

## Surfactant protein gene A, B, and D marker alleles in chronic obstructive pulmonary disease of a Mexican population

X. Guo\*, H-M. Lin<sup>#</sup>, Z. Lin\*, M. Montañó<sup>¶</sup>, R. Sansores<sup>¶</sup>, G. Wang\*, S. DiAngelo\*, A. Pardo<sup>+</sup>, M. Selman<sup>¶</sup>, J. Floros\*.<sup>§</sup>

*Surfactant protein gene A, B, and D marker alleles in chronic obstructive pulmonary disease of a Mexican population. X. Guo, H-M. Lin, Z. Lin, M. Montañó, R. Sansores, G. Wang, S. DiAngelo, A. Pardo, M. Selman, J. Floros. ©ERS Journals Ltd 2001.*

**ABSTRACT:** Chronic obstructive pulmonary disease (COPD) is characterized by chronic inflammation. It is most likely the result of complex interactions of environmental and genetic factors. Because pulmonary surfactant components play important roles in normal lung function, innate host defence, and inflammation in the lung, this study investigated the hypothesis that the surfactant protein genes are involved in certain cases of COPD.

Genotype analysis of surfactant protein (SP)-A, SP-B, SP-B-linked microsatellite, and SP-D marker alleles was performed in patients with COPD (n=97) and smoker (n=82) or nonsmoker (n=99) controls. Univariate and multiple logistic regression analyses were performed.

The regression analysis results between COPD and smokers revealed several COPD susceptibility alleles (AA62\_A, B1580\_C, D2S388\_5), based on an odds ratio (OR >2.5). The predictive ability of this model for developing COPD is good (c=0.926). Allele-allele (B1580\_C and D2S388\_5) and allele-environment (*i.e.* smoking) interactions were detected. When smoker controls were compared to nonsmoker controls, marker D2S388\_5 appeared to be smoking-independent (p=0.874), whereas marker alleles AA62\_A (p=0.045) and B1580\_5 (p=0.007) were smoking-dependent. Males were at higher risk (OR=6.05, p=0.001), and smoking (>50 packs·yr<sup>-1</sup>) increased risk (OR=5.38, p=0.007). Males and alleles of loci flanking SP-B were associated with more severe cases (forced expiratory volume in one second/forced vital capacity ≤40%).

The present results indicate that the surfactant protein alleles may be useful in chronic obstructive pulmonary disease by either predicting the disease in a subgroup and/or by identifying disease subgroups that may be used for therapeutic intervention. These observations should now be confirmed in a larger study, designed according to strict epidemiological criteria.

*Eur Respir J 2001; 18: 482–490.*

Depts of \*Cellular and Molecular Physiology, <sup>§</sup>Pediatrics, and <sup>#</sup>Health Evaluation Sciences, The Pennsylvania State University College of Medicine, Hershey, PA, USA. <sup>¶</sup>National institute of Respiratory Diseases. <sup>+</sup>Dept of Science National Autonomous University of Mexico, Mexico.

Correspondence: J. Floros, Dept of Cellular and Molecular Physiology, The Pennsylvania State University College of Medicine, 500 University Drive, H166 Hershey, PA 17033, USA. Fax: 1 7175317667

Keywords: Chronic obstructive pulmonary disease  
gene interaction  
surfactant protein  
surfactant proteins A, B and D

Received: May 12 2000  
Accepted after revision May 9 2001

This work was supported by NIH R37 HL34788 and the General Clinical Research Center at the Pennsylvania State University College of Medicine.

Chronic obstructive pulmonary disease (COPD) is a major medical problem and a leading cause of morbidity and mortality among the adult population. The disease is characterized by progressive airflow obstruction due to chronic bronchitis, emphysema, or both [1], and some COPD patients develop predominantly airway disease and others may have predominantly parenchymal disease. Although the most important risk factor by far for the development of COPD is cigarette smoking, only 10–20% of heavy cigarette smokers develop COPD [1, 2], suggesting that other environmental or genetic factors contribute to COPD. Environmental risk factors include childhood respiratory infections, occupational exposures, ambient air pollution, lower birth weight, and diet [3].

Evidence for genetic susceptibility factors is derived from studies of twins where the lung functional response to cigarette smoking was studied [4], and from studies showing an aggregation of COPD in families [5]. Genetic risk factors for COPD may include the inherited deficiency of  $\alpha_1$ -antitrypsin in

individuals with the Z allele [6], which is uncommon and explains only a very small proportion (<1%) of the COPD cases [1]. Genetic polymorphisms in the xenobiotic enzyme [7], and vitamin D binding protein [8] appear to be associated with an increased or decreased risk for COPD, respectively.

Although pulmonary surfactant or its components have the potential to be contributors to the pathogenesis of COPD, very little work has been done in this regard. Pulmonary surfactant, a lipoprotein complex, is essential for normal lung function and surfactant, or components of surfactant, play a role in bronchiolar stability, innate host defence, and the regulation of the inflammatory processes in the lung [9]. The surfactant proteins (SPs) play important roles in surfactant function, structure, and metabolism [9], and SP-A and SP-D are involved in host defence and/or the regulation of inflammatory processes of the lung.

The human SP-A locus is located on chromosome 10 and consists of two functional genes in opposite transcriptional orientation [10]. The human SP-D is

linked to the SP-A locus and is located proximal to the centromere at about 80–100 kb from the SP-A2 gene [10]. A number of alleles have been characterized for each SP-A gene. The most commonly observed alleles for the SP-A1 gene are the 6A, 6A<sup>2</sup>, 6A<sup>3</sup>, 6A<sup>4</sup> alleles, and for the SP-A2 gene are the 1A, 1A<sup>0</sup>, 1A<sup>1</sup>, 1A<sup>2</sup>, 1A<sup>3</sup>, 1A<sup>5</sup> [11–13], (unpublished observations). Functional differences between SP-A1 and SP-A2 alleles, and possibly among alleles of each gene have been demonstrated with *in vitro* expressed human SP-A alleles [14]. Moreover, splice variation and/or polymorphisms at the 5' [12] and 3' untranslated regions of these alleles, respectively, point to regulatory differences [15]. Several polymorphisms have also been identified for SP-D [13] and an association between an SP-D allele, as well as SP-A alleles and tuberculosis, have been observed [16]. The human SP-B locus is located on chromosome 2p12-p11.2. A number of polymorphisms have been characterized for SP-B and some of these have been associated with disease [17–20]. In addition, a number of microsatellite markers flanking the SP-B locus have been characterized [21].

Because COPD is characterized by a chronic inflammation and abnormal lung function and is probably the result of complex interactions of environmental and genetic factors, this report has investigated whether genetic variation in surfactant proteins contributes to COPD by assessing any associations with surfactant protein marker alleles. Allele associations may help to genetically identify more homogeneous COPD subgroups to study mechanisms involved in the pathogenesis of COPD, and/or study the response to therapeutic interventions.

### Patients and methods

#### *Study population*

The study was conducted at the National Institute of Respiratory Diseases (INER) in Mexico City, according to institutional protocol for human subjects. INER is one of the National Institutes of Health in Mexico and is a tertiary referral and research centre.

Patients and healthy smoker controls were sequentially enrolled from the Smoking Cessation Program at INER, according to their willingness to participate in the study. Individuals with a smoking history of >10 pack-yrs, forced expiratory volume in one second (FEV<sub>1</sub>) <70% predicted, FEV<sub>1</sub>/forced vital capacity (FVC) <70% and with no history suggestive of asthma were considered COPD patients (cases). Subjects with smoking history of >10 pack-yrs, but with FEV<sub>1</sub> >70% predicted, and FEV<sub>1</sub>/FVC >70% were considered smoker controls. In addition, sequential, unrelated nonsmoking healthy blood donors from INER were included. The transfusion service at INER serves the geographical area from which the patients were recruited. Patients and controls had >3 previous generations born in Mexico.

#### *Pulmonary function test*

*Spirometry.* FEV<sub>1</sub> and FVC were determined by using a turbine spirometer (Pony Cosmed Inc., Rome, Italy).

According to the criteria of the American Thoracic Society (ATS), the best of three acceptable forced expiratory manoeuvres is used to select FEV<sub>1</sub> and FVC while the best FEV<sub>1</sub> and the best FVC from either tracing are used for the FEV<sub>1</sub>/FVC ratio [1].

The Pony spirometer used in this study fulfils the equipment recommendations by the ATS in terms of range/accuracy for FVC, FEV<sub>1</sub>, flow range (0–14 L·s<sup>-1</sup>), time, resistance and back-pressure. In addition, in patients with COPD, a good concordance was observed between the body plethysmography and spirometric results, which further supports the present data.

Based on the present authors' experience, the readings of the calibration sessions, which were made by the same technicians twice a week, had <3% differences. In addition, a biological control is used (one of the technicians is submitted two- or three-times a week to the test). The accumulated results from this control test over a year indicate that differences in reproducibility are <3%.

*Lung volumes and diffusing capacity.* Total lung capacity (TLC) and residual volume (RV) were obtained by using a body plethysmograph (MasterLab, Jaegger, Frankfurt, Germany). All subjects were previously familiarized with the equipment before the tests. For carbon monoxide diffusing capacity of the lung (DL<sub>CO</sub>) manoeuvres, >2 tests with ≤5% variability, were considered to be acceptable. Patients with FVC <1 L and/or unable to follow instructions or to hold their breath for 10 s, were not submitted to this test. The coefficient of variation and intraclass correlation coefficient (ICC) for FEV<sub>1</sub> and FVC were 3% and 0.98 respectively; for TLC, RV and DL<sub>CO</sub>, these were in the range of 5–12% and 0.90–0.92, respectively. For all the pulmonary function tests, the reference values reported by QUANIER [22] were used.

*Arterial blood gases.* These were obtained through a radial puncture while patients were breathing room air for >30 min and measured in a gasometer (IL 1310, Instrumentation Laboratory Inc., Lexington, MA, USA).

#### *Genotype analysis for surfactant protein-A, -B, and -D using a polymerase chain reaction-based converted restriction fragment length polymorphism method*

The converted polymerase chain reaction (PCR) and converted restriction fragment length polymorphism/genotype analysis (cRFLP) approaches are based on the notion that a site containing a single base polymorphism (if it does not represent a natural enzyme recognition site) is converted into a restriction enzyme recognition site by simply using a PCR primer that contains the necessary mismatched base. Because the amplified fragment will contain PCR products from both alleles, it is expected that in the case of homozygotes for either one or the other allele, the PCR products will either both be digested or not by the appropriate enzyme. In the case of heterozygotes, one allele would be digested and the other will remain intact. Following electrophoresis and visualization of the digested PCR products, one can assess the genotype at the particular site for the given individual.

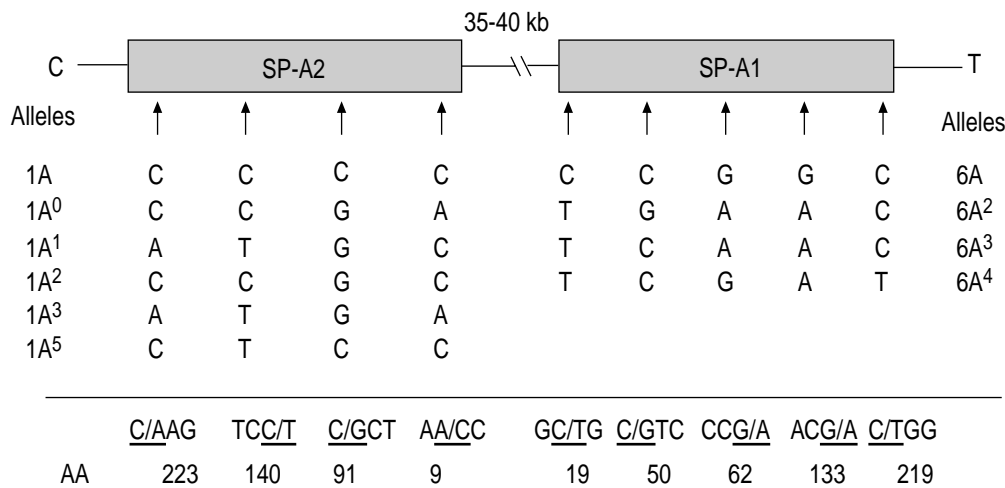


Fig. 1. – Surfactant protein (SP)-A2 and SP-A1 marker loci used in genotype analysis. The location of SP-A2 and SP-A1 genes relative to the centromere (C) and telomere (T) is shown. The SP-A1 and SP-A2 genes are in opposite transcriptional orientation [10]. The amino acid location of each bi-allelic SP-A2 and SP-A1 marker is shown on the bottom line, *i.e.* AA9 (the first A stands for SP-A and the second A for amino acid; the number denotes the actual amino acid). Above each amino acid location, the corresponding codon is shown and the single base polymorphism genotyped at each amino acid location is underlined in the corresponding codon. For SP-A2 alleles, bi-allelic markers at four amino acid locations (AA9, AA91, AA140 and AA223) are genotyped using converted polymerase chain reaction restriction fragment length polymorphism analysis. For example, allele 1A has the C polymorphism at all four amino acid locations. Each SP-A2 allele (1A, 1A<sup>0</sup>, 1A<sup>1</sup>, *etc.*) is determined by the combined pattern of the four bi-allelic markers, as shown. Similarly, for the SP-A1 alleles, bi-allelic markers at five amino acid locations (AA19, AA50, AA62, AA133 and AA219) are genotyped to determine SP-A1 alleles (6A, 6A<sup>2</sup>, *etc.*). C, A, G, T: nucleotide bases cytosine, adenine, guanine and thymine, respectively.

For all PCR reactions, except the 11 kb fragment of SP-B (see later), the thermocycler, ERICOMP TwinBlock™ system was used. For the 11 kb, the Perkin Elmer Cetus DNA Thermal Cycle (Applied Biosystems, Foster City, CA, USA) was used.

**Genotype analysis.** Surfactant protein-A. Genomic deoxyribonucleic acids (DNAs) were used as a template to first amplify a 3.3 kb SP-A1 or SP-A2 specific fragment, which then served as a template for subsequent converted PCR reactions.

cRFLP analysis was then performed for nucleotide changes at codons of five (AA19, AA50, AA62, AA133, and AA219) and of four (AA9, AA91, AA140 and AA223) amino acids, in order to score

SP-A1 and SP-A2 alleles, respectively (fig. 1), as described in detail by DIANGELO *et al.* [13].

**Surfactant protein-B.** The SP-B genotyping was carried at four nucleotide positions Q<sub>-18</sub> (A/C), 1013 (A/C), 1580 (C/T), and 9306 (A/G) of the SP-B gene (fig. 2) and the scoring of alleles was made according to the DNA patterns as previously described [18, 20].

**Surfactant protein-B linked microsatellite markers.** Genotyping was carried out for four (D2S2232, D2S388, (AAGG)<sub>n</sub> and GATA41E01) SP-B flanking microsatellite markers (fig. 2), as described previously [20, 21]. For each marker, a specific end labelled <sup>32</sup>P-γ adenosine triphosphate primer was used for PCR, the

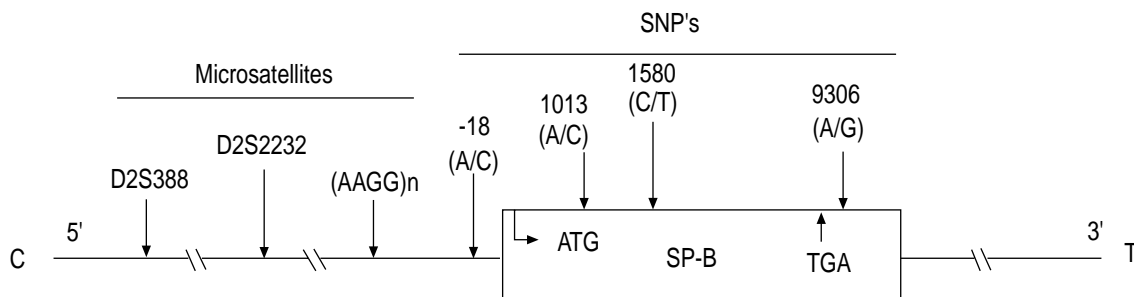


Fig. 2. – Surfactant protein (SP)-B and microsatellite (SP-B-linked) marker loci. The location of the SP-B microsatellite markers relative to the centromere (C) and telomere (T) is shown. The relative position of D2S388 and D2S2232 is according to TNG3 high resolution radiation hybrid panel analysis [21] and the draft (June 6, 2001) of the human genome map of chromosome 2 [23]. The distance between SP-B and microsatellite marker loci D2S388 and D2S2232, according to the human genome map, is 158 kb and 62 kb, respectively. The position of (AAGG)<sub>n</sub> is, according to KALA *et al.* [21], at a distance of about 27 kb from SP-B. There is a discrepancy in the position of GATA41E01. The G3 medium resolution radiation hybrid panel analysis [21] places it at the telomeric side of SP-B and the human genome map at the centromeric side of SP-B, distal to the D2S388 marker, at about 270 kb from SP-B. The nucleotide location of the four SP-B bi-allelic marker loci or single nucleotide polymorphisms (SNPs) is shown [18]. ATG: translation start codon; TGA: translation stop codon.

PCR products were run on a 6% polyacrylamide gel containing formamide, the gel dried, and exposed to X-AR film. The scoring of alleles was made according to control allele DNA patterns as previously described [21].

Surfactant protein-D. For SP-D genotyping, two fragments (376 bp and ~1 kb) were amplified from genomic DNA. One (376 bp) contained the codon ATG or ACG for amino acid 11 (DA11), and the other contained the codon ACA or GCA for amino acid 160 (DA160). The underlined nucleotide is the polymorphic nucleotide. Each fragment served as a template for converted PCR at DA11 or DA160 sites, as previously described [13].

### Statistical analyses

The analyses performed for the genotype data include univariate analysis and multiple logistic regression analysis. The three groups of individuals are: 1) nonsmoker control, 2) smoker control, and 3) COPD, all of whom were smokers. The genotype fields considered in the analyses are summarized in table 1 and the location and characteristics of these are depicted in figures 1 and 2.

### Univariate analyses

The Chi-squared test was used for comparison of the distributions of alleles between two groups (*i.e.* control and COPD, smoker and nonsmoker), except when the expected frequency of the markers was too small, in which case the Fisher's Exact test was used. For each of the bi-allelic SP-A, SP-B, and SP-D markers, analyses were also performed using logistic regression as described previously [16], adjusting for age, smoking (>50 packs·yr<sup>-1</sup>) and sex. In this analysis allele dose-effect was also evaluated, *i.e.* being homozygous for a given allele has a different effect than being heterozygous. Because only a few tests

Table 1.—Surfactant proteins (SP)-A, SP-B, SP-D, and SP-B-linked microsatellite markers used in statistical analyses

SP-A2	SP-A1	SP-B	SP-D	Micro-satellites		
AA9	1A	AA19	6A	B_18A/C	DA11C/T	D2S388
AA91	1A <sup>0</sup>	AA50	6A <sup>2</sup>	B1013A/C	DA160A/G	D2S2232
AA140	1A <sup>1</sup>	AA62	6A <sup>3</sup>	B1580C/T		AAGG
AA223	1A <sup>2</sup>	AA133	6A <sup>4</sup>	B9306A/G		GATA
	1A <sup>3</sup>	AA219	6A <sup>13</sup>			
	1A <sup>5</sup>		6A <sup>14</sup>			
	1A <sup>8</sup>					
	1A <sup>9</sup>					

The markers of the left subcolumn of the SP-A2 (*i.e.* AA9, *etc.*) and of SP-A1 (*i.e.* AA 19 *etc.*) as well as the SP-B and SP-D markers are all single nucleotide polymorphisms. The markers on the right subcolumn of SP-A2 (*i.e.* 1A, 1A<sup>0</sup>, *etc.*) and of SP-A1 (*i.e.* 6A, 6A<sup>2</sup>, *etc.*) are haplotypes as depicted in figure 1, according to DIANGELO *et al.* [13]. The micro-satellites are size polymorphisms [21].

were significant for dose-effect, it was reasoned that 1 allele has the same effect as 2 identical alleles. For the multi-allelic SP-A1 and SP-A2 markers and for microsatellites, similar analyses were performed as for the bi-allelic markers [16].

### Multivariate analyses

Assuming no dose-effect for the alleles, multivariate logistic regression models were built using backward selection methods with staying significance level equal to 0.05. The alleles that entered the model were preselected, *i.e.* only the alleles shown to be significant in the univariate analysis ( $p < 0.1$ ) were considered in the model. Sex, age, and smoking (>50 packs·yr<sup>-1</sup>) were forced to be included as variables in the model that compares COPD and healthy smokers, because further analysis showed them to be confounders for certain alleles. For smoker/nonsmoker comparison, only sex was included because age was unknown for nonsmokers.

## Results

### Chronic obstructive pulmonary disease versus smoker control

The clinical and functional characteristics of COPD patients and control smokers are shown in table 2, and the impact of pack-yrs on those who develop COPD and those who do not ( $p = 0.001$ ) is shown in table 3. An association (Chi-squared 16.9,  $p < 0.001$ ) between heavy smokers (>50 packs·yr<sup>-1</sup>) with COPD ( $n = 36$ ), and without COPD ( $n = 9$ ) was observed. An association between the intensity of smoking and sex ( $p = 0.012$ ) was also observed when smokers with or without COPD were considered. Within the COPD group there was no correlation between the intensity of smoking and severity of COPD, or between sex and severity of COPD.

Marker alleles that showed significant differences ( $p < 0.1$ ) in the univariate analysis (table 4) were

Table 2.—Clinical and functional characteristics of chronic obstructive pulmonary disease (COPD) patients and smoker controls

Parameter	Smokers		p-value
	With COPD	Without COPD	
n	101	81	
Age yrs	67±8	53±10	<0.01 <sup>#</sup>
Sex M:F	86:15	44:37	<0.0001 <sup>†</sup>
P <sub>a</sub> O <sub>2</sub> <sup>+</sup> mmHg	56±11	77±8	<0.001
RV % pred	209±60	ND	
TLC % pred	126±23	ND	
DL <sub>CO</sub> % pred	58±29	ND	

Data are presented as mean±SD; <sup>†</sup>: Fisher's exact test was used as opposed to t-test; <sup>+</sup>: normal arterial oxygen tension (P<sub>a</sub>O<sub>2</sub>) in Mexico City is 67±3 mmHg. M: male; F: female; ND: not done; RV: residual volume; TLC: total lung capacity; DL<sub>CO</sub>: carbon monoxide diffusion capacity of the lung. <sup>#</sup>: If nonparametric test is used  $p < 0.0001$ .

Table 3. – Intensity of smoking in individuals with or without chronic obstructive pulmonary disease (COPD)

Pack·yrs <sup>-1</sup>	Smokers	
	With COPD	Without COPD
10–20	18	29
21–30	14	19
31–40	16	10
41–50	17	14
51–60	10	3
61–70	5	2
>71	21	4

Smokers with  $\leq 50$  versus  $\geq 51$  packs·yr<sup>-1</sup>, with (n=101) or without (n=81) COPD; Chi-squared: 16.9,  $p < 0.001$ .

included in the multivariate analysis. A number of marker alleles were significant. These included one SP-A1 (AA62\_A), one SP-B (B1580\_C), and one microsatellite (D2S388\_5) (table 5). Based on the OR, all three marker alleles, as well as sex (males), age, and smoking ( $>50$  packs·yr<sup>-1</sup>) appear to be associated with increased risk for COPD, ranging from 1.17-fold (age) to 7.42-fold (AA62\_A). Having CCA at the marker locus AA62, instead of CCG, or having 1580 ACT at B, instead of ATT, increases the susceptibility by 7.42- and 3.65-fold, respectively. The transition from A to G at AA62 does not change the encoded amino acid, but the transition of C to T at B1580 changes the encoded amino acid from a threonine to a serine, resulting in the elimination of a potential N-linked glycosylation site.

The *c*-statistic, a measure as to how well the model predicts an individual's expected probability of developing COPD, is 0.926, suggesting that this is a good model. The *c*-statistic should be between 0.5–1.0; the larger the *c*-statistic, the better the predictive ability of the model would be [24]. Similar results to those described earlier were obtained when the COPD group was compared to the entire control group (smokers and nonsmokers).

#### Allele interactions and confounding factors

Interaction exists when the presence of marker A would affect the susceptibility of having COPD for

Table 5. – Chronic obstructive pulmonary disease (COPD) versus smoker control multiple logistic regression analysis

Variable*	Odds ratio	95% confidence interval	p-value
Male	6.05	2.11–19.17	0.001
Age, yrs	1.17	1.11–1.25	<0.0001
>50 pack·yr <sup>-1</sup>	5.38	1.67–20.40	0.007
AA62_A	7.42	1.40–51.66	0.028
B1580_C	3.65	1.19–12.12	0.027
D2S388_5	5.75	1.87–20.87	0.004

COPD: n=94; smoker: n=77; \*: Variables (except sex) were significant ( $p < 0.10$ ) in univariate tests (table 4); *c*=0.926.

individuals with another marker B. Note that in the interaction analysis, again sex, smoking, and age were adjusted for in the COPD/smoker comparison. In the COPD and healthy smoker comparison, an interaction ( $p=0.055$ ) was observed between B1580\_C and D2S388\_5. The combined presence of two of the alleles studied here may better predict COPD in certain cases (fig. 3). If both B1580\_C and D2S388\_5 are present, the OR for COPD is 24.3 ( $p=0.0004$ ) compared to the situation where both markers are absent (A, fig. 3). No significant differences were observed for situations B and C in figure 3 when either of these was compared to A.

Interactions between genotype and smoking were also tested within the COPD and the smoker groups and none was found to be statistically significant, after including pack-yrs and individual genotypes into the model. However, the failure to detect any significant interaction term may only imply that the association between genotype and COPD is not affected by the amount of smoking, but it does not necessarily mean that the gene-disease association cannot be modified by the exposure of smoking. Because both the COPD cases and the smoker controls were all exposed to smoking, this latter possibility could not be tested directly, but when the COPD cases were compared to nonsmoking controls, two different alleles (6A<sup>4</sup>/D160\_G) were significant (unpublished data) providing partial indirect support to this hypothesis. Thus, to gain further insight into present study COPD/control comparisons, a univariate analysis of smoker versus nonsmoker control was performed

Table 4. – Chronic obstructive pulmonary disease (COPD) versus smoker controls (univariate analysis)

Allele	Smoker		COPD		OR*	95% CI*	p-value*	OR**	95% CI**	p-value**
	n	(%)	n	(%)						
AA62_A	82	84.2	97	91.8	2.10	(0.82–5.34)	0.12	4.09	(1.05–15.92)	0.04
B1580_C	79	67.1	95	82.1	2.25	(1.11–4.55)	0.02	3.39	(1.18–9.77)	0.02
D2S388_4	81	33.3	97	22.7	0.59	(0.30–1.14)	0.11	0.39	(0.15–0.98)	0.05
D2S388_5	81	25.9	97	43.3	2.18	(1.15–4.13)	0.02	4.71	(1.62–13.70)	<0.01
D2S2232_1	78	41.0	93	26.9	0.53	(0.28–1.00)	0.05	0.41	(0.17–1.00)	0.05
D2S2232_3	78	34.6	93	54.8	2.29	(1.23–4.26)	0.01	1.98	(0.84–4.71)	0.12
AAGG_3	81	72.8	96	60.4	0.57	(0.30–1.08)	0.08	0.40	(0.16–1.00)	0.05

Smoker and COPD data are presented as total number of samples analysed and percentage of n with at least one copy of the given allele. \*: Not adjusted for sex, age, smoking; \*\*: adjusted for sex, age, and smoking ( $>50$  packs·yr<sup>-1</sup>) using logistic regression. OR: odds ratio; CI: confidence interval.

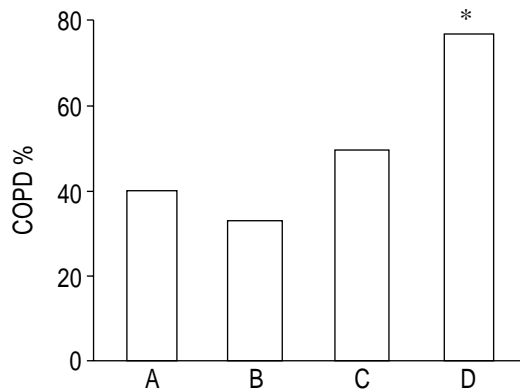


Fig. 3.—Interaction between marker alleles B1580\_C (1) and D2S388\_5 (2) in chronic obstructive pulmonary disease (COPD) *versus* smoker controls after adjusting for age, sex and number of pack-yrs. A: negative for both alleles 1 and 2; 13 from 31 (42%) subjects were diagnosed with COPD. B: negative for allele 1 and positive for allele 2; 4/12 (33%) subjects had COPD. C: positive for allele 1 and negative for allele 2; 40/81 (49%) subjects had COPD. D: positive for both alleles 1 and 2; 38/50 (76%) subjects had COPD. The p-values are based on logistic regression after the above adjustments. The two marker alleles appear to interact. When both marker alleles are present (D), the odds ratio (OR) for COPD is considerably increased (OR=24.3, (\*) p<0.001) compared to when both markers are absent (A). No significant differences are observed when one marker is present and the other absent (OR for B=0.73, p=0.81; OR for C=1.9, p=0.31).

(table 6), assuming no dose effect, for the three alleles (AA62\_A, D2S388\_5, B1580\_C) that remained in the model of COPD *versus* smokers. The frequency of alleles AA62\_A and B1580\_C was significantly different between smokers and nonsmoker controls, after adjusting for sex (p=0.045 and p=0.007, respectively, and ORs of 0.337 and 0.346, respectively). However, the frequency of allele D2S388\_5 was not significantly different in these two control groups (OR=0.94, p=0.874). These findings indicate that the two control groups (smokers and nonsmokers), based on the marker alleles used in this study, are different but overlapping populations. Also, the results of the smoker/nonsmoker comparison identify the D2S388\_5 marker allele as a "smoking-independent" marker allele and the AA62\_A and B1580\_C as "smoking-dependent", pointing to the complexity of the underlying interactions. How smoking affects these marker alleles remains to be determined.

This study also investigated the possibility of

Table 6.—Smoker\* *versus* nonsmoker# control (univariate analysis)

Variable	Odds ratio	95% CI of OR <sup>†</sup>	p-value
AA62_A	0.38	0.15–0.98	0.045
B1580_C	0.35	0.16–0.75	0.007
D2S388_5	0.94	0.46–1.95	0.874

\*: Number of smokers: n=81 for AA62\_A and D2S388\_5 and n=79 for B1580\_C; #: Number of nonsmokers: n=99 for AA62\_A and B1580\_C and n=98 for D2S388\_5; †: Wald confidence intervals (CI) of odds ratios (OR), adjusting for sex.

confounders. A confounder would lead to a biased estimate of the OR, but it may not affect the homogeneity of the ORs between the two groups. For example, without adjusting for the confounder (say marker B), the OR for marker A is 3. After adjusting for the confounder, the OR for marker A becomes 5. In this example, the OR of 3 is a biased estimate.

In the COPD/smoker comparison, the OR for the AA62\_A marker allele in the univariate analysis is 4.09 (table 4), but in the multivariate analysis is 7.42 (table 5), whereas the ORs for B1580\_C and D2S388\_5 are similar between univariate and multivariate analyses (tables 4 and 5). These observations indicate that the AA62\_A marker allele is confounded by some other marker(s). Further analysis indicated that B1580\_C may be a confounder for AA62\_A, but D2S388\_5 is not a confounder for AA62\_A.

#### Chronic obstructive pulmonary disease subgroups based on severity

To determine whether marker alleles are associated with severity in COPD, the COPD population was separated into two subgroups based on FEV<sub>1</sub>/FVC values. One group included patients (n=46) with FEV<sub>1</sub>/FVC>40% and the other group, patients (n=47) with FEV<sub>1</sub>/FVC ≤40%. Multiple logistic regression analysis was performed that included only marker alleles shown to be significant in the univariate tests. Age and smoking appear to be nonsignificant risk factors, but sex appears to be significant. Therefore, sex was forced as a variable in the model. Two marker alleles (B\_18\_C and D2S388\_4), as well as sex (males) were significant (table 7). All three factors appear, based on ORs, to be associated with severe cases of COPD by approximately 3–4-fold. These results were also confirmed in a multivariate analysis (not shown) that included all marker alleles depicted in table 1, without preselection of significant (p<0.1) marker alleles.

#### Discussion

As a way of gaining insight into the possible role of the surfactant proteins in the pathogenesis of COPD, this study carried out case-control association studies to determine the usefulness of these markers in identifying COPD subgroups with increased or

Table 7.—Chronic obstructive pulmonary disease (COPD) subgroups: forced expiratory volume in one second (FEV<sub>1</sub>)/forced vital capacity (FVC)\* ≤40% *versus* >40%

Variable	Odds ratio	95% confidence interval	p-value
Male	3.60	1.07–14.6	0.048
B_18_C	3.03	1.06–9.50	0.044
D2S388_4	4.11	1.42–13.5	0.012

\*: FEV<sub>1</sub>/FVC: >40% (n=46); ≤40% (n=47); c=0.719.

decreased susceptibility. The findings from the comparison of patients with COPD and smoker controls revealed that SP-A1 (AA62\_A), SP-B (B1580\_C) and microsatellite (D2S388\_5) marker alleles may increase susceptibility to COPD and that sex (male), smoking, and age are important susceptibility factors for COPD. In addition, certain marker alleles (B\_18\_C and D2S2388\_4) and sex (male) were associated with severe cases of COPD. A comparison of smoker and nonsmoker controls revealed both differences and similarities in the three significant marker alleles (AA62\_A, B1580\_C, D2S388\_5) between COPD and smoker controls. It is possible that a combination of markers (surfactant proteins and/or others) that are both "smoking-dependent" and "smoking-independent" are required for development of COPD in certain subgroups. In the present study, the COPD susceptibility D2S388\_5 marker allele is independent of smoking, as no significant differences were observed between smoker and nonsmoker controls, whereas marker alleles AA62\_A and B1580\_C are "smoking-dependent". Thus, the surfactant protein genetic marker alleles either alone and/or along with other marker alleles may be useful in identifying disease subgroups. Such subgroups of patients may provide adequate genetic homogeneity to study mechanisms involved in the pathogenesis of COPD subgroups and/or be useful in studies of therapeutic intervention when drug response is being investigated.

The regression analysis data revealed that smoking is an important variable for COPD, and this is consistent with the observed association ( $p < 0.001$ ) between heavy smokers ( $>50$  packs·yr<sup>-1</sup>) and COPD (table 3). Although, smoking is a major determinant of COPD, only a fraction of smokers develop COPD [1, 2]. Smokers with normal FEV<sub>1</sub> have been shown to have increased inflammation, but the characteristics of this inflammation differ between smokers who develop COPD and those who do not [25]. Genetic factors have been implicated in smoking [26] and perhaps yet unknown genetic factors may determine which subgroup of smokers has increased susceptibility to COPD. The surfactant protein marker alleles may help to distinguish such a subgroup. A subgroup of smokers, identified by marker alleles (AA62\_A and B1580\_C), have increased susceptibility to COPD, compared to smokers who lack these alleles.

Whether SP-A1 (AA62\_A), SP-B (B1580\_C), or SP-B-linked microsatellite (D2S388\_5) marker loci themselves, or linked loci, contribute to COPD remains to be determined. Of interest, heterozygous SP-B mice have half the amount of SP-B (compared to homozygous SP-B) and under normal conditions, exhibit small physiological lung abnormalities [27]. It is possible that certain SP-B alleles may associate with compromised lung function under certain circumstances. For example, the marker allele B1580\_C, associated with COPD, encodes a threonine (instead of isoleucine) at the marker locus B1580. A threonine at this marker locus can result in an additional N-linked glycosylation site. Although the functional role of such a change is not known, it may have an impact on the processing of the precursor SP-B

molecule resulting in low SP-B content, which in turn may compromise lung function. Conversely, the SP-A1 AA62\_A single nucleotide polymorphism, present in alleles 6A<sup>2</sup> and 6A<sup>3</sup> (fig. 1), and which does not change the encoded amino acid, may identify regulatory differences between AA62\_A and other SP-A1 alleles and/or chromosomes or groups of alleles that contribute to COPD. Similarly, the microsatellite D2S388 marker locus is an unknown marker locus and the D2S388\_5 marker allele may be linked to a gene that contributes to COPD.

SP-A and SP-B are functionally interactive, because both are necessary for aspects of surfactant such as the structural form of surfactant called tubular myelin [9]. It is possible that the functional capabilities of certain surfactant protein allele combinations may compromise, significantly more, host defence and/or alveolar integrity, through networks of additive and/or epistatic interactions [28–30] at the molecular, cellular, or tissue level. However, in the present study, no interaction was observed between AA62\_A (or 6A<sup>2</sup> and 6A<sup>3</sup>; fig. 1) and either of the two (B1580\_C, D2S388\_5) significant markers (not shown), suggesting that the presence of one marker allele does not affect susceptibility to COPD for individuals with the other marker. Conversely, smoking may have an impact on alleles AA62\_A and B1580\_C, since these alleles appeared to be "smoking-dependent". Although the underlying mechanisms are unknown, genetic factors playing a role in cigarette smoking have been previously suggested [26].

Furthermore, surfactant protein marker alleles may be useful in separating subgroups of COPD based on severity. Two marker alleles on chromosome 2 (B\_18\_C and D2S2388\_4) were associated with severe cases of COPD. Males were also more likely (OR=3.6) to be in the severe (FEV<sub>1</sub>/FVC  $\leq 40\%$ ) COPD subgroup, which is also consistent with the strong association between males ( $p < 0.0001$ , table 2) and COPD. Of relevance is a recent preliminary finding where an intron 4 variant of the SP-B gene [17] was observed in higher frequency in acute respiratory failure in COPD of a German population. The frequency of this variant was higher in females in this severe subgroup of COPD [31]. These data together, suggest that sex is important in COPD and that different marker alleles may identify increased severity for males and females. Smoking was no longer a variable in the COPD severity subgroups, which is consistent with the lack of interaction of smoking and genotype within the COPD group. These data indicate that the amount of smoking may not play a role in the severity of COPD but it may modify gene-disease interactions, and through this modification, smoking contributes to COPD in individuals with certain genetic backgrounds. In this regard, the surfactant protein marker alleles may be useful in identifying COPD subgroups.

For the markers tested, there is a possibility that some of the results are significant by chance alone. For example, out of the 33 markers considered in table 1, if each marker is tested individually at a statistical significance level of  $p < 0.05$ , the probability of observing at least one significant marker by chance

is 0.82, of three or more markers by chance is 0.23, and of six or more markers by chance is 0.05. Four and six markers were identified in table 4, with  $p < 0.05$  from the Chi-squared test and the logistic regression, respectively. Thus, some of the markers found in table 4 may be expected to be due to chance alone and some may not. Although, the probabilities (0.23 and 0.05) are low, indicating that the probability of the "chance" factor alone being a major factor is indeed low, the observations made need to be confirmed in a larger study designed according to strict epidemiological criteria.

In summary, the data presented implicate surfactant proteins in chronic obstructive pulmonary disease, and point to the usefulness of surfactant protein marker alleles in studies of the disease. Because chronic obstructive pulmonary disease is a disease with complex aetiology, it is possible that difficulties exist identifying homogeneous clinical subgroups for study. The surfactant protein marker alleles may be useful in this regard, and contribute to the identification of a significant chronic obstructive pulmonary disease subgroup. Such a subgroup, along with other subgroups may be used, for example, in studies of drug response, or other therapeutic intervention. Furthermore, the use of genetic markers as additional parameters to identify more homogeneous subgroups may facilitate the study of mechanisms involved in the pathogenesis of specified chronic obstructive pulmonary disease subgroups, that may in turn help in the understanding of why, for example, only a fraction of smokers develop the disease.

**Acknowledgements.** The authors thank S. Phillips, B. Phillips, and A. Matthews for their contributions, and S. Myers for typing this manuscript.

### References

1. ATS Statement. Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease. Definitions, Epidemiology, Pathophysiology, Diagnosis, and Staging. *Am J Respir Crit Care Med* 1995; 152: S78–S83.
2. Bascom R. Differential susceptibility to tobacco smoke: possible mechanisms. *Pharmacogenetics* 1991; 1: 102–106.
3. Higgins MW, Thom T. Incidence, prevalence, and mortality: intra- and inter-county differences. In: Hensley MJ, Saunders NA, eds. Clinical epidemiology of chronic obstructive pulmonary disease. New York, Marcel Dekker., 1990; pp. 23–43
4. Webster PM, Lorimer EG, Man SFP, Woolf CR. Pulmonary function in identical twins: comparison of nonsmokers and smokers. *Am Rev Respir Dis* 1979; 119: 223–228.
5. Khoury MJ, Beaty TH, Tockman MS, Self SG, Cohen BH. Familial aggregation in chronic obstructive pulmonary disease: Use of the log linear model to analyze intermediate environmental and genetic risk factors. *Genet Epidemiol* 1985; 2: 155–166.
6. Sandford AJ, Weir TD, Spinelli JJ, Pare PD. Z and S mutations of the alpha1-antitrypsin gene and the risk of chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol* 1999; 20: 287–291.
7. Smith CAD, Harrison DJ. Association between polymorphism in gene for microsomal epoxide hydrolase and susceptibility to emphysema. *Lancet* 1997; 350: 630–633.
8. Horne SL, Cockcroft DW, Dosman JA. Possible protective effect against chronic obstructive lung disease by the GC 2 allele. *Hum Hered* 1990; 40: 173–176.
9. Floros J, Phelps DS. Pulmonary Surfactant. In: Biebuyck J, Lynch IIC, Maze M, Saidman LJ, Yaksh TL, Zapol WM, eds. Anaesthesia: Biologie Foundations. New York, Lippincott-Raven, 1997; pp 1257–1279.
10. Hoover RR, Floros J. Organization of the human SP-A and SP-D loci at 10q22-23. Physical and radiation hybrid mapping reveals gene order and orientation. *Am J Respir Cell Mol Biol* 1998; 18: 353–362.
11. Floros J, DiAngelo S, Koptides M, et al. Human SP-A locus: Allele frequencies and linkage disequilibrium between the two surfactant protein A genes. *Am J Respir Cell Mol Biol* 1996; 15: 489–498.
12. Karinch AM, Floros J. 5' splicing and allelic variants of the human pulmonary surfactant protein A genes. *Am J Respir Cell Mol Biol* 1995; 12: 77–88.
13. DiAngelo S, Lin Z, Wang G, et al. Novel, non radioactive, simple and multiplex PCR-cRFLP methods for genotyping human SP-A and SP-D marker alleles. *Dis Markers* 1999; 15: 269–281.
14. Wang G, Phelps DS, Umstead TM, Floros J. Human SP-A protein variants derived from one or both genes stimulate TNF- $\alpha$  production in the THP-1 cell line. *Am J Physiol* 2000; 278: L946–L954.
15. Hoover RR, Floros J. SP-A 3'UTR is involved in the glucocorticoid inhibition of human SP-A gene expression. Differential allelic regulation. *Am J Physiol* 1999; 276: L917–L924.
16. Floros J, Lin HM, Garcia A, et al. Surfactant protein genetic marker alleles identify a subgroup of tuberculosis in a Mexican population. *J Infect Dis* 2000; 182: 1473–1478.
17. Floros J, Veletza SV, Kotikalapudi P, et al. Dinucleotide repeats in the human surfactant protein B gene and respiratory distress syndrome. *Biochem J* 1995; 305: 583–590.
18. Lin Z, deMello DE, Wallot M, Floros J. An SP-B gene mutation responsible for SP-B deficiency in fatal congenital alveolar proteinosis: Evidence for a mutation hotspot in Exon 4. *Mol Genet Metab* 1998; 64: 25–35.
19. Lin Z, Wang G, deMello DE, Floros J. An alternatively-spliced surfactant protein B mRNA in normal human lung: Disease implication. *Biochem J* 1999; 343: 145–149.
20. Pearson C, Lin Z, Pison U, Floros J. SP-B gene polymorphisms and acute respiratory distress syndrome in a German population. *Am J Respir Crit Care Med* 1999; 159: A507.
21. Kala P, Koptides M, DiAngelo S, et al. Characterization of markers flanking the human SP-B locus. *Dis Markers* 1997; 13: 153–167.
22. Quanjer P. Standardized lung function testing. *Bull Eur Physiopathol Respir* 1983; 19: Suppl. 5, 1–95.
23. Chromosome 2. Master: Gene on sequence map.



- www.ncbi.nlm.nih.gov/cgi-bin/Entrez/efetch?db=Genome&gi=2.
24. Cox DR, Wermuth N. A comment on the coefficient of determination for binary response. *American Statistician* 1992; 46: 1–4.
  25. Saetta M. Airway inflammation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1999; 160: S17–S20.
  26. Lerman C, Caporaso NE, Audrain J, et al. Evidence suggesting the role specific genetic factors in cigarette smoking. *Health Psychol* 1999; 18: 14–20.
  27. Clark JC, Weaver TE, Iwamoto HS, et al. Decreased lung compliance and air trapping in heterozygous SP-B- deficient mice. *Am J Respir Cell Mol Biol* 1997; 16: 46–52.
  28. Frankel WN, Schork NJ. Who's afraid of epistasis? *Nat Genet* 1996; 14: 371–373.
  29. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994; 265: 2037–2048.
  30. Schork NJ. Genetics of complex disease: approaches, problems and solutions. *Am J Respir Crit Care Med* 1997; 156: S103–S109.
  31. Seifart C, Plagens A, Brodje D, Muller B, von Wichert P. Intron 4 SP-B gene variation and association with acute respiratory failure in COPD. *Eur Respir J* 1999; 14: P3509.