ELSEVIER

Biochimica et Biophysica Acta 1408 (1998) 312–322

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

Review

Genetics of the hydrophilic surfactant proteins A and D

Joanna Floros [a,b,*], Russell R. Hoover [a]

[a] *Department of Cellular and Molecular Physiology H166, Pennsylvania State University College of Medicine, P.O. Box 850, 500 University Dr., Hershey, PA 17033, USA*

[b] *Department of Pediatrics, Pennsylvania State University College of Medicine, Hershey, PA, USA*

Received 3 March 1998; received in revised form 27 April 1998; accepted 27 April 1998

Abstract

The use of candidate genes has increased the ability to identify genetic factors involved in diseases with complex and multifactorial etiology. The surfactant proteins (SP) A and D are involved in host defense and inflammatory processes of the lung, which are often components of pulmonary disease. Therefore, the SP-A and SP-D genes make particularly good candidates to study factors contributing to pulmonary disease etiopathogenesis. Moreover, SP-A also plays a role in the surface tension lowering abilities of pulmonary surfactant, which is essential for normal lung function. Although genetic variability at the SP-D locus may exist among humans, allelic variants have not yet been characterized. On the other hand, the human SP-A genes (SP-A1 and SP-A2) are characterized by genetically dependent splice variants at the 5′ untranslated region and allelic variants. The polymorphisms that give rise to SP-A1 and SP-A2 alleles are contained within coding regions, potentially having an effect on protein function. There appears to be a correlation between SP-A genotype and SP-A mRNA content. Furthermore, one SP-A2 allele ($1A^0$) shown to associate with low SP-A mRNA levels is found with higher frequency in a subgroup with respiratory distress syndrome. The evidence gathered thus far indicates that SP-A, possibly by interacting with other surfactant components, may play a role (e.g. be a susceptibility factor) in the development of respiratory disease.

Keywords: Surfactant protein; Collectin; SP-A; SP-D; Allele; Genetics; Mannose binding protein

Contents

* Corresponding author, at address a. Fax: (717) 5317667; E-mail: jxf19@psu.edu

0925-4439/98/$ – see front matter
PII: S0925-4439(98)00077-5

1. Genetic analysis of complex disease

Linkage analysis has been used successfully to elucidate the genetic cause of a number of diseases that are inherited in a classic Mendelian fashion and are the result of a mutation(s) in a single gene (e.g. cystic fibrosis [1] and Huntington's disease [2]). However, it has become increasingly apparent that traditional linkage analysis is not as useful in the identification of the genetic components of diseases with complex and multifactorial etiology (e.g. hypertension, diabetes, or schizophrenia). Complex diseases are most likely the result of multiple interactions between both genetic and non-genetic factors (see Fig. 1). Incomplete penetrance, polygenic inheritance, and locus heterogeneity (reviewed in [3]) are among the reasons that complex diseases often display phenotypic heterogeneity. The underlying complexity often presents limitations for the use of traditional linkage analysis because complex diseases do not exhibit classic Mendelian recessive or dominant inheritance attributable to a single gene locus [4]. Moreover, the large multigenerational families necessary to perform linkage analysis are most times not available.

In recent years, a number of methods have been used to study the genetics of complex diseases including association studies, transmission/disequilibrium tests, affected sib-pair analysis, and studies of twins. Association studies [5] compare the frequency of an allele between two populations: cases (unrelated individuals who have the disease) and controls (unrelated individuals who do not have the disease). If the frequencies of certain alleles for a genetic locus differ significantly between cases and controls, it is said that the particular alleles are associated with the disease. One disadvantage of association studies is the potential for false positives due to population admixture (i.e. the populations under study are not genetically homogeneous). The transmission/disequilibrium test (TDT) can be used to help overcome this limitation.

TDT analysis is based on the notion that a parent, heterozygous for an allele associated with disease and an allele not associated with disease, would transmit the associated allele to the affected offspring more frequently than expected by chance [6]. Advantages of TDT analysis include the use of the non-transmitted parental allele as an internal control and the ability to test for disease alleles that have modest effects. The latter may not be easily detected using other methods [6–8].

Studies of affected sib-pairs and of twins are also useful approaches in the investigation of complex diseases. The major advantage of affected sib-pair (ASP) analysis over traditional pedigree linkage analysis is that the ASP does not require assumptions regarding the mode of disease inheritance. Also, increased allele sharing among affected relatives occurs even in the presence of genetic heterogeneity, incomplete penetrance, phenocopy, and high frequency of disease alleles. However, ASP and twin studies can only detect major gene effects whereas association studies and TDT analysis can identify susceptibility loci with a small or modest contribution to the etiopathogenesis of disease [5,6]. Moreover, TDT and association studies, unlike affected sib-pair or classical linkage analysis, do not require families with multiple affected siblings. Furthermore, an extension of the TDT can be used to address the possibility that each genetic marker allele is associated with the disease to a different extent [8].

The genetic markers used in these various genetic approaches are often microsatellites, nucleotide repeats of variable length that are found throughout

the human genome. Recently, single nucleotide polymorphisms (SNPs) have also been suggested as useful genetic markers [9]. However, the power of TDT analysis and association studies is greatly increased when polymorphisms (e.g. SNPs) within a candidate gene are used [4,10]. A candidate gene is a gene that by way of derangement of either its function or expression, as a result of a mutation or polymorphism, could cause and/or modify the disease under study. It is within this context that the genes encoding the surfactant proteins (SP), particularly SP-A, have become the focus of recent investigations into the genetic causes of respiratory disease [3].

2. Pulmonary surfactant and disease

Pulmonary surfactant is a lipoprotein complex synthesized by the lungs and is essential for normal lung function (reviewed in [11]). Pulmonary surfactant not only reduces the surface tension at the alveolar air liquid interface (thus preventing alveolar collapse) but is also involved in host defense and inflammatory processes of the lung. The importance of surfactant for normal lung function is exemplified by the fact that prematurely born babies deficient in surfactant can develop respiratory distress syndrome (RDS). The etiology of RDS is complex in that a number of factors play a role in its development including gestational age, race, sex, and the hormonal milieu to which the fetus is exposed [3,12]. A genetic contribution to the etiopathogenesis of RDS has been suggested by an increased concordance of the disease between monozygotic twins compared to dizygotic twins [13].

Alteration in pulmonary surfactant function, content, or composition has been associated with a number of other respiratory diseases such as adult respiratory distress syndrome (ARDS), idiopathic pulmonary fibrosis (IPF), pulmonary bacterial and viral infection, and alveolar proteinosis (AP) (reviewed in [14]). Like RDS, these diseases have complex etiologies and probably result from a number of environmental and genetic influences. It is less likely that the genetic components of these diseases would involve genes responsible for the production of surfactant lipids, because their gene products are not lung specific and therefore a change in their activity could result in a broader phenotype (not limited to the lungs). Thus, the lung specific (for the most part) genes encoding the surfactant proteins become good candidate genes for diseases where abnormalities of pulmonary surfactant function occur. The progression of many respiratory diseases involve inflammation and/or host defense. Therefore, the surfactant proteins, namely SP-A and SP-D, known to have immunomodulatory functions [14] are even more likely to play a role in pulmonary disease pathogenesis.

3. SP-A, SP-D, and disease

The hydrophilic proteins of pulmonary surfactant, SP-A and SP-D, are members of the C-type lectin (also known as collectin) family of proteins in that they contain a carbohydrate recognition domain as well as a collagen-like domain [15]. Another member of this protein family is the mannose binding protein (MBP). Collectins are generally involved in the immune process and it seems that SP-A and SP-D are no exception. A number of studies have shown that SP-A is involved in the activation of alveolar macrophages, proliferation of lymphocytes, and the production of inflammatory cytokines [14]. SP-A, by way of its collectin characteristics, binds to bacteria and viruses, leading to their elimination. Indeed, SP-A 'knockout' mice have problems in clearing bacteria from their lungs [16]. SP-D has been shown to interact with a number of the same pathogens as SP-A; however, the exact immunomodulatory functions of the two proteins may be distinct [17]. For example, SP-D, but not SP-A, is able to cause agglutination of a pathogenic yeast (*C. neoformans*) [18] while SP-A, but not SP-D, is able to promote phagocytosis of influenza A virus [19].

Interestingly, the levels of SP-A or SP-D are altered in various diseases. Changes in SP-A levels are seen in such diseases as IPF, *Pneumocystis carinii* pneumonia, RDS, and ARDS [14]. Similarly, levels of SP-D are increased in the serum and bronchoalveolar lavage fluid of patients with IPF, interstitial pneumonia with collagen disease, and pulmonary alveolar proteinosis [20]. These findings suggest a correlation between SP-A and/or SP-D levels with diseases where host defense or inflammatory processes

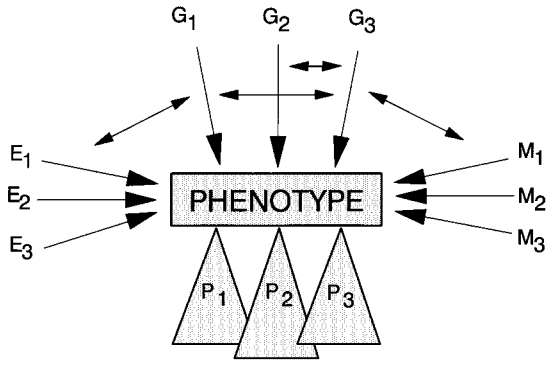

Fig. 1. Schematic representation of genetic and non-genetic interactions in the development of complex disease. Most diseases are multifactorial in that their development is a result of multiple interactions between the environment (E_1, E_2, and E_3) and genetic factors, such as disease susceptibility genes (G_1, G_2, and G_3) and genes that modify disease expression (M_1, M_2, and M_3). Double head arrows indicate interaction. Box represents the entire spectrum of disease phenotype. Triangles represent phenotypic subgroups (P_1, P_2, and P_3) of a given disease, and the overlap of the triangles indicates that distinctions between subgroups may be dependent upon disease severity or stage. (From [3] with permission.)

are involved. Whether the alteration in SP-A or SP-D levels is a contributing factor or merely a consequence of disease is currently unknown.

4. The chromosome 10 collectin genes

The genes encoding SP-A and SP-D have been mapped to human chromosome 10q21-q24 [21–23]. The human SP-A locus [24] contains two very similar but non-identical genes, SP-A1 [25] and SP-A2 [26], as well as a non-functional gene (i.e. a pseudogene) [27]. It is generally thought that two SP-A1 gene products and one SP-A2 gene product associate through their collagen-like domains [28]. Six of these trimers then combine to form the final SP-A multimer that resembles a bouquet of flowers. Baboons [29] are the only other species reported to have two SP-A genes; rats, mice, rabbits, and dogs have only one SP-A gene (that is not entirely identical to SP-A1 or SP-A2). The human SP-D locus, on the other hand, contains a single gene [23].

SP-A1 and SP-A2 have been shown to be in linkage disequilibrium [30], indicating close physical association. This linkage was verified by the physical and radiation hybrid mapping of the SP-A locus (Fig. 2), which revealed that SP-A1 and SP-A2 are separated by about 35–40 kb and are in opposite transcriptional orientation [31]. The SP-A pseudogene was found to be in between the two functional genes and in the same orientation as SP-A2. The fairly close proximity of SP-A1 and SP-A2, and their opposite orientation hold the potential for shared *cis*-acting regulatory elements.

Radiation hybrid mapping also showed that the human SP-D locus is tightly linked to the SP-A locus and lies closer to SP-A2 than to SP-A1 (Fig. 2). SP-D is about 80–100 kb away from SP-A2 and is more proximal (closer to the centromere) than the SP-A locus [31]. The transcriptional orientation of SP-D relative to either of the SP-A genes is currently unknown.

Although the gene encoding another collectin, MBP, has also been mapped to the long arm of chromosome 10 [32], MBP is not linked to either SP-D or SP-A [31]. The MBP locus, however, is linked to a genetic marker (D10S567) about 25 cM (25 000 kb) proximal to SP-D and SP-A [31]. The organization of the SP-A, SP-D, and MBP loci is illustrated in Fig. 2.

Nevertheless, the SP-A, SP-D, and MBP genes share a number of characteristics such as similar genomic structural organization (see Fig. 3) and interrupted glycine codons [23,32]. They also share the

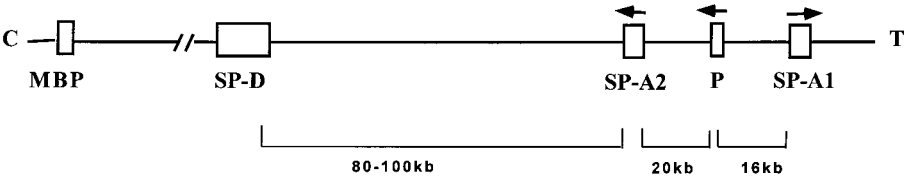

Fig. 2. Organization of the chromosome 10 collectin locus. Physical and radiation hybrid mapping revealed that SP-D, SP-A2, SP-A pseudogene, and SP-A1 are tightly linked. The gene for MBP, although structurally related to SP-D and SP-A, is not linked to either locus as indicated by the break in the line. P represents the SP-A pseudogene, and C and T represent the centromere and telomere, respectively. The arrows indicate transcriptional orientation.

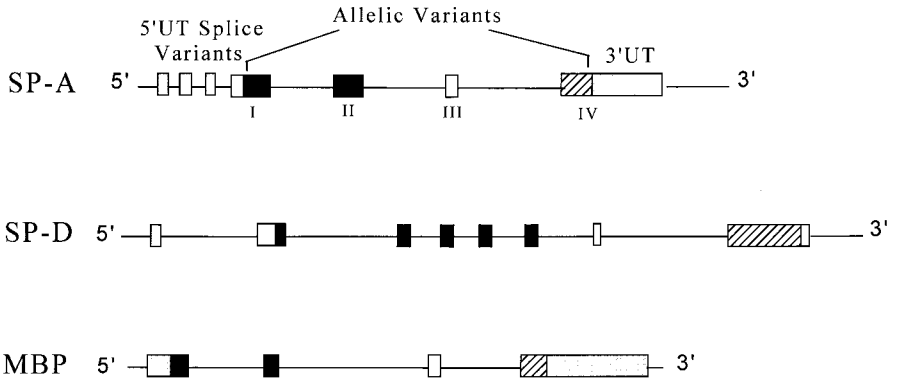

Fig. 3. Genomic structural organization of the collectin genes (SP-A, SP-D, and MBP) located on human chromosome 10. Light gray boxes represent untranslated exons. Black boxes indicate exons encoding the collagen-like domain and hatched boxes indicate exons encoding the carbohydrate recognition domain. SP-A represents either SP-A1 or SP-A2. The nature of the SP-A variability (splice or allelic variants) is noted above the SP-A gene while coding exons (I–IV) are numbered below. SP-A allelic variants are classified based on nucleotide differences within coding regions. Sequence variability within the SP-A 3′ untranslated (3′UT) region has also been observed.

possibility of evolving through the recombination of an ancient non-fibrillar collagen gene and a gene that codes for carbohydrate binding [23,32]. Moreover, a correspondence between spatial distance and phylogenetic tree data appears to exist; for example, SP-D, which is more related to SP-A than MBP in evolution [21], is also physically closer to SP-A than MBP.

The physical linkage of SP-D, SP-A2, SP-A pseudogene, and SP-A1 greatly improves the ability to perform genetic studies, because polymorphisms in any of these loci can be combined and used together in the form of haplotypes.

5. Complexity of the human SP-A genes

As noted in Fig. 3, each functional SP-A gene exhibits extensive complexity. This complexity includes splice variability of 5′ untranslated (5′UT) exons, allelic variability (based on polymorphisms within coding regions), and sequence variability within the 3′UT region. SP-D splice or allelic variants have not yet been described; however, Crouch et. al. [23] have noted differences between SP-D genomic clones and the published cDNA SP-D sequence. Although the possibility for SP-D allelic variation does exist, SP-D alleles have not yet been characterized; thus we focus our attention below on the genetic variability of SP-A.

5.1. SP-A splice variants

Each SP-A gene has a number of 5′UT exons that splice in different configurations giving rise to a number of different SP-A1 and SP-A2 transcripts [33], as shown in Fig. 4A,B (some rare transcripts [33,34] are not shown). The most common splice variant of SP-A1, the AD′ transcript, is observed at a frequency of approx. 0.8 [33]. The most common splice variants of SP-A2, the ABD and ABD′ transcripts, are observed with varied frequency among humans [33,35]. This variation depends on the SP-A2 genotype of the individual, and may be accounted for by a single nucleotide change at the splice recognition site and another one at the extended splice recognition site sequence (Fig. 4C). The former nucleotide change (at the splice recognition site) results in splice differences among SP-A2 alleles with some alleles producing only the ABD′ transcript and others producing both the ABD and ABD′ transcripts. Specifically, at the fourth untranslated exon, SP-A2 alleles $1A^0$ and $1A^1$ have two potential splice sites, D and D′ (Fig. 4C), whereas alleles 1A and $1A^2$ have only one possible splice site, D′ [35]. The second nucleotide change (at the extended splice recognition site) results in splice transcript preference, with the ABD transcripts being preferred over the ABD′ transcripts when both splice sites (D and D′) are present, which is the case for the $1A^0$ and $1A^1$ alleles. This preference leads to an abundance of ABD transcripts and

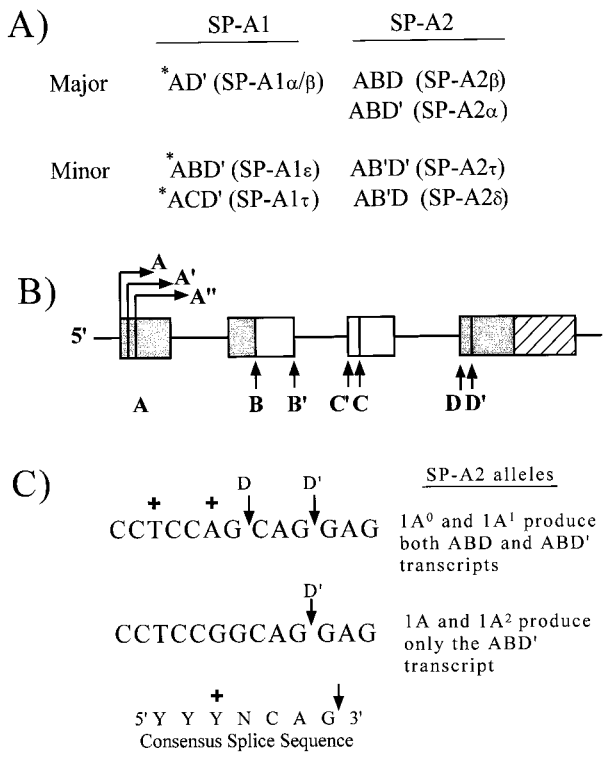

Fig. 4. 5′ splice variants of the SP-A1 and SP-A2 genes. (A) The names of the transcripts arising from alternate splicing of the 5′UT exons are shown. Designations are of Karinch and Floros [33] and in parentheses are those of McCormick et al. [34]. There are three transcription start sites for SP-A1 (see below) and *A represents transcripts derived from any of the three start sites. SP-A2 has only one transcription start site (see below) [33]. (B) Schematic representation of 5′UT exons A–D. Bent arrows show the relative positions of the three SP-A1 transcription start sites (A, A′, and A″). The transcription start site for SP-A2 corresponds to start site A of SP-A1 (plus one additional 5′ nucleotide). The shaded boxes represent exons or portions of exons found in the major SP-A1 or SP-A2 transcripts. Arrows point to the alternate splice sites. Hatched box represents coding exon I. (C) Sequence comparison among SP-A2 alleles shows the presence of two potential splice sites (D and D′) for the $1A^0$ and $1A^1$ alleles and the presence of one splice site (D′) for the 1A and $1A^2$ alleles. The consensus extended splice sequence is also shown. For alleles $1A^0$ and $1A^1$ the fifth nucleotide 5′ of the D′ splice site, which is a purine (A), and the fifth nucleotide (T) 5′ of the D splice site, which is a pyrimidine (T), as well as the fifth nucleotide (Y) of the consensus splice sequence, are marked with a cross (+). Y represents a pyrimidine and N represents any nucleotide. A pyrimidine in the fifth position from the actual splice site is preferred over a purine. This preference may account for the observed high (>1) ABD/ABD′ ratio for the $1A^0$ and $1A^1$ alleles [35].

Gene	Allele	Amino Acid Differences
SP-A1	6A	N A L M D I C P R Q
	$6A^2$	N V V M D I C P R Q
	$6A^3$	N V L M D I C P R Q
	$6A^4$	N V L M D I C P W Q
	$6A^5$	N A L M D I C P W Q
SP-A2	1A	T A V T N V R P R Q
	$1A^0$	N A V T N V R A R Q
	$1A^1$	T A V T N V R A R K
	$1A^2$	T A V T N V R A R Q
	$1A^3$	N A V T N V R A R K
	$1A^4$	T A V T N V R A R K

9 19 50 66 73 81 85 91 219 223

Amino Acid Number

Fig. 5. SP-A1 and SP-A2 alleles. Allelic designations for SP-A1 and SP-A2 are based on differences within coding regions. The positions of the relevant amino acids are at the bottom of the figure. The 'core' amino acids, that distinguish SP-A1 from SP-A2, are boxed. The $6A^5$ and $1A^4$ alleles are described in [40] and the remaining in [30].

may be accounted for by nucleotide differences at the fifth position 5′ (noted with '+' in Fig. 4C) from the D and D′ sites [35]. The consensus sequence for the splice recognition site (shown at the bottom of Fig. 4C) indicates that a pyrimidine, a T (the case for the D site), is preferred over a purine, an A (the case for the D′ site).

The SP-A1 and SP-A2 transcripts are translated both in vitro [33] and in vivo [36], as assessed by the presence of these messages in polysomes. Currently, it is not entirely clear whether the relative efficiency of translation for the various SP-A transcripts is controlled by specific regulatory factors or even whether efficiency of translation is altered under compromised conditions (where presumably an imbalance of key regulatory molecules occurs). Nevertheless, the available data [34,36] suggest that if translation is a control point for SP-A gene regulation it is unclear whether translation would be affected by the characteristics of the splice variants alone. In other words, control of translation may come about by the interaction of specific SP-A splice variants and other factors.

We speculate that variability at the transcript level reflects an example of genetic parsimony (i.e. genetic

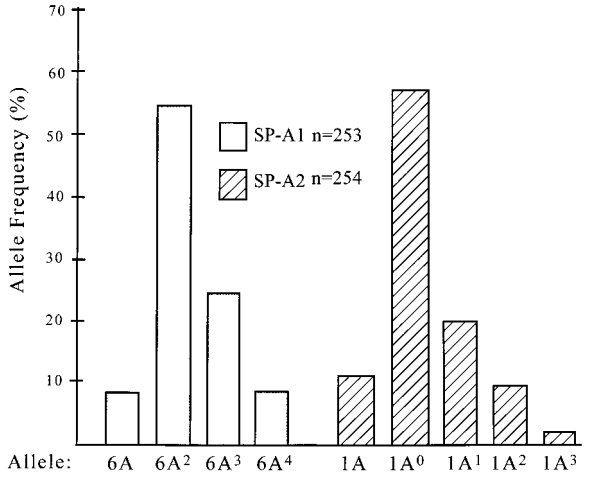

Fig. 6. SP-A1 and SP-A2 allele frequencies in the population. The $6A^2$ and the $1A^0$ alleles are the most frequent SP-A1 and SP-A2 alleles, respectively. Data for $6A^5$ and $1A^4$ are not available, but these alleles are rare and their frequencies are expected to be very low.

economy), where stability at the gene level is maintained but a number of variable forms evolve (i.e. alternatively spliced transcripts). These variable forms can then be subjected to a variety of regulatory mechanisms that can modulate the response to various stimuli and at the same time minimize perturbation of function.

5.2. SP-A alleles

A number of studies have shown extensive variability in SP-A mRNA levels [35,37] and SP-A protein content [38,39] among individuals. It is possible that this variability is a consequence of genetic heterogeneity. Based on the available coding sequences for each SP-A gene, a number of alleles have been characterized [30,40]. To date, five alleles for SP-A1 (6A, $6A^2$, $6A^3$, $6A^4$, and $6A^5$) and six alleles for SP-A2 (1A, $1A^0$, $1A^1$, $1A^2$, $1A^3$, and $1A^4$) have been characterized (see Fig. 5). The alleles of SP-A1 are distinguished from those of SP-A2 by a 'core' of invariant nucleotides or amino acids in coding exon II [33]. Alleles of either gene are distinguished from one another by nucleotide differences throughout the coding region [30,40] that produce changes in the amino acid sequence (Fig. 5). The frequencies of the SP-A alleles in the population are presented in Fig. 6. The most common SP-A1 allele is $6A^2$ and the most common SP-A2 allele is $1A^0$ [30]. Presently, it is unknown whether there are functional differences among the alleles; however, changes in either the collagen-like domain or carbohydrate recognition domain could affect trimerization or host defense functions.

Recently, an association has been made between mRNA content and the most frequent genotype, $6A^26A^21A^01A^0$ [35,40]. This genotype appears to correlate with low to moderate levels of mRNA content, suggesting that either one or both ($1A^0$, $6A^2$) alleles are 'low producers'. Based on the SP-A1 to SP-A2 ratio of mRNA content, the $1A^0$ allele appears to be the one that correlates with low mRNA levels compared to $6A^2$, and probably compared to most (if not all) other SP-A alleles [35]. Of interest, this ($1A^0$) 'low producing' allele is found with higher frequency in a subgroup of RDS [41] (see below), raising the speculation that a prematurely born infant with $1A^0$ genotype is at a higher risk of developing RDS compared to the gestational age-matched counterpart with a different SP-A2 genotype.

6. RDS and surfactant protein polymorphisms

The etiology of RDS is likely to be multifactorial and/or multigenic. Several factors play a role in the development of the disease and evidence exists for a genetic contribution (in at least certain cases) [3,12]. Therefore RDS, like other complex diseases, can result from interactions among disease causing genes, the environment and/or modifier genes (Fig. 1). Different or overlapping sets of interactions are thought to be responsible for (or underlie) each phenotypic RDS subgroup. To identify the genetic contribution in diseases with complex etiology (such as RDS) it is important to study well characterized phenotypic subgroups. Subgrouping (e.g. severity for hypertension or age of onset for Alzheimer's disease), has facilitated the identification of genetic components for other complex diseases [4].

In the context of Fig. 1, and because the surfactant proteins play important roles in surfactant physiology and lung host defense, the hypothesis that the

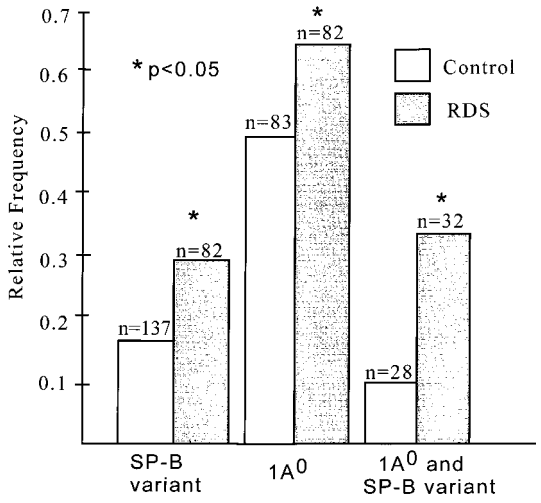

Fig. 7. Association of $1A^0$ with RDS. The frequencies of an SP-B variant and the $1A^0$ allele either singly or in combination in a control population and a subgroup of RDS population are shown. Note the synergistic increased frequency in RDS of the combined polymorphisms. (Adapted from [41] with permission.)

surfactant proteins are contributors to the etiology of RDS is being examined. Association studies have shown that the frequency of a polymorphism in intron 4 of the SP-B gene [42] and the frequency of the $1A^0$ allele of the SP-A2 gene are higher in a subgroup of RDS [41] (Fig. 7). Although an association between an allele of each locus (SP-B or SP-A) and RDS is observed, the frequency of the combined polymorphisms in the RDS population is significantly higher [41] (Fig. 7). More importantly, the nature of the increased frequency of the combined polymorphisms appears to be that of synergism rather than addition, suggesting an interaction between the SP-A and SP-B loci [41]. Interaction between two unlinked loci with gene products that interact functionally (such as SP-A and SP-B) has also been observed in other systems [43].

In light of an observed correlation between levels of SP-A and severity of RDS [44], it is rather intriguing that the SP-A2 allele ($1A^0$), shown to associate with low to moderate mRNA content, is found with higher frequency in RDS (Fig. 8). Moreover, reduced levels of SP-A as well as of SP-B and SP-C have also been noted by immunohistochemistry of lung tissues from babies that died from RDS [45,46].

7. Lessons from the mannose binding protein

As members of the collectin family, SP-A and SP-D share structural and functional similarities with MBP. In this context, it is of interest to note that certain MBP alleles have been shown to associate with diseases involving infection and inflammation. Associations have been found between MBP alleles and systemic lupus erythematosus [47], persistent childhood infections [48], and chronic hepatitis B infection [49]. One MBP allele, MBPD (resulting from a polymorphism at codon 54 that changes a glycine to an aspartate), is dysfunctional [50]. Although the MBPD allele retains the collectin ability to form multimers and bind bacteria, it cannot, unlike the wild type MPB allele, activate the complement cascade [50]. Thus it appears that, within the population, MBP alleles exist that have different functional capabilities; such a situation may also be true for SP-A and/or SP-D.

Genotype/Allele		Observations	Reference
$6A^2 6A^2 1A^0 1A^0$	→	associates with low to moderate SP-A mRNA levels	[35,40]
$6A^2 6A^2 / 1A^0 1A^0$	→	ratio is high to moderate	[35]
$1A^0$ or $1A^0 1A^0$	→	higher frequency in RDS	[41]

Summary: the $1A^0$ allele associates with low levels of SP-A mRNA and appears with higher frequency in the RDS population

Fig. 8. Summary of observations concerning the $1A^0$ allele.

8. Unresolved issues and future perspectives

A significant percentage (>30%) of the genetic loci that have been studied thus far are polymorphic. A locus is said to be polymorphic if the less frequent allele(s) is found in the population at a frequency of at least 0.01 and the heterozygotes with that particular allele occur at a frequency of 0.02. Stable polymorphisms can be used as markers of genetic diversity, and thus they can play a very important role in the identification of genetic factors that may contribute to a particular trait or disease [51].

Furthermore, current wisdom suggests that most (if not all) of the stable polymorphisms are the result of selection for advantageous alleles, although in most cases the selective pressures are currently unknown. In the case of the human SP-A locus, it is clear that both functional SP-A genes are polymorphic. The distribution of SP-A1 and SP-A2 alleles, shown in Fig. 6, indicates that the most frequent alleles are $6A^2$ and $1A^0$ [30]. These alleles have been shown to associate with low to moderate total SP-A mRNA content [35,40], and the $1A^0$ allele is found in higher frequency in a subgroup of the RDS population [41]. If one assumes that the mRNA content reflects protein levels, the observed higher frequency of $6A^2$ and $1A^0$ alleles in the general population is counterintuitive. In other words, the selective advantage for the increased frequency of the 'low or moderately producing alleles' is not readily obvious.

However, SP-A is involved in two major groups of functions. One group of functions involves the surface-tension lowering properties of surfactant. The other group relates to innate host defense and inflammatory processes of the lung. In the former role, SP-A appears to be important only when various cellular factors or other surfactant components that are necessary for optimal surfactant function are not at optimal concentrations [52,53]. In this circumstance one can speculate that the 'low or moderately producing' SP-A alleles are not advantageous for the prematurely born infant. However, such alleles must be advantageous (as judged by their high frequency in the population) in some other role of SP-A. One may argue that this other role is host defense and/or inflammatory processes of the lungs. Conversely, the 'high producing' alleles may be disadvantageous with regards to innate host defense and inflammatory processes of the lung but advantageous for the surface-tension lowering properties of surfactant.

One good example of a stable or balanced polymorphism is the allele(s) of hemoglobin responsible for the sickle trait. Homozygous individuals for the 'sickle' allele are affected by a severe hemolytic anemia and usually die before they reach reproductive age whereas heterozygous individuals (carriers) are more resistant to malaria. The 'sickle' allele is maintained in the population and is found in high frequency in parts of the world where malaria infection is common. In other words, the small selective advantage (increase of biological fitness) of heterozygotes has outweighed the major disadvantage (death) of homozygotes and has thus provided the necessary pressure to maintain the 'sickle' allele in high frequency in malaria affected regions.

Therefore, it is possible that the high frequency of the 'low/moderately producing' SP-A alleles reflect a selective advantage as a result of an increased biological fitness. To assess the validity of such speculation, and to explore others, it will be necessary to develop reagents in order to measure each SP-A gene product individually and to determine whether mRNA levels reflect protein levels. We need to investigate what happens when an imbalance occurs with regard to the content of each SP-A gene product. We will also need to study the regulation and the function of each allele of each SP-A gene. By doing so we will gain insight into the potential role of SP-A alleles in the susceptibility of various diseases. We also need to understand how the products of a particular set of alleles at two separate unlinked loci (e.g. SP-A and SP-B) can interact to produce an unexpected result with regard to a given phenotype (an example of epistasis). Finally, we need to understand the complementary and overlapping roles of SP-A and SP-D in innate host defense and the potential contribution (if any) of SP-A and SP-D alleles to the susceptibility of disease.

9. Final thoughts

The benefit derived from the cumulative knowledge of multidisciplinary approaches will reach new heights regarding the hydrophilic surfactant proteins.

Our thinking and problem solving capacities will be pleasantly challenged as never before. As parts of the puzzle become clearer, unanticipated paths and questions will most undoubtedly come to view and refuel our scientific curiosity.

Acknowledgements

This work was supported by NIH R37 HL34788 and HL49823.

References

[1] A. Beaudet, A. Bowcock, M. Buchwald, L. Cavalli-Sforza, M. Farrall, M.C. King, K. Klinger, Am. J. Hum. Genet. 39 (1986) 681–693.
[2] J.F. Gusella, N.S. Wexler, P.M. Conneally, S.L. Naylor, M.A. Anderson, R.E. Tanzi, P.C. Watkins, K. Ottina, M.R. Wallace, A.Y. Sakaaguchi, A.B. Young, I. Shoulson, E. Bonilla, J.B. Martin, Nature 306 (1983) 234–238.
[3] J. Floros, P. Kala, Annu. Rev. Physiol. 60 (1998) 365–384.
[4] E.S. Lander, N.J. Schork, Science 265 (1994) 2037–2048.
[5] N. Risch, K. Merikangas, Science 273 (1996) 1516–1517.
[6] R.S. Spielman, R.E. McGinnis, W.J. Ewens, Am. J. Hum. Genet. 52 (1993) 506–516.
[7] R.S. Spielman, R.E. McGinnis, W.J. Ewens, Am. J. Hum. Genet. 54 (1994) 559–560.
[8] P.C. Sham, D. Curtis, Ann. Hum. Genet. 59 (1995) 323–336.
[9] F.S. Collins, M.S. Guyer, A. Chakravarti, Science 278 (1997) 1580–1581.
[10] D.A. Greenberg, Am. J. Hum. Genet. 52 (1993) 135–143.
[11] J. Floros, D.S. Phelps, in: T.L. Yaksh et al. (Eds.), Anesthesia: Biologic Foundations, Lippincott-Raven, Philadelphia, PA, 1997, pp. 1259–1280.
[12] J. Floros, A. Karinch, in: B. Robertson, H.W. Taeusch (Eds.), Surfactant Therapy for Lung Disease, Marcel Dekker, New York, 1995, pp. 95–106.
[13] N.C. Myrianthopoulos, J.A. Churchill, A.J. Baszynski, Acta Genet. Med. Gemellol. 20 (1971) 199–204.
[14] D.S. Phelps, Appl. Cardiopulm. Pathophysiol. 5 (1995) 221–229.
[15] K. Reid, Biochem. Soc. Trans. 21 (1993) 464–468.
[16] A.M. LeVine, M.D. Bruno, K.M. Huelsman, G.F. Ross, J.A. Whitsett, T.R. Korfhagen, J. Immunol. 158 (1997) 4336–4340.
[17] K.L. Hartshorn, M.R. White, V. Shepherd, K. Reid, J.C. Jensenius, E.C. Crouch, Am. J. Physiol. 273 (1997) L1156–L1166.
[18] S. Schelenz, R. Malhotra, R.B. Sim, U. Holmskov, G.J. Bancroft, Infect. Immun. 63 (1995) 3360–3366.
[19] C.A. Benne, B. Benaissa-Trouw, J.A.G. van Strijp, C.A. Kraaijeveld, J.R.F. van Iwaarden, Eur. J. Immunol. 27 (1997) 886–890.
[20] Y. Honda, Y. Kuroki, E. Matsuura, H. Nagae, H. Takahashi, T. Akino, S. Abe, Am. J. Respir. Crit. Care Med. 152 (1995) 1860–1866.
[21] K. Kölble, J. Lu, S.E. Mole, S. Kaluz, K.B.M. Reid, Genomics 17 (1993) 294–298.
[22] G. Bruns, H. Stroh, G.M. Velman, S.A. Latt, J. Floros, Hum. Genet. 76 (1987) 58–62.
[23] E. Crouch, K. Rust, R. Veile, H. Donis-Keller, L. Grosso, J. Biol. Chem. 268 (1993) 2976–2983.
[24] J. Floros, A.M. Karinch, Am. J. Physiol. 268 (1995) L162–L165.
[25] R.T. White, D. Damm, J. Miller, K. Spratt, J. Schilling, S. Hawgood, B. Benson, B. Cordell, Nature 317 (1985) 361–363.
[26] S.L. Katyal, G. Singh, J. Locker, Am. J. Respir. Cell. Mol. Biol. 6 (1992) 446–452.
[27] T.R. Korfhagen, S.W. Glasser, M.D. Bruno, M.J. McMahan, J.A. Whitsett, Am. J. Respir. Cell. Mol. Biol. 4 (1991) 463–469.
[28] T. Voss, K. Melchers, G. Scheirle, K.P. Schäfer, Am. J. Respir. Cell. Mol. Biol. 4 (1991) 88–94.
[29] E. Gao, Y. Wang, S.M. McCormick, J. Li, S.R. Seider, C.R. Mendelson, Am. J. Physiol. 271 (1996) L617–L630.
[30] J. Floros, S. DiAngelo, M. Koptides, A.M. Karinch, P. Rogan, H. Nielson, R. Spragg, K. Watterberg, G. Dieter, Am. J. Respir. Cell. Mol. Biol. 15 (1996) 489–498.
[31] R.R. Hoover, J. Floros, Am. J. Respir. Cell. Mol. Biol. 18 (1998) 353–362.
[32] K. Sastry, G.A. Herman, L. Day, E. Deignan, G. Bruns, C.C. Morton, R.A.B. Ezekowitz, J. Exp. Med. 170 (1989) 1175–1189.
[33] A.M. Karinch, J. Floros, Am. J. Respir. Cell. Mol. Biol. 12 (1995) 77–88.
[34] S.M. McCormick, V. Boggaram, C.R. Mendelson, Am. J. Physiol. 266 (1994) L354–L366.
[35] A.M. Karinch, J. Floros, Biochem. J. 321 (1997) 39–47.
[36] A.M. Karinch, J. Floros, Biochem. J. 307 (1995) 327–330.
[37] J. Floros, D.S. Phelps, D.E. deMello, J. Longmate, H. Harding, B. Benson, T. White, Exp. Lung Res. 17 (1991) 91–104.
[38] I.R. Doyle, M.E. Jones, H.A. Barr, S. Orgeig, A.J. Crockett, C.F. McDonald, T.E. Nicholas, Am. J. Respir. Crit. Care Med. 149 (1994) 1619–1627.
[39] H. Takahashi, Y. Honda, Y. Kuroki, K. Imai, T. Akino, S. Abe, Clin. Chim. Acta 239 (1995) 213–215.
[40] A.M. Karinch, G. Deiter, P.L. Ballard, J. Floros, Biochim. Biophys. Acta 1398 (1998) 192–202.
[41] P. Kala, T.T. Have, H. Nielson, M. Dunn, J. Floros, Pediatr. Res. 43 (1998) 169–177.
[42] J. Floros, S.V. Veletza, P. Kotikalapudi, L. Krizkova, A.M. Karinch, C. Friedman, S. Buchter, K. Marks, Biochem. J. 305 (1995) 583–590.
[43] K. Kajiwara, E.L. Berson, T.P. Dryja, Science 264 (1994) 1604–1607.

[44] F.R. Moya, H.F. Montes, V.L. Thomas, A.M. Mouzinho, J.F. Smith, C.R. Rosenfeld, Am. J. Respir. Crit. Care Med. 150 (1994) 1672–1677.
[45] D.E. deMello, D.S. Phelps, G. Patel, J. Floros, D. Lagunoff, Am. J. Pathol. 134 (1989) 1285–1293.
[46] D.E. deMello, S. Heyman, D.S. Phelps, J. Floros, Am. J. Pathol. 142 (1993) 1631–1640.
[47] K.E. Sullivan, C. Wooten, D. Goldman, M. Petri, Arthritis Rheum. 39 (1996) 2046–2051.
[48] J.A. Summerfield, M. Sumiya, M. Levin, M.W. Turner, Br. Med. J. 314 (1997) 1229–1232.
[49] H.C. Thomas, G.R. Foster, M. Sumiya, D. McIntosh, D.L. Jack, M.W. Turner, J.A. Summerfield, Lancet 348 (1996) 1417–1419.
[50] M. Matsushita, R.A. Ezekowitz, T. Fujita, Biochem. J. 311 (1995) 1021–1023.
[51] T.D. Gelehrter, F.S. Collins, Principles of Medical Genetics, Williams and Wilkins, Baltimore, MD, 1990.
[52] T.R. Korfhagen, M.D. Bruno, G.F. Ross, K.M. Huelsman, M. Ikegami, A.H. Jobe, S.E. Wert, B.R. Stripp, R.E. Morris, S.W. Glasser, C.J. Bachurski, H.S. Iwamoto, J.A. Whitsett, Proc. Natl. Acad. Sci. USA 93 (1996) 9594–9599.
[53] S. Schürch, F. Possmayer, S. Cheng, A.M. Cockshutt, Am. J. Physiol. 263 (1992) L210–L218.