Protein composition of bronchoalveolar lavage fluid and airway surface liquid from
newborn pigs

Jennifer A. Bartlett¹, Matthew E. Albertolle², Christine Wohlford-Lenane¹, Alejandro A. Pezullo³, Joseph Zabner³, Richard K. Niles², Susan J. Fisher², Paul B. McCray Jr.¹, Katherine E. Williams²

¹Department of Pediatrics, ³Department of Internal Medicine, Carver College of Medicine, University of Iowa, Iowa City, IA 52242
²Department of Obstetrics, Gynecology and Reproductive Sciences and, University of California, San Francisco, CA 94143-0665

Author contributions
Experimental design: J.A.B., J.Z., P.B.M., S.J.F., K.E.W.
Data analysis: R.K.N., M.E.A., K.E.W.

Running title: Proteomics of porcine lung fluid

Correspondence may be addressed to:
Katherine Williams
521 Parnassus Avenue, C-18
San Francisco, CA, 94143-0665
415-476-2330
Email: katherine.williams@ucsf.edu
Abstract

The airway mucosa and the alveolar surface form dynamic interfaces between the lung and the external environment. The epithelial cells lining these barriers elaborate a thin liquid layer containing secreted peptides and proteins that contribute to host defense and other functions. The goal of this study was to develop and apply methods to define the proteome of porcine lung lining liquid, in part, by leveraging the wealth of information in the *Sus scrofa* database of Ensembl gene, transcript, and protein model predictions. We developed an optimized workflow for detection of secreted proteins in porcine bronchoalveolar lavage (BAL) fluid and in methacholine-induced tracheal secretions (airway surface liquid, ASL). We detected 674 and 3,858 unique porcine-specific proteins in BAL and ASL, respectively. This proteome was composed of proteins representing a diverse range of molecular classes and biological processes, including host defense, molecular transport, cell communication, cytoskeletal, and metabolic functions. Specifically, we detected a significant number of secreted proteins with known or predicted roles in innate and adaptive immunity, microbial killing, or other aspects of host defense. In greatly expanding the known proteome of the lung lining fluid in the pig, this study provides a valuable resource for future studies using this important animal model of pulmonary physiology and disease.

Keywords: proteomics, airway surface liquid, bronchoalveolar lavage, porcine lung
Introduction

The passageways of the mammalian respiratory tract are bathed by secretions that play critically important roles in protecting the organism from microbial and environmental insults and promoting normal lung function. In the conducting airways, this fluid layer is known as airway surface liquid (ASL). ASL, a complex mixture of secreted proteins, peptides, and mucins as well as electrolytes and water, represents the combined contributions of the surface and submucosal gland epithelia of the conducting airways (5, 85). The varied roles of these gene products include modulating inflammation, promoting wound healing, maintaining ASL volume homeostasis, and the transport of various nutrients and lipids in the extracellular milieu. In particular, ASL possesses a redundant and polyfunctional array of secreted peptides and proteins with host defense and immunomodulatory properties (reviewed in (7)). In the gas exchange regions of the lung, epithelia secrete a liquid known as the alveolar subphase, which contains pulmonary surfactant, a potent mixture of lipids and proteins necessary for proper mechanical functioning of the lung. In addition, like the ASL of the conducting airways, alveolar secretions include a number of innate immune factors involved in pulmonary host defenses. The coordinated functions of these secreted products help maintain lung health.

To better understand the functions of these compartments, proteomics approaches have been used to help identify the secreted proteins that comprise the ASL and alveolar subphase (collectively termed lung lining liquid). In past studies, researchers have used two dimensional gel electrophoresis or other fractionation methods coupled with mass spectrometry (MS) to investigate lung lining liquid composition in a variety of fluids sampled from the human respiratory tract: nasal lavage fluid (49, 50), bronchoalveolar lavage (BAL) fluid (14, 28, 44, 47, 48, 50, 53, 58, 83, 84, 87), induced sputum (38, 60), and apical washings from cultured human bronchial epithelia (38). A serous cell model of airway submucosal glands using apical fluid from Calu-3 cells was used to investigate the regulation of antiprotease and antimicrobial secreted proteins (37). Notably, these proteomic tools provide a means to assess global changes in the composition of respiratory secretions in response to various inflammatory stimuli (12, 47-50) or in association with certain disease states (44, 53, 58, 83, 87).
Proteomic techniques have also been used to identify the BAL fluid proteins from wild-type mice (28, 31) and in murine models of lung injury and disease (13, 17, 28, 89). Here, we used a shotgun MS approach to define the repertoire of secreted peptide and protein components of the porcine lung.

The pig offers several advantages as an animal model for studies of lung biology and disease states. Porcine lungs share many similarities with human lungs in terms of size and structure (71). Also, unlike mice, the relatively long lifespan of pigs (10-20 years on average) makes them well suited for studies of progressive lung diseases (71). Porcine lungs have previously been used in many areas of lung research, including studies of surfactant composition and function (76), lung development (29), infectious diseases including influenza and porcine specific infections (27, 51, 61, 65), lung transplantation (46, 52, 59), responses to various types of lung injury (11, 32), and testing of gene therapy vectors (20, 57). Additionally, pigs have been used to model a number of lung diseases and conditions, including pulmonary hypertension (9), bronchiolitis obliterans (2), asthma (82), and cystic fibrosis (CF) (15, 43, 71, 77). Notably, the CF pig model has provided insights into the development of CF lung disease, an aspect that was previously difficult or impossible to study using murine CF models (15, 62, 66, 77).

In contrast to the human and the mouse, the lung lining liquid composition of pig airways is poorly characterized. The components of the ASL and BAL in the normal pig lung are largely unknown and it is likely that there are hundreds of proteins and peptides that have never been identified. Proteomic analysis of BAL and ASL has posed formidable challenges due to variable protein composition and small sample volumes, which make it difficult to identify relevant biomarkers against the background of intracellular proteins from cell shedding and other processes (28, 31). Additionally, the number of well annotated porcine protein sequences lags far behind the available genomic information. In this study, a global proteomic analysis to identify the proteins in wild-type porcine ASL and BAL was performed to establish a foundation for future research. To create a useful pig-specific protein database for MS data analysis, we annotated Ensembl *Sus scrofa* FASTA
Proteomics of porcine lung fluid protein entries with protein and gene names. These data provided a comprehensive profile of lining liquid components in healthy lung and new insights into the biology of this important animal model. This repository is an important resource for future comparative studies of the alterations in secreted factors that may occur in association with CF and other porcine models of pulmonary disease states.
Materials and Methods

Animal protocols and collection of bronchoalveolar lavage (BAL) and airway surface liquid (ASL)

Samples were collected from wild-type pigs as previously described (62, 71, 72, 77). All experimental techniques were approved by the Institutional Animal Care and Use Committee of the University of Iowa.

For BAL collection, 6 newborn pigs were euthanized within 12 hours of birth by administering Euthasol (90 mg/kg, IV) and lungs excised using aseptic technique. To lavage, 1/16 inch diameter sterile polyethylene tubing was inserted into the mainstem bronchi and lungs were washed with 5 mL normal saline. This procedure was repeated 3 times for each excised lung and the collected washes from an individual animal were immediately pooled and placed on ice. Then each pooled BAL was centrifuged at a low speed (228 x g) and the supernatant transferred to a fresh tube. Clarified BAL was buffer exchanged against 100 mM tetraethylammonium bicarbonate (TEAB; Sigma-Aldrich, St. Louis, MO, USA) using Amicon Ultra-15 3 kDa molecular weight cut-off filters (Millipore, Billerica, MA, USA). Total protein concentrations were estimated by the Bradford assay. Samples were stored frozen at -80°C until use.

Porcine ASL was collected from 5 wild-type pigs within 12 hours of birth. For this procedure, we used an established bronchoscopic microprobe method (23) to collect native secretions from the trachea and first generation bronchi. Methacholine was used to stimulate secretion after initial studies revealed that it was not possible to collect secretions from newborn animals without stimulation (66). We reasoned that this method would allow us to collect the net contributions of airway and submucosal gland cells in response to a strong neurochemical stimulus. Pigs were anesthetized with a mixture of ketamine (20 mg/kg, IM) and xylazine (2 mg/kg, IM), followed by propofol (2 mg/kg, IV); saline was given intravenously to prevent dehydration. Once an animal had reached the proper plane of anesthesia, the neck was dissected to expose the trachea. Tracheal secretion was stimulated by administering methacholine (2.5 mg/kg, IV). After approximately 5 minutes, tracheal secretions were collected by making a small incision in the tracheal wall and inserting a sterile...
polyester tipped applicator (Puritan Medical Products Co., Guilford, ME, USA) to swab the lumen of the trachea. Then the probe was inserted into a microcentrifuge tube and secretions were recovered by centrifugation. This procedure generally resulted in recovery of approximately 10-20 µL of ASL from each animal. Samples were immediately placed on ice and frozen at -80°C until use. Following ASL collection, pigs were euthanized with Euthasol.

Protein preparation and mass spectrometry

Bronchoalveolar lavage. The workflow for BAL preparation and processing is shown in Figure 1. Upon thawing, 100 ug of protein was denatured with trifluoroethanol (TFE), trypsin digested (81), divided into 2 aliquots, and fractionated by isoelectric focusing (16, 25, 34) and alkaline reverse phase high pressure liquid chromatography (RP HPLC) (21, 22, 24). For BAL1 and BAL2, 8 HPLC fractions were collected and for BAL3-6 a total of 30 fractions were collected. Peptide fractions were analyzed by LCMS on a QSTAR Elite mass spectrometer (AB Sciex, Foster City, CA, USA).

Airway surface liquid. The workflow for ASL preparation and processing is shown in Figure 1. ASL proteins were solubilized in sodium dodecyl sulfate (SDS), reduced, alkylated, and trypsin digested using either a filter assisted solubilization protocol (FASP), or by in-solution digestion followed by SDS removal using strong cation exchange (SCX) (68, 75). Peptides were fractionated offline (n=20) by using alkaline RP HPLC followed by LCMS on an LTQ Velos Orbitrap (Thermo Scientific, San Jose, CA, USA).

Pig protein sequence database development and protein identification

The Ensembl Sus scrofa 10.2.67.pep.all protein FASTA database, containing 23,118 entries, was annotated with protein and gene names as follows. First, a program was developed to query all Ensembl entries for each protein accession code. The gene name, description, database source (e.g., UniProt, NCBI, HGNC), and entry name, if present, were parsed out and assembled to replace the original Ensembl annotation. For those entries for which
the description was “uncharacterized protein” or “novel transcript”, the gene name, if present, was used to search the human Uniprot Knowledgebase v2012_07 and the human protein description used. The source for these entries is designated UniprotKB(Hu). The final database contained protein sequences and Ensembl accession codes for all of the original 23,118 entries with 18,664 entries fully annotated with descriptive protein names.

Protein identification was accomplished by using ProteinPilot 4.0 software (AB Sciex) and the integrated false discovery rate (FDR) analysis function (79) with a concatenated reversed database. Search parameters were trypsin enzyme specificity, carbamidomethyl cysteine, and thorough search effort. Proteins with $\leq 5\%$ local FDR and peptides with $\leq 1\%$ global FDR were reported. For pig Ensembl entries that did not contain a protein name, the gene name was mapped to the human protein name. For the novel transcripts and uncharacterized proteins lacking a gene name that were detected at an FDR threshold of $\leq 5\%$, a sequence similarity search was performed using BLAST (4) and the protein with the highest score was reported. If equivalent top-scoring BLAST matches occurred, the human match was reported whenever present. A subset of the data was also searched using mammalian sequences in the Uniprot SwissProt database. For both BAL and ASL, proteins detected from each individual sample were aligned to a master search result comprised of all data using the Protein Alignment Template V2.000p beta (78). The master search was a reference protein identification list produced by searching the MS data from all samples to produce a single result. To perform the analysis of the intersection of protein identifications, the threshold for the master search was set at 1% global FDR and the threshold for the individual samples set to 5% local FDR. These settings were chosen to ensure that high quality identifications from each set were matched. The annotation of protein molecular function and biological processes was performed using PANTHER Gene Ontology (GO) (80).

SDS-PAGE and immunoblotting
To visualize proteins in lung lining fluid, BAL and ASL samples (2 µg total protein per lane) were electrophoresed through 4-20% Tris-HCl gradient gels (Bio-Rad Laboratories, Hercules, CA, USA) and silver stained using the Silver Stain Plus kit (Bio-Rad Laboratories). To immunoblot for PLUNC and SP-D, total protein from pig BAL and ASL was separated on 4-20% Tris-HCl gels (20 µg/lane for BAL and 5 µg/lane for ASL) and transferred to PVDF membranes, followed by blocking overnight in TBS-Tween containing 2% BSA.

To detect PLUNC protein, membranes were incubated with a monoclonal antibody recognizing human and porcine palate, lung, nasal epithelium clone (PLUNC; R&D Systems, Minneapolis, MN, USA) diluted 1:250 in TBS-Tween, for 1.5 hours at room temperature. Membranes were washed 4 times using TBS-Tween, then incubated with secondary antibody (Immunopure goat anti-mouse conjugated to horseradish peroxidase; Thermo Fisher Scientific) at a 1:20,000 dilution for 1 hour. After 5 more washes in TBS-Tween, protein bands were detected using SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific). For SP-D detection, membranes were incubated with anti-porcine SP-D diluted 1:500 in TBS-Tween for 1.5 hours at room temperature, followed by 4 washes in TBS-Tween. Membranes were incubated with secondary antibody (Immunopure goat anti-rabbit conjugated to horseradish peroxidase; Thermo Fisher Scientific) at 1:20,000 for 1 hour, followed by 5 washes in TBS-Tween and chemiluminescent detection as described above.

Immunoblotting for MUC5AC was carried out under nonreducing conditions as described by Lacunza and colleagues (45). Briefly, BAL and ASL samples were diluted in SDS-PAGE sample buffer containing 25% SDS (125 mM Tris-HCl, pH 6.8; 25% SDS; 20% glycerol; 0.004% bromophenol blue) prior to boiling at 95°C. Samples were loaded onto 7.5% Tris-HCl gels (5 µg/lane) and electrophoresed for approximately 2 hours. Then proteins were transferred to PVDF membranes by electroblotting overnight at 4°C, by using methanol-free transfer buffer containing 0.1% SDS. To block nonspecific immunoreactivity, the membranes were incubated for 1-2 hours at room temperature in TBS-Tween containing 2% BSA. Then they were incubated for 1.5 hours with a monoclonal antibody recognizing MUC5AC (Clone 45M1; Thermo Fisher Scientific) diluted 1:500 in TBS-Tween. Next they were washed 4 times in TBS-Tween, and incubated for 1 hour with secondary antibody (Immunopure goat anti-mouse conjugated to horseradish peroxidase; Thermo Fisher Scientific) at a 1:20,000 dilution.
dilution. After 5 washes in TBS-Tween, bands were detected as described above for PLUNC. Electrophoresis and immunoblotting of MUC5AC under these conditions is expected to produce a predominant band of approximately 200 kDa (45).
Results

Diversity of proteins in BAL and ASL.

To investigate the composition of the lung lining fluid in the newborn pig, we sampled the alveolar subphase and the conducting airways by BAL as well as by direct collection of ASL from the trachea. The overall abundance and complexity of proteins in the BAL and ASL samples was assessed by separating proteins using SDS-PAGE and visualizing by silver stain. As shown in Figure 2, both the BAL and ASL samples displayed a diverse complement of proteins across a wide range of molecular weights. Total protein concentrations were 0.050-0.42 mg/mL in newborn pig BAL samples and 5-15 mg/mL in ASL.

Shotgun proteomics detection of BAL proteins.

To better understand the composition of BAL, 6 samples were processed for LCMS analysis (as summarized in Figure 1) using trifluoroethanol (TFE) for denaturation prior to trypsin digestion. To gain maximum coverage of the proteome, peptide mixtures were fractionated offline using two orthogonal methods, isoelectric focusing and alkaline pH reversed phase HPLC, prior to LCMS data acquisition. For the first 2 samples, a total of eight offline alkaline pH HPLC fractions were collected. The remaining BAL samples were more extensively fractionated for a total of 30 fractions. To maximize protein identification, spectra were searched against the pig Ensembl pep.all database, the super-set of all translations resulting from Ensembl known or novel gene predictions. The general format of these entries was “ID Sequence type:Status:Location:Gene:Transcript”. To make this database useful for protein identification, we developed software to configure the header information for each protein to include the protein name, the gene name, and the source database. The annotated database is available for download at https://wiki.library.ucsf.edu/x/vSzWAw.

The number of proteins detected at \( \leq 5\% \) FDR in individual BAL samples ranged widely, from 46-558, with a total of 674 distinct proteins identified from all BAL samples (Table 1). More extensive fractionation at the peptide level prior to LCMS resulted in a significant increase in the number of peptides and proteins detected.
Proteomics of porcine lung fluid (BAL 1 and 2 vs. BAL 3-6). Over 89% of the proteins identified in the Ensembl database were mapped to porcine species-specific entries in Swiss-Prot, RefSeq, or TrEMBL (Supplemental Table 1). A sequence similarity search using BLAST (4) was performed for the 70 confidently detected sequences listed as “uncharacterized” or “novel transcript”, containing neither a protein name nor a gene name. For all but 2 proteins, very high scoring homologous sequences were identified (Supplemental Table 2).

As expected, the BAL dataset included many well recognized lung gene products, such as α-1-antitrypsin, the alveolar surfactant proteins (surfactant proteins B [SP-B] and C [SP-C]), and mucins such as MUC1 and MUC16 (Supplementary Table 1). Of interest, we identified numerous proteins associated with mammalian host defense, including lactoferrin; lipocalin 2; the collectins; SP-A and SP-D; the PLUNC protein and related family members; lactoperoxidase; S100-A8 (calgranulin-A) and S100-A9 (calgranulin-B); histones; and several components of the complement system. Additionally, we identified the pig-specific cathelicidin antimicrobial protein protegrin-1. Other proteins expected to be found in BAL, e.g., PR-39, PMAP-37, and the highly abundant MUC5AC and MUC5B, were not present in the Ensembl database. To verify the presence of these abundant BAL proteins in our samples, we searched the Uniprot pig database, which contained full-length sequences for PR-39 and PMAP-37 and a partial sequence of MUC5AC (MUC5B was not in this database). PMAP-37 and MUC5AC were detected with multiple high scoring peptides (Table 2). The relatively small size and high arginine content (10 of 39 amino acids) of PR-39 limited the number of observable tryptic peptides to one, which could explain why this abundant antimicrobial protein was not detected.

To independently verify our results, we selected a subset of BAL proteins (PLUNC, SP-D, and MUC5AC) to confirm by immunoblotting (Figure 3). These proteins were selected for their known roles in the lung lining liquid milieu and the availability of cross-reacting antibodies. All BAL samples contained these proteins as determined by immunoblotting.
ASL protein analysis

ASL from the conducting airways was harvested using a modified collection approach that was designed to avoid sampling the alveolar contribution. Based on experience gained with the BAL analysis, we refined our workflow as outlined in Figure 1, using SDS as a denaturant with trypsin digestion in-solution or on filters (55, 86). The number of distinct peptides identified was 6.5-fold higher with the in-solution trypsin digestion protocol, averaging 18,872, compared to 2,888 with the filter-assisted digestion protocol. This translated to an average of 2,407 protein groups as compared to 969, respectively (Table 1). The presence of highly abundant glycosylated, difficult to proteolyze proteins in ASL could potentially clog the filter and reduce peptide recovery. Overall, the ASL analyses resulted in a significantly larger number of unique protein identifications as compared with BAL, with 3,858 distinct proteins detected in toto (Supplemental Table 3). Confidently detected Ensembl sequences lacking a protein or gene name were searched using BLAST as described for BAL proteins (Supplemental Table 4). Additional lung and host defense proteins detected in ASL included MUC-4, -13, and -19, LPLUNC2, lipocalin 1, antileukoproteinase (SLPI), PMAP-23, PLAT, protegrin 3, SPPI, and serpinB5. A subset of proteins (PLUNC, SP-D, and MUC5AC) was confirmed by immunoblotting and all ASL samples contained these proteins (Figure 3B).

Functional classification of BAL and ASL proteins.

We observed significant overlap between the BAL and ASL proteomes; 88% of BAL-associated proteins were also detected in ASL. Using PANTHER GO classifications, proteins in both datasets were categorized by biological process, protein class, molecular function, and cellular compartment (Figure 4). The distribution of proteins was similar between BAL and ASL, and included a broad range of protein classes associated with diverse processes such as responses to stimuli, immune system functions, cell communication, and transport. In addition to extracellular and secreted proteins, a number of intracellular proteins such as metabolic enzymes were identified.
We previously profiled mRNA expression in trachea, bronchus tissue and well-differentiated primary cultures of tracheal and bronchial epithelia isolated from wild-type and cystic fibrosis transmembrane conductance regulator (CFTR) null newborn piglets (66). A set of 973 genes (secreted host defense proteins as well as those involved in inflammation, bacterial responses and wounding) was compared to the proteins detected in BAL and ASL. A total of 314 proteins were found, 87 of which were secreted host defense proteins (Table 3).
**Discussion**

Here we developed techniques to identify proteins from the porcine lung and conducting airways by using BAL fluid and ASL as starting materials. To our knowledge, this study provides the most comprehensive database to date of proteins in porcine lung lining fluid. Because BAL has traditionally been the method of choice for sampling lung and airway secretions for proteomic studies, we began by characterizing BAL from newborn pigs. While earlier studies used proteomics-based approaches to monitor pig BAL for changes in response to bacterial (33) and viral (88) pathogens, ours is the first attempt to compile the proteome of normal pig BAL, which resulted in the identification of 674 proteins. This level of detection was in line with previous reports that used mass spectrometry-based methods to identify BAL proteins in human (28, 44, 53, 58, 87), mouse (13, 17, 28, 31, 67), rat (74), cow (8), and horse (10). For example, Guo and colleagues identified 297 proteins in BAL from a single mouse (31), and analysis of BAL from 6 healthy horses yielded 582 identifications (10). Notably, protein discovery has often been greatest in studies designed to investigate differential protein expression in the context of various pulmonary diseases or challenges. To date, the greatest number of BAL proteins reported for any species was 959, observed in a study investigating the effects of cigarette smoke inhalation in mice (67). In a study of biomarkers in lung transplant patients with chronic graft dysfunction (CGD), Kosanam et al. reported a total of 531 proteins in human BAL (44). In another study, as many as 889 proteins were identified in BAL from antigen-challenged asthmatic patients and control individuals (87).

One drawback of using BAL fluid as the analyte for proteomics is that the collection technique necessarily involves significant and variable dilution, potentially complicating efforts to make quantitative comparisons among study subjects. As an alternative, several groups attempted to characterize lung lining fluid in humans using induced sputum, which can be collected without dilution of the sample (38, 60). We addressed this issue by collecting methacholine-induced tracheobronchial secretions directly from pig airways (ASL). We originally collected tracheal ASL using the capillary technique, however it was not possible to collect sufficient material
for MS analysis from individual animals. Though the composition of secretions with methocholine induction might be different than steady state, the proteins detected are those that the epithelium might release under resting conditions, or in response to neurohumoral or environmental stimuli (18, 19, 36, 40).

The number of protein identifications increased markedly when we shifted our focus from BAL samples to ASL. However, it is not clear whether the ASL samples are intrinsically more protein-rich and diverse, or whether this is a consequence of refinements in our methods of sample collection and preparation, peptide fractionation, and/or LCMS instrumentation, making a direct comparison of the data difficult. For example, increasing the number of BAL peptide fractions analyzed from 8 to 30 resulted in an average 4.5-fold increase in protein identifications on the QSTAR Elite. Though fewer ASL fractions (n=20) were analyzed, more proteins were detected by using the LTQ Orbitrap Velos. It is not possible to determine the how much of this effect was due to the MS platform or inherent to the sample type. We note that both instruments undergo quality control measures assuring 1 fmol protein digest sensitivity. A confounding factor in the analysis of lung lining fluids is the presence of highly abundant proteins (e.g., albumin, serotransferrin, α2-macroglobulin, α1-acid glycoprotein) that contribute to the dynamic range of these complex mixtures and confound detection of medium and low abundant species. Affinity depletion strategies have been used to ameliorate this issue in human BAL, though no pig-specific antibodies are currently available for this purpose. We found that 75% of the proteins detected in BAL were also found in ASL. However, since other studies have reported ~900 proteins in BAL (67, 87) the overlap of common proteins in these fluids could be greater. We speculate that avoidance of the sample dilution associated with BAL collection may have enhanced the number of proteins detected. In addition to minimizing sample dilution, this approach also afforded the opportunity to investigate the protein repertoire of the tracheobronchial epithelium, representative of the conducting airways, separately from the alveolar compartment. While this sampling technique appears to be unique to our study, other investigators have attempted to examine the composition of ASL by using mass spectrometry to identify proteins in apical secretions from air-liquid interface cultures of human primary airway epithelia (3, 12, 38) or cell lines (37).
In the course of these experiments, we optimized database searching methods for identification of proteins in pig-derived samples. A critical factor was selection of the protein database. The porcine protein databases still have large gaps, though the number of entries is increasing monthly due to continued progress on the pig genome (30). The UniProt Knowledgebase (KB) version used in this study contained only 1,406 manually curated, non-redundant (SwissProt) entries for *Sus scrofa* proteins, as compared with 20,248 entries in the human and 65,102 in the mammalian databases (54). To overcome the paucity of fully annotated entries, MS data from porcine samples are often searched against either human plus pig or mammalian protein databases (6, 35, 56, 70). An additional UniProtKB database, TrEMBL, contains protein sequences associated with computationally generated annotation, but is not reviewed nor curated for redundancy. Protein fragments, isoforms and variants, encoded by the same gene, are found in separate entries. We annotated the 23,118 Ensembl *Sus scrofa* 10.2.67.pep.all database entries with protein name, gene name, and source database. Then we compiled databases that contained various combinations of the curated and non-curated pig and mammalian entries and tested various search parameters to determine the optimal strategy for identifying the greatest number of proteins with low FDRs for peptide and protein detection. In summary, searching the Ensembl *Sus scrofa* protein FASTA database explained the greatest number of spectra at a given FDR and reduced the number of redundant proteins found as compared to searching the larger mammalian databases. We detected many pig peptides and proteins whose existence was previously only predicted from available genomic (DNA) and transcriptomic (mRNA, EST) data. The Ensembl protein database is not complete, and we noted that some known lung proteins, such as the mucins MUC5AC and MUC5B as well as the pig-specific antimicrobial proteins PR-39 and PMAP-37, are not present. However, MUC5AC and PMAP-37 were identified in BAL and ASL using the pig Uniprot SwissProt and TrEMBL. The abundant MUC5B, which has been detected on the luminal surface of the bronchiolar epithelium of piglets airways using immunohistochemistry (41) is not present in any of the porcine protein databases thus precluding MS-based identification in these data sets.
The most abundant proteins in both the BAL and ASL datasets included plasma proteins: serotransferrin, serum albumin, complement factor C3, alpha-2-HS-glycoprotein (fetuin-A), alpha-fetoprotein, alpha-1-acid glycoprotein 1, alpha-2-macroglobulin, immunoglobulins, fetuin-B, and haptoglobin. In this respect, our study is in very good agreement with earlier proteomic analyses, which consistently reported plasma proteins in BAL and other airway fluids (14, 28, 44, 48, 50, 53, 58, 83, 84, 87). This finding may reflect physiologic transudates of plasma proteins from the circulation into the airways or alveoli and/or may be due to low level blood contamination introduced during sample collection procedures. However, there is also evidence that at least some of these proteins are normal secretory products of airway epithelia. Proteomic studies of apical secretions from cultured human airway epithelia have documented release of complement factor C3, alpha-2-macroglobulin, alpha-2-HS-glycoprotein, and serum albumin from these cells (3, 12, 38).

As shown in Figure 4, we found that the proteins detected in BAL and ASL were similarly distributed across protein classes, biological processes and molecular functions. Interestingly, the largest proportion (~25%) of proteins in both sample types was categorized as having a role in metabolic processes, likely reflecting the fact that a significant proportion of the detected proteins were predicted to be cytoplasmic or intracellular. The detection of intracellular proteins has been noted in earlier studies of airway secretions (3, 12, 38, 87). It is possible that some of these proteins may be derived from apoptotic cells, the result of normal turnover of airway epithelium, or possibly from cell debris due to disruption of the epithelium during sample collection. However, a review of the literature suggests that a certain number of intracellular proteins might actually be expected in the lung lining fluid of mammals, due to the secretory machinery of the airway epithelium and submucosal glands. Studies in humans, mice, and rabbits have provided evidence for both merocrine and apocrine secretion by goblet cells of the airway epithelium (and Clara cells in mice) (26, 42, 63, 64, 73). In merocrine secretion, secretory products are packaged into vesicles and exocytosed directly to the extracellular milieu; apocrine secretion involves release of cargo as part of membrane-bound vesicles that bud off the apical surface of the cell. Thus, it is anticipated that apocrine secretion could result in disruption of cell membranes and introduction of...
of intracellular and membrane proteins into the ASL. In support of this possibility, a recent proteomic study demonstrated that membrane-bound granules secreted from airway epithelial goblet cells contain cytoskeletal and regulatory proteins as well as mucins, suggesting that such granules may be a source for some of the intracellular proteins detected in ASL (69). Additionally, several groups have reported the presence of exosomes, microvesicles implicated in host defense and cell communication, in human BAL and cell culture secretions (1, 39). Exosomes have been shown to be associated with mucins, cytoskeletal proteins, and cytosolic enzymes, suggesting that these structures may also be a source of some of the intracellular proteins we identified.

Among the lung lining liquid (BAL and ASL) proteins identified were many associated with host defense (Supplemental Tables 1 and 3), including histone fragments, lactoferrin, lactoperoxidase, lysozyme, lipocalin 1, lipocalin 2, PLUNC, LPLUNC1, LPLUNC2, SLPI, surfactant proteins A and D, S100A8 (calgranulin-A), S100A9 (calgranulin-B), and the pig specific antimicrobial cathelicidin proteins (PMAP-23, PMAP-37), and protegrins-1 and -3. It is also interesting to note that several proteins of the complement family were present in lung lining liquid. These included complement components (C1r, C4, CB, C5, C6, C7, C9) and complement factors (B, D, H, and I). The complement system is recognized for its roles in innate immunity, with activities that include opsonization, chemotaxis, and membrane destruction. However, the functions of complement protein components in lung lining liquid are not well studied. Several protease inhibitors (e.g., the serine protease inhibitors, serpins) and proteases (e.g., the cathepsins) were also resident in lung lining liquid. This speaks to the importance of a balance of these forces in modulating protein function. Mucociliary clearance is an important component of airway host defenses and both tethered and secreted components of the mucin layer were identified, including Muc1, -2, -3A, -4, -5AC, -13, -16, and -19. Our data focus on an important early time point after birth. Thus, the BAL and ASL proteome may change over time as the animal develops.
In summary, this study is the first to define the proteome of the lung lining fluid in the newborn pig. For this task, we used both BAL fluid as well as methacholine-stimulated tracheal secretions in an effort to ensure that our results encompass both the conducting airways and the gas-exchange regions of the lung. In doing so, we greatly expanded the known proteome of the porcine lung, while also contributing to the greater body of literature documenting the composition of airway secretions across all mammalian species. Our database of porcine airway proteins should provide a framework for future studies utilizing porcine models of airway infection, disease or injury. In particular, it can serve as a reference for proteomic studies of pathogenesis and/or progressive lung changes in porcine models of cystic fibrosis and other inflammatory lung conditions. Additionally, these findings may also be useful for studies investigating responses to economically relevant pig pathogens such as Porcine Reproductive and Respiratory Syndrome (PRRS) and influenza. Thus, this “pig airway proteome” is a resource that will enhance the utility of the pig as an animal model for studies of lung biology, disease, and therapeutics.
Acknowledgments

We thank Paula Ludwig for technical support. Polyclonal antisera against porcine SP-D was generously provided by Dr. Henk Haagsman (Utrecht University, The Netherlands).

Grants

We acknowledge support from the Cystic Fibrosis Foundation RDP (PM and SF), NIH P50 HL-61234 (PM) and P01 HL-091842 (PM), as well as the Roy J. Carver Charitable Trust (PM). The UCSF Sandler-Moore Mass Spectrometry Core Facility is partially funded by a National Cancer Institute Cancer Center Support Grant, the Sandler Family Foundation, and the Gordon and Betty Moore Foundation.

At the request of the author(s), readers are herein alerted to the fact that additional materials related to this manuscript may be found at the institutional website of one of the authors, which at the time of publication they indicate is: https://wiki.library.ucsf.edu/x/vSzWAw. These materials are not a part of this manuscript, and have not undergone peer review by the American Physiological Society (APS). APS and the journal editors take no responsibility for these materials, for the website address, or for any links to or from it.
References


34. **Hubner NC, Ren S, and Mann M.** Peptide separation with immobilized pl strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics* 8: 4862-4872, 2008.


Figure Captions

Figure 1. Workflow for collection, processing, and mass spectrometric analyses of pig bronchoalveolar lavage (BAL) and airway surface liquid (ASL).

Figure 2. SDS-PAGE separation/silver staining of bronchoalveolar lavage (BAL) and airway surface liquid (ASL) proteins. A) BAL fluid from 6 individual newborn wild-type pigs, 2 µg protein/lane. B) ASL collected from 5 individual newborn wild-type pigs, 2 µg/lane. MW = molecular weight standard (SDS-PAGE Molecular Weight Standard, Low Range; Bio-Rad Laboratories, Hercules, CA).

Figure 3. Verification of selected proteins identified in LC-MS/MS analysis. A) BAL fluid from wild-type newborn pigs was electrophoretically separated by SDS PAGE under reducing conditions (top and middle panel) and non-reducing conditions (bottom panel). Proteins were transferred to PVDF membranes prior to immunoblotting with anti-sera specific for the PLUNC protein (top panel), SP-D (middle panel), and MUC5AC (bottom panel). Lanes 1-6 contained lavage samples from 6 individual pigs. Secretions from the apical surface of pig primary airway epithelial cultures (2 µL) served as a positive control for the PLUNC antibody (lane 7). Recombinant porcine SP-D (50 ng) was a positive control for the SP-D antibody (lane 7; noncontiguous lane from the same blot). Total protein from pig stomach scraping (10 ng) served as a positive control for the MUC5AC antibody (lane 7). B) The strategy described for Panel A was used for the immunoanalyses of five individual porcine ASL samples from wild-type newborn pigs.

Figure 4. Classifications for proteins identified in porcine BAL and ASL. Using PANTHER (80), proteins were classified by Gene Ontology terms describing: A) Biological Process, B) Protein Class, C) Molecular Function, and D) Cellular Compartment. Results are displayed as percent of genes classified to a category over the total number of class hits. Class hit means independent ontology terms and if a gene was classified to more
than one independent ontology terms that are not parent or child to each other, it counts as multiple class hits.

The total number of class hits for each category is shown in parentheses after the sample type.
Proteomics of porcine lung fluid

Tables

Table 1. Number of unique peptides and proteins detected in porcine BAL and ASL. Data are presented for each individual BAL and ASL sample. Distinct peptides at the ≤1% global false discovery rate (FDR) and proteins with a ≤5% local FDR are shown. The “Master” search result of all BAL or ASL sample sets searched together was used to disambiguate the protein groups and align the proteins across data sets (Supplemental Tables 1 and 3).

Table 2. MUC5AC and PMAP-37 peptides that were used for protein identification. MS data from BAL samples were searched using the pig Uniprot SwissProt and TrEMBL databases. Peptides detected with confidence levels ≥95% are shown. Coverage is the percent of matching amino acids (AAs) from identified peptides with any confidence divided by the total number of amino acids in the sequence. CAM, carbamidomethyl; Dea, deamidation; Cleavages, non-tryptic cleavage sites.

Table 3. Secreted host defense proteins detected in BAL and ASL. Proteins detected in BAL and ASL were cross referenced to a curated list of mRNA transcripts with recognized host defense and antimicrobial functions (66).

Supplemental Table 1. List of proteins detected in BAL.

Supplemental Table 2. Sequence similarity search results for the 70 BAL uncharacterized protein entries.

Supplemental Table 3. List of proteins detected in ASL.

Supplemental Table 4. Sequence similarity search results for 486 ASL uncharacterized protein entries.

Table 1
<table>
<thead>
<tr>
<th>Sample</th>
<th>Distinct Peptides</th>
<th>Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAL1</td>
<td>826</td>
<td>119</td>
</tr>
<tr>
<td>BAL2</td>
<td>149</td>
<td>46</td>
</tr>
<tr>
<td>BAL3</td>
<td>4096</td>
<td>384</td>
</tr>
<tr>
<td>BAL4</td>
<td>3898</td>
<td>161</td>
</tr>
<tr>
<td>BAL5</td>
<td>4270</td>
<td>422</td>
</tr>
<tr>
<td>BAL6</td>
<td>4957</td>
<td>558</td>
</tr>
<tr>
<td>Master BAL</td>
<td>9313</td>
<td>674</td>
</tr>
<tr>
<td>ASL1</td>
<td>3513</td>
<td>1134</td>
</tr>
<tr>
<td>ASL2</td>
<td>2316</td>
<td>803</td>
</tr>
<tr>
<td>ASL3</td>
<td>3207</td>
<td>975</td>
</tr>
<tr>
<td>ASL4</td>
<td>18012</td>
<td>2959</td>
</tr>
<tr>
<td>ASL5</td>
<td>22888</td>
<td>3289</td>
</tr>
<tr>
<td>Master ASL</td>
<td>30591</td>
<td>3858</td>
</tr>
</tbody>
</table>
## Table 2

<table>
<thead>
<tr>
<th>Accession</th>
<th>Name</th>
<th>Coverage (%)</th>
<th>Conf. (%)</th>
<th>Sequence</th>
<th>Cleavages</th>
</tr>
</thead>
<tbody>
<tr>
<td>O97866</td>
<td>MUC5AC (Fragment, 357 AA)</td>
<td>56.9</td>
<td>99</td>
<td>AESFPDTPLQALGQDVIC(CAM)DK</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WFDVFPSGPHGDFETYSNILR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LGQVVEC(CAM)RPEVGLVC(CAM)R</td>
<td></td>
</tr>
<tr>
<td>P49932</td>
<td>Antibacterial peptide PMAP-37</td>
<td>76.7</td>
<td>99</td>
<td>AVDRLNEQSSEANLYR</td>
<td>Missed R-L@4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LLELDQPPKADEDPGTPKPSFTVK</td>
<td>Missed K-A@9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LNEQSSEANLYR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LLELDQPPK</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ADEDPGTPKPSFTVK</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RPPELC(CAM)DFKEN(Dea)GR</td>
<td>W-R@N-term</td>
</tr>
<tr>
<td>Protein Identified in BAL/ASL</td>
<td>Gene from MicroArray</td>
<td>Protein Identified in BAL/ASL</td>
<td>Gene from MicroArray</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------</td>
<td>-------------------------------</td>
<td>----------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha-2-macroglobulin</td>
<td>A2M</td>
<td>High mobility group protein B1</td>
<td>HMGB1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP-binding cassette F1</td>
<td>ABCF1</td>
<td>Hermansky-Pudlak syndrome 5 protein</td>
<td>HPS5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAGE Protein</td>
<td>AGER</td>
<td>60 kDa heat shock protein, mitochondrial</td>
<td>HSPD1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angiotensinogen</td>
<td>AGT</td>
<td>Interleukin-6 receptor subunit beta</td>
<td>IL6ST</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum albumin</td>
<td>ALB</td>
<td>Integrin, beta 4</td>
<td>ITGB4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein AMBP</td>
<td>AMBP</td>
<td>Kininogen-1</td>
<td>KNG1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annexin A1</td>
<td>ANXA1</td>
<td>Alpha-lactalbumin</td>
<td>LALBA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annexin A5</td>
<td>ANXA5</td>
<td>Lipopolysaccharide-binding protein</td>
<td>LBP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apolipoprotein A-I</td>
<td>APOA1</td>
<td>Protein-lyase 6-oxidase</td>
<td>LOX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apolipoprotein A-IV</td>
<td>APOA4</td>
<td>Lactoperoxidase</td>
<td>LPO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apolipoprotein E</td>
<td>APOE</td>
<td>Lactotransferrin</td>
<td>LTF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-2-glycoprotein 1</td>
<td>APOH</td>
<td>Lysozyme C-2</td>
<td>LYSC2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-2-microglobulin</td>
<td>B2M</td>
<td>Lysozyme-like protein 6</td>
<td>LYZL6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bactericidal permeability-increasing protein</td>
<td>BPI</td>
<td>Macrophage migration inhibitory factor</td>
<td>MIF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complement C1r subcomponent C1R</td>
<td>Growth/differentiation factor 8</td>
<td>PLAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4b-binding protein alpha chain</td>
<td>C4BPA</td>
<td>Tissue-type plasminogen activator</td>
<td>PLAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complement C5a anaphylatoxin C5</td>
<td>Plasminogen</td>
<td>PLG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complement component C6 C6</td>
<td>BPI fold-containing family A member 1</td>
<td>PLUNC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complement component C7 C7</td>
<td>Peroxiredoxin-5, mitochondrial</td>
<td>PRDX5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carbonic anhydrase 2 CA2</td>
<td>Vitamin K-dependent protein C</td>
<td>PROC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>monocyte differentiation antigen CD14 CD14</td>
<td>Saposin-B-Val Saposin-B</td>
<td>PSAP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD59 glycoprotein CD59</td>
<td>Protein S100-A8 S100A8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD97 antigen CD97</td>
<td>Protein S100-A9 S100A9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complement factor D CFD</td>
<td>Serum amyloid A2 SAA4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complement factor I CFI</td>
<td>Alpha-1-antitrypsin SERPINA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Properdin CFP</td>
<td>Plasma serine protease inhibitor SERPINA5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clusterin CLU</td>
<td>Thyroxine-binding globulin SERPINA7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cystatin-C CST3</td>
<td>Leukocyte elastase inhibitor SERPINB1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cystatin-M CST6</td>
<td>Serpin B5 SERPINB5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cystatin-B CSTB</td>
<td>Heparin cofactor 2 SERPIN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lysosomal protective protein CTS A</td>
<td>Plasminogen activator inhibitor 1 SERPINE1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cathepsin B CTSB</td>
<td>alpha-2 antipasmin 1 SERPINF1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cathepsin C CTSC</td>
<td>alpha-2 antipasmin 2 SERPINF2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro-cathepsin H CTSH</td>
<td>Plasma protease C1 inhibit SERPING1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cathepsin L1 CTSL1</td>
<td>Pulmonary surfactant-associated protein D SFTPD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coxsackievirus and adenovirus receptor CXADR</td>
<td>Superoxide dismutase SOD1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acyl-CoA-binding protein DBI(32-86) DBI</td>
<td>Superoxide dismutase SOD1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epidermal growth factor EGF</td>
<td>Osteopontin SPP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coagulation factor XI F12</td>
<td>Serotransferrin TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coagulation factor V F5</td>
<td>Thyroglobulin TG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fibrinogen beta chain Fibrinopeptide B</td>
<td>FGB</td>
<td>Thrombospondin 1 THBS1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fibronectin FN1</td>
<td>Tenascin TNC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vitronectin VTN</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Proteomics of porcine lung fluid
Newborn pig

Lavage 3X

Swab collection

ice

Centrifuge 228 x g

BAL

ASL

-80 °C

TFE
denaturation

SDS
denaturation

Trypsin digest
in solution

Trypsin digest
in solution

Trypsin digest
on filter

RP HPLC
pH 10

IEF

RP HPLC
pH 10

LCMS

Total 4018 proteins identified