analysis showed that polymorphisms in the 6.6 kb class of fragments did not differ between RDS and control populations, and therefore the 6.6 kb polymorphisms were not studied further. The focus of the present study is the 2.5 kb class of fragments that appeared with different frequency in the populations studied. A 2.5 kb fragment is widely represented in the population (Figure 1, lanes 1–9), and we will refer to it as the 'invariant' fragment. Variably sized fragments were also present in some of the genomic DNAs. There were two classes of variant fragments: either slightly smaller than the 2.5 kb fragment (lane 4) or slightly larger (lanes 5 and 6). The variant fragments appeared with higher frequency in DNAs from individuals with RDS than from controls (results not shown).

To localize precisely the site of variability in the SP-B gene, we prepared ³²P-labelled PCR-generated truncated SP-B cDNA probes (A, B and C; see the bottom of Figure 2). These probes were used in genomic Southern plots, and the results are depicted, respectively, in Figures 2(a), 2(b) and 2(c). The truncated probe B (Figure 2b) hybridizes only to the 2.5 kb variant sized fragments, but not to the 6.6 kb fragment. This finding suggests that the 2.5 kb fragment spans the 5' region of the SP-B gene that includes exons 1–5 and a portion of exon 6. The truncated probe C hybridizes (Figure 2c) with the 2.5 kb group of fragments and

Figure 1 *Bam*HI RFLP

Genomic DNAs were digested with *Bam*HI. The digests were size-fractionated by FIGE, transferred on to a nylon membrane, and the blot was hybridized with ³²P-labelled human SP-B cDNA. All lanes show two classes of fragments, one of about 6.6 kb and the other of about 2.5 kb. Lanes 1–9 show a band of about 2.5 kb that we refer to as the 'invariant' band. Lanes 4–6 depict, in addition to the 2.5 kb band, a variant form of this band. Lane 4 shows a smaller band, and lanes 5 and 6 show a larger band than the 2.5 kb one.

Figure 2 Localization of the polymorphism to the first half of the SP-B gene

Genomic DNAs were digested with *Bam*HI as in Figure 1. Lanes 1–3 in (c) contain the same samples as lanes 2–4 in (a) and (b). Blots in (a) were hybridized with the entire human SP-B cDNA probe. Blots in (b) and (c) were hybridized with probes B and C respectively, which were generated by PCR using SP-B cDNA as template. The genomic organization of the SP-B gene [28] is shown at the bottom of this Figure. The boxes (I–X) denote all coding exons, and the hatched box denotes the 3' untranslated region. The *Bam*HI recognition sites are marked. A scale in kb is shown below the SP-B gene.

the 6.6 kb because it spans the *Bam*HI recognition site within exon 6 [28].

In subsequent experiments we amplified different DNA regions of the 5' half of the SP-B gene and analysed the PCR products by agarose-gel electrophoresis. With this series of experiments, the genetic variability was narrowed down to an interval including positions 2591–3467. Further PCR analysis with primers 172 and 161 mapped the variable fragment to the first half of intron 4. Using this 'targetted' PCR approach (primers 172/161) the invariant DNA fragment assumes a size of about 600 bp and the variant allele may be either smaller or larger than the invariant allele. The PCR products were first analysed by electrophoresis on high-resolution Nusieve agarose gels. An example is shown in Figure 3(a). Genomic DNAs in lanes 1–3 contain only the invariant fragment, and the remaining lanes, in addition to the invariant fragment, contain either a smaller (lanes 4 and 5) or a larger (lanes 6–8) band. To assess more accurately the size difference between the variant and invariant fragments, we analysed the 172/161-primer ³²P-labelled PCR products on polyacrylamide sequencing gels. An example is shown in Figure 3(b). This enhanced resolution revealed a range of sizes for the variant band, both larger and smaller than the invariant band. The allele sizes were estimated by comparison with a sequencing ladder (see below).

Frequency of the variant fragment in genomic DNAs from control and RDS individuals

The frequency of the variant (both small and large) fragment was

Figure 3 Variability of Intron 4

A segment of the SP-B intron 4 was amplified using primers 172 and 161, and an aliquot was analysed by agarose-gel electrophoresis (a). Lanes 1–3 depict only the invariant band, lanes 4–5 show an additional smaller band, and lanes 6–8 show an additional larger band. Lane M depicts DNA markers of a ϕ X174-RF *Hae*III fragments. (b) Depicts 32 P-labelled PCR products of the same samples shown in (a), but run on a polyacrylamide gel for enhanced resolution.

Table 1 Population characteristics

Characteristic	Control	RDS
Gender: male/female	1.4	1.3
Race: Caucasian/African American	2.8	2.9
Mean gestational age (weeks)	37.3	31.6

determined in 219 DNA samples. Of these specimens, 82 were obtained from individuals with RDS and 137 were obtained from individuals without RDS. The characteristics of both groups are shown in Table 1. The ratio of male to female and Caucasians to African Americans is similar in the control and the RDS groups. The mean gestational age was 37.3 weeks in the control group and 31.6 weeks in the RDS group. The frequency of the variant fragment was significantly higher in the RDS group compared with the control group (Figure 4). The variant fragment was present in 29.3 % of the RDS and in 16.8 % of the control group ($P < 0.05$).

These data suggest that the intron-4 variability associates with RDS. Alternatively, because the gestational age of the RDS group was lower than that in the control group, this variability may correlate with premature birth, which may or may not lead to RDS. To minimize the effect of gestational age and attempt to determine whether the polymorphism increases the incidence of RDS independent of gestational age, we further analysed specimens of control ($n = 21$) and RDS ($n = 59$) groups with gestational age of ≤ 34 weeks. The difference in the mean gestational age between the two groups was approx. 2 weeks: mean gestational age in the RDS group was 29.8 weeks and in the

Figure 4 Frequency of polymorphism in control and RDS populations

The cross-hatched and the hatched bars depict the frequency of a variant band in control and RDS groups respectively. A χ -square analysis was performed ($P < 0.05$).

control group was 32 weeks. The frequency of the variant fragment was still significantly higher in the RDS group (32.2 %) than in the control group (9.5 %) ($P = 0.048$, Fisher's exact test).

Of the samples with the variant fragment, all were heterozygous for a variant and the invariant fragment, except for three samples: one was homozygous for a larger fragment and two were homozygous for a smaller fragment. Furthermore, among the samples with the variant fragment, 55.3 % carried a smaller, and 44.7 % carried a larger, fragment. Specimens from both African Americans and Caucasians were included in this analysis. When the data were analysed as a function of race, a trend appeared in which, among those individuals who had a variant fragment, African Americans were more likely to have a larger variant (in about 80 % of the cases), whereas a smaller variant was more frequent ($\sim 80\%$) in Caucasians, suggesting that some race-related differences exist.

Meiotic stability

To determine the meiotic stability of this intron-4 variability (that is, its ability to pass unaltered from one generation to the next), we used the 'targetted' approach (primers 172/161) and PAGE to study DNAs derived from a two-generation family (Figure 5c). For comparison, the same samples were analysed by RFLP analysis (Figure 5b). Individual 4 (Figure 5a) was born prematurely and suffered from RDS. This individual had both the invariant and a variant fragment, as depicted in Figure 5(b) (RFLP approach) and Figure 5(c) (targetted PCR approach). She inherited the invariant fragment from her mother (lane 2) and the variant one from her father (lane 1). Her siblings, individuals 3 and 5 (Figure 5a), were not born prematurely and did not suffer from RDS. Individual 3 is homozygous for the invariant 600 bp fragment. Individual 5 (lane 5) is a heterozygote and inherited the invariant fragment from his father (lane 1) and a larger variant from his mother (lane 2). These findings suggest that variability within intron 4 is meiotically stable, at least in this family, because both variant fragments, lanes 1 and 4 (smaller) and lanes 2 and 5 (larger), were transmitted without any change in size. The stability of one of the variant fragments was also confirmed by DNA sequence analysis. The sequence of the smaller variant fragments of individuals 1 and 4 was found to be identical. Sequence data for individual 1 are depicted in Figure 6 (below).

Figure 5 Meiotic stability of variant band

(b) Depicts a genomic Southern blot (as described in Figure 1) of a two-generation family (a). The daughter in lane 4 was born prematurely (< 30 weeks) and had RDS. This individual has a smaller variant band which she inherited from her father (lane 1, lower arrow). Her siblings in lanes 3 and 5 were not born prematurely, did not have RDS, and lack the smaller variant band. (c) Depicts the 172/161 PCR products from the same family as in (a) following electrophoresis on a 7 M urea/6% acrylamide gel. This higher-resolution approach revealed that the mother (lane 2) was not homozygous for the invariant band (upper arrow), as suggested by the RFLP analysis (b). She contains both the invariant band (upper arrow) and a larger band, the latter being transmitted to her son (lane 5). The upper arrow indicates the invariant fragments in each panel.

Characterization of the intron-4 variability

To study the nature of this variability, we cloned and sequenced PCR products (like those shown in Figure 3) generated with primers 172 and 189 using seven genomic DNAs as template. The 172/189-primer segments included two invariant fragments: one variant with a larger size and four variants with the smaller size. Sequence data from the invariant fragment are shown in Figure 6(b), and data for the larger and one of the smaller variants are shown in Figures 6(c) and 6(a) respectively. The broken lines shown in Figure 6(b) bracket the region of the invariant fragment that is missing from the smaller variant in Figure 6(a). The point at which the two broken lines meet marks the break-points on Figure 6(a) where a deletion (162 bp) appears to have occurred. The continuous lines on Figure 6(c) bracket the inserted segment that appears to account for the larger variant. The point at which the two continuous lines meet marks the position on Figure 6(b) where the insertion appears to have occurred.

It has been previously reported that the first half of intron 4 consists of repeated motifs [28]. Each motif consists of two elements: a 20 bp segment and a variable number of (CA)_n repeats. On the basis of sequence-analysis data for the seven segments (see above), we found that specific single base differences (underlined in Figure 7a) within each 20 bp element identify each 20 bp element as an element specific to the first or the second motif, and so on. The 11 motifs are depicted in the order of their appearance in Figure 7(a). Motif 10 appears to have a considerably truncated 20 bp element.

Three of the smaller variants (the sequence of one is shown in Figure 6a) lacked five motifs, 5–9 (Figure 7a). The break-points of this deletion are: at the 5' end, after the (CA)_n of motif 4, and at the 3' end, within the (CA)_n of motif 9 (marked with continuous lines in Figure 7a). The fourth smaller variant lacked eight motifs, 2–9 (Figure 7a). In this case, although the break-point at

Figure 6 Sequence data for the non-variant and variant DNA segments

(b) Depicts the sequence of the invariant band of individual 1 of Figure 5. (a) Depicts the smaller variant of the same individual, and (c) depicts the larger variant of another individual. The deleted region is identified with broken lines and the inserted region with continuous lines. The point where the broken lines meet in (a) and the point where the continuous lines meet in (b) identify the break-points where the deletion and insertion respectively occurred. Loading order: GATC.

the 5' end differed from the variants lacking five motifs, the break-point at the 3' end was the same. For the larger variant the break-point for the insertion (Figure 7b), the sequence of which is shown in Figure 6(c), is again within the (CA)_n of motif 9 that was found to be the stable (same) break-point for all four deletion variants. These data suggest that the break-points of deletions and insertions within the variant intron-4 locus occur within the (CA)_n element of the motif and not the 20 bp repeat segment. In the samples studied here, the (CA)_n element of motif 9 is involved in both insertion and deletion events.

The insertion found in the larger variant involves a number of degenerate repetitive motifs; the first portion of the insertion includes a long stretch of CA repeats, followed by a series of motifs (Figure 7b). The last four (CA)_n elements of the inserted motifs are identical with those in motifs 5–8, whereas the 20 bp elements of the inserted motifs are degenerate versions of the corresponding 20 bp elements of motifs 4–9. The 3' end of each 20 bp inserted element appears to be conserved and therefore recognizable as an element of a particular motif, but the 5' end of each 20 bp element is degenerate, and for the most part the degeneracy appears to be due to truncation. This larger variant lacked motif 4.

The size of the deleted or inserted region differed among individuals. The size range of these insertions and deletions as well as their distribution was determined by PCR and PAGE. The number of inserted or deleted motifs in samples where a

variant fragment was present was then approximated using a sequencing ladder as a marker. The approximation of inserted motifs was based on the average motif size of 25–30 bp. This

develop RDS ($n = 23$). However, prematurity itself is a risk factor for RDS, and it is possible that the polymorphism is a marker of prematurity and does not directly affect the development of RDS. In an attempt to dissociate the polymorphism and prematurity we restricted the analysis to specimens from infants with gestational age of ≤ 34 weeks. The difference in the mean gestational age between the two groups was then reduced to approx. 2 weeks (compared with approx. 6 weeks when the entire data set was analysed), and the frequency of the variant fragment was still significantly higher in the RDS group. This finding suggested that the polymorphism is associated with development of RDS, at least in part, independent of prematurity. It is also possible that the variant locus may itself not have any impact on SP-B expression, but may be linked to some other gene that contributes to the aetiology of RDS, or that intron-4 variability contributes to RDS in combination with other gene products. At present, we cannot distinguish among these possible alternatives. Therefore further studies are required to determine definitively whether the polymorphism described here is a marker of RDS and/or prematurity or whether the polymorphism and prematurity co-vary and both contribute to development of RDS.

Some of the variants (alleles) of the marker locus examined here were found more frequently in Caucasians (i.e., the smaller fragments) and others were found more frequently in African Americans (i.e., the larger fragments), suggesting a race difference in this SP-B polymorphism. Differences in the levels of SP-A and SP-B in amniotic fluids between Caucasians and African Americans have been reported [38]. Moreover, several epidemiological studies [13,17,18] and one biochemical study [19] have previously suggested a race difference in the risk for RDS. Although our present findings support the hypothesis that race factors are involved in the aetiology of RDS, additional studies are warranted to understand fully the implications (if any) of race.

The results of the present study suggest that genetic loci in addition to the SP-B intron-4 locus contribute to RDS because approx. 70% of the prematurely born individuals with RDS lacked the (CA)_n variant fragment. This possibility is consistent with the hypothesis that the aetiology of RDS is multigenic/multifactorial and requires further investigation. Studies of twins and of families with recurrence of RDS could elucidate the genetic basis of these outcomes. Development and use of additional genetic markers could also enhance future studies in this regard.

It is becoming evident that the genetic homogeneity or heterogeneity of a disease cannot be determined *a priori* on the basis of the homogeneity or heterogeneity of its clinical presentation. Diseases previously thought to be multifactorial and polygenic (e.g., osteoporosis) can have a single gene defect that predisposes towards their expression. In the case of osteoporosis, a defect in the 3' untranslated region of the vitamin-D-receptor gene (which is key for normal absorption of Ca^{2+} from the gut) correlates with reduced bone density in about three-quarters of the tested individuals [39]. Preliminary evidence suggests that the molecular mechanism involves alteration of the vitamin-D-receptor mRNA level, which, in turn, affects bone density via imbalance of Ca^{2+} homeostasis. On the other hand, several phenotypes may result from mutations in one gene. For example, mutations in the RET proto-oncogene have been associated with four syndromes, namely Hirschsprung's disease [40,41], multiple endocrine neoplasia and sporadic medullary thyroid carcinomas [42] and defects in the kidney and enteric nervous system of mice [43]. To what extent RDS is genetically homogeneous or heterogeneous is not known.

In summary, we have identified and characterized a variant locus within the SP-B gene that correlates with RDS. Further studies are necessary to determine whether this locus itself contributes to the pathogenesis of RDS or is linked to a gene that does.

We thank Dr. D. S. Phelps, Dr. P. K. Rogan and Dr. T. E. Weaver for their contributions, Gina Deiter and Susan DiAngelo for excellent technical assistance, the OB-GYN staff at the Brigham and Women's Hospital for their help with sample collection and Beth Ditzler for typing this manuscript. This work was supported by National Institutes of Health grant NIH HL34788.

REFERENCES

- 1 Suzuki, Y., Fujita, Y. and Kogishi, K. (1989) *Am. Res. Respir. Dis.* **140**, 75–81
- 2 Williams, M. C., Hawgood, S. and Hamilton, R. L. (1991) *Am. J. Respir. Cell Mol. Biol.* **5**, 41–50
- 3 Robertson, B., Kobayashi, T., Ganzuka, M., Grossmann, G., Li, W. Z. and Suzuki, Y. (1991) *Pediatr. Res.* **30**, 239–243
- 4 Rider, E. D., Ikegami, M., Whitsett, J. A., Hull, W., Absolom, D. and Jobe, A. H. (1993) *Am. Rev. Respir. Dis.* **147**, 669–676
- 5 Mathialagan, N. and Possmayer, F. (1990) *Biochim. Biophys. Acta* **1045**, 121–127
- 6 Seeger, W., Günther, A. and Thede, C. (1992) *Am. J. Physiol. (Lung Cell. Mol. Physiol.)* **5** **261**, L286–L291
- 7 Häfner, D., Kilian, U., Bühler, R., Beume, R. and Habel, R. (1993) *Pulmon. Pharmacol.* **6**, 15–25
- 8 Hall, S. B., Venkitaraman, A. R., Whitsett, J. A., Holm, B. A. and Notter, R. H. (1992) *Am. Rev. Respir. Dis.* **145**, 24–30
- 9 van Golde, L. M. G., Batenburg, J. J. and Robertson, B. (1994) *News Physiol. Sci.* **9**, 13–20
- 10 Avery, M. E. and Mead, J. (1959) *Am. J. Dis. Child.* **97**, 517–523
- 11 Usher, R. H., Allen, A. C. and McLean, F. H. (1971) *Am. J. Obstet. Gynecol.* **111**, 826–832
- 12 Farrell, P. M. and Avery, M. E. (1975) *Am. Rev. Respir. Dis.* **111**, 657–688
- 13 Richardson, D. K. and Torday, J. S. (1994) *Am. J. Obstet. Gynecol.* **170**, 1273–1278
- 14 Graven, S. N. and Mesenheimer, H. R. (1965) *Am. J. Child Dis.* **109**, 489–494
- 15 Myrlandhopoulos, N. C., Churchill, J. A. and Baszynski, A. J. (1971) *Acta Genet. Med. Gemellol.* **20**, 199–204
- 16 Lankenau, H. M. (1976) *Eur. J. Pediatr.* **123**, 167–177
- 17 Fujikawa, T. and Froehlich, L. A. (1966) *Am. J. Obstet. Gynecol.* **95**, 572–578
- 18 Farrell, P. M. and Wood, R. E. (1976) *Pediatrics* **58**, 167–176
- 19 Olowe, S. A. and Akinkugbe, A. (1978) *Pediatrics* **62**, 38–41
- 20 Rishi, A., Hatzis, D., McAlmon, K. and Floros, J. (1992) *Am. J. Physiol. (Lung Cell. Mol. Physiol.)* **6** **262**, L566–L573
- 21 Krizkova, L., Sakthivel, R., Olowe, S. A., Rogan, P. and Floros, J. (1994) *Am. J. Physiol. (Lung Cell. Mol. Physiol.)* **10** **266**, L519–L527
- 22 deMello D. E., Phelps, D. S., Patel, G., Floros, J. and Lagunoff, D. (1989) *Am. J. Pathol.* **134**, 1285–1293
- 23 deMello, D. E., Heyman, S., Phelps, D. and Floros, J. (1993) *Am. J. Pathol.* **142**, 1631–1640
- 24 Nogue, L. M., deMello, D. E., Dehner, L. P. and Colten, H. R. (1993) *New Engl. J. Med.* **328**, 406–410
- 25 Emrie, P. A., Jones, C., Hofman, T. and Fisher, J. H. (1988) *Somat. Cell Mol. Genet.* **14**, 105–110
- 26 Vamvakopoulos, N. C., Modi, W. S. and Floros, J. (1995) *Cytogenetics Cell Genet.* **68**, 8–10
- 27 Jacobs, K. A., Phelps, D. S., Steinbrink, R., Fisch, J., Kriz, R., Mitsch, L., Dougherty, J. P., Taeusch, H. W. and Floros, J. (1987) *J. Biol. Chem.* **262**, 9808–9811
- 28 Pilot-Matias, T. J., Kister, S. E., Fox, J. L., Kropp, K., Glasser, S. W. and Whitsett, J. A. (1989) *DNA* **8**, 75–86
- 29 Todd, S. and Naylor, S. L. (1991) *Nucleic Acids Res.* **19**, 3756
- 30 The Huntington's Disease Collaborative Research Group (1993) *Cell* **72**, 971–983
- 31 Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutter, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barceló, J., O'Hoy, K., LeBlond, S., Earle-MacDonald, J., DeJong, P. J., Wieringa, B. and Korneluk, R. G. (1992) *Science* **255**, 1253–1255
- 32 Fu, Y.-H., Pizzuti, A., Fenwick, R. G., Jr., King, J., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., DeJong, P., Wieringa, B., Korneluk, R., Perryman, M. B., Epstein, H. F. and Caskey, C. T. (1992) *Science* **255**, 1256–1258
- 33 Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., Verkerk, A. J. M. H., Holden, J. J. A., Fenwick, R. G., Jr., Warren, S. T., Oostra, B. A., Nelson, D. L. and Caskey, C. T. (1991) *Cell* **67**, 1047–1058
- 34 Orr, H. T., Chung, M.-Y., Banfi, S., Kwiatkowski, T. J., Jr., Servadio, A., Beaudet, A. L., McCall, A. E., Duvick, L. A., Ranum, L. P. W. and Zoghbi, H. Y. (1993) *Nature Genet.* **4**, 221–226

-
- 35 Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., Saito, M., Tomoda, A., Miike, T., Naito, H., Ikuta, F. and Tsuji, S. (1994) *Nature Genet.* **6**, 9–13
- 36 Aaltonen, L. A., Peltomäki, P., Leach, F. S., Sistonen, P., Pylkkänen, L., Mecklin, J.-P., Järvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., Petersen, G. M., Kinzler, K. W., Vogelstein, B. and de la Chapelle, A. (1993) *Science* **260**, 812–816
- 37 Smith, G. R. and Stahl, F. W. (1985) *BioEssays* **2**, 244–249
- 38 Pryhuber, G. S., Hull, W. M., Fink, I., McMahan, M. J. and Whitsett, J. A. (1991) *Pediatr. Res.* **30**, 597–605
- 39 Morrison, N. A., Qi, J. C., Tokita, A., Kelly, P. J., Crofts, L., Nguyen, T. V., Sambrook, P. N. and Eisman, J. A. (1994) *Nature (London)* **367**, 284–287
- 40 Romeo, G., Ronchetto, P., Luo, Y., Barone, V., Seri, M., Ceccherini, I., Pasini, B., Bocciardi, R., Lerone, M., Kääriäinen, H. and Martucciello, G. (1994) *Nature (London)* **367**, 377–378
- 41 Edery, P., Lyonnet, S., Mulligan, L. M., Pelet, A., Dow, E., Abel, L., Holder, S., Nihoul-Fékété, C., Ponder, B. A. J. and Munnich, A. (1994) *Nature (London)* **367**, 378–380
- 42 Hofstra, R. M. W., Sansvater, R. M., Ceccherini, I., Stulp, R. P., Stelwagen, T., Luo, Y., Pasini, B., Höppener, J. W. M., van Amstel, H. K. P., Romeo, B., Lips, C. J. M. and Buys, C. H. C. M. (1994) *Nature (London)* **367**, 375–376
- 43 Schuchardt, A., D'Agati, V., Larsson-Blomberg, L., Costantini, F. and Pachnis, V. (1994) *Nature (London)* **367**, 380–383
-

Received 11 July 1994/2 September 1994; accepted 6 September 1994