Available online at www.sciencedirect.com
ScienceDirect
Journal of Chromatography B, 864 (2008) 95–101
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Characterization of the mouse bronchoalveolar lavage proteome by micro-capillary LC–FTICR mass spectrometry

Joel G. Pounds [a], Jason W. Flora [c], Joshua N. Adkins [a], K. Monica Lee [b], Gaurav S.J.B. Rana [c], Tapas Sengupta [c], Richard D. Smith [a], Willie J. McKinney [c,*]

[a] *Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, 99354, United States*
[b] *Battelle Toxicology Northwest, Richland, WA, 99354, United States*
[c] *Philip Morris USA, Richmond, VA, 23261, United States*

Received 11 July 2007; accepted 27 January 2008
Available online 14 February 2008

Abstract

Bronchoalveolar lavage fluid (BALF) contains proteins derived from various pulmonary cell types, secretions and blood. As the characterization of the BALF proteome will be instrumental in establishing potential biomarkers of pathophysiology in the lungs, the objective of this study was to contribute to the comprehensive collection of Mus musculus BALF proteins using high resolution and highly sensitive micro-capillary liquid chromatography (μLC) combined with state-of-the-art high resolution mass spectrometry (MS). BALF was collected from ICR and C57BL/6 male mice exposed to nose-only inhalation to either air or cigarette smoke. The tandem mass spectra were analyzed by SEQUEST for peptide identifications with the subsequent application of accurate mass and time tags resulting in the identification of 1797 peptides with high confidence by high resolution MS. These peptides covered 959 individual proteins constituting the largest collection of BALF proteins to date. High throughput monitoring profiles of this extensive collection of BALF proteins will facilitate the discovery and validation of biomarkers that would elucidate pathogenic or adaptive responses of the lungs upon toxic insults.

Keywords: BAL; BALF; Bronchoalveolar; Lavage; Mus musculus; Proteome; FTICR; Cigarette smoke

1. Introduction

New genomic and proteomic technologies now provide the opportunity to measure hundreds or thousands of expressed genes and proteins in biological samples. The advantage of these technologies is that they provide unbiased, discovery driven characterization of the gene or protein expression within a biological matrix without *a priori* knowledge [1,2]. Assessment of gene expression by microarrays is currently a more mature technology than mass spectrometry (MS) based proteomics and has the advantage of relatively easier quantitation, ability to measure all expressed genes, and high throughput. The inherent limitation of characterizing gene expression is that it requires RNA from target or surrogate target tissues and its phenotypic changes are not as predictable. Proteomic approaches are attractive since diseases are generally manifested at the protein level and body fluids which do not contain RNA from diseased organs and can be easily sampled. Thus, changes in the levels of protein function resulting from alterations in proteins levels, cellular localization, post-translational modifications, and protein–protein interactions can define the physiological state of cells and tissues [3]. However, proteomic approaches are more labor intensive, require more sophisticated instrumentation, complex statistical analysis, bioinformatics, and have lower throughput than microarray expression measurements [4]. In addition, proteomes are conceptually much more complex than genomes, with alternative splicing and post-translational modifications resulting in perhaps approximately 10 times as many proteins as genes coding for proteins [5]. Nonetheless instrumentation, methods, and approaches for measuring many proteins simultaneously are rapidly developing and promise to become increasing powerful tools for investigations in molecular biology [4].

* Corresponding author at: Philip Morris USA, P.O. Box 26603, Richmond, VA, 23261, United States. Tel.: +1 804 743 2434; fax: +1 804 743 2498.
E-mail address: Willie.J.McKinney@pmusa.com (W.J. McKinney).

1570-0232/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2008.01.044

Currently, the most common proteomic analysis involves the use of two-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (2D-SDS-PAGE) [6]. The technique has been applied to the search for protein markers of pulmonary disease in bronchoalveolar lavage fluid (BALF) [7–14], nasal lavage fluid [15], and breath condensate [16]. These efforts pioneered the proteomic evaluation of respiratory system fluids as a source of information to understand disease etiology and potential biomarkers for diagnosis, susceptibility, and disease progression. In particular, bronchoalveolar lavage allows collection of pulmonary fluid that has become a useful technique to investigate pathogenesis in the lower respiratory tracts [17]. Certain cytokines including chemokines are detected in BALF as mediators for the acute or chronic inflammatory responses in the asthma or chronic obstructive pulmonary disease (COPD) animal models. These biochemical components of BALF such as cytokines, enzymes, and surfactants, in addition to cytological assessments, are reported as biomarkers for pulmonary diseases in experimental animals as well as humans [18–21]. Despite the high potential, however, the protein mediators in BALF represent only a small portion of the total secreted proteins from epithelial and inflammatory cells and from blood under a specific condition. The types and quantity of proteins vary with the physiological status of the lungs [17]. Because of generally low levels of some proteins (i.e., ∼pg/mL), the limited availability of BALF samples, and the lack of specific antibodies, the number of components that can be assessed using a traditional method, such as Enzyme-Linked ImmunoSorbent Assay (ELISA), is very limited. Similarly the common 2D-SDS-PAGE-MS approaches to BALF samples, which are complementary to liquid chromatography (LC) coupled to MS based approaches, do not provide an adequate dynamic range (the ability to detect low abundance proteins in the presence of higher abundance proteins) and lack sensitivity to detect low abundance proteins. To date, only about 100 proteins have been identified by 2D-PAGE-MS method of human BALF (http://us.expasy.org/ch2d/2d-index.html) [22]. A recent report has greatly expanded the Mus musculus proteome to almost 300 unique proteins using low resolution MS (ion trap) and combining both 1D-SDS-PAGE-MS and 2D liquid chromatography (LC) MS (2D-LC-MS) [23]. These studies illustrate that the protein composition of BALF provides important knowledge of lung structure and function at the molecular level and to the study of lung disease at the clinical level [24].

The objective of this study was to contribute to the comprehensive collection of Mus musculus BALF proteins using high resolution and highly sensitive micro-capillary liquid chromatography (μLC) combined with state-of-the-art high resolution MS. In this project, BALF was collected from control and cigarette smoke-exposed mice from a concomitant study. Two strains of mice selected for investigation were based on findings that exposure to cigarette smoke (3 cigarettes/day, 5 days/week) for 7 months produced emphysema in C57BL/6 but not in ICR mice [26]. However, the experiments discussed *vida infra* did not result in the histopathological outcome previously reported [26]: this apparent discrepancy is beyond the scope of this manuscript and the reason differential expression is not evaluated [27]. However, this sample pool provided an excellent opportunity from which to obtain a comprehensive collection of Mus musculus BALF proteins from both control and cigarette smoke-exposed mice potentially representing a variety of physiological states.

Utilization of an accurate mass and time (AMT) tag strategy [25] and sensitive μLC–MS/MS and μLC–FTICR-MS enabled the identification of a large number of peptides/proteins in BALF to significantly contribute to the search for biomarkers for toxic insult and pulmonary disease. Multiple μLC–MS/MS measurements of BALF samples were made for both control and smoke-exposed mice with differing chromatographies and MS instrument settings to enhance proteome coverage. The resulting peptide identifications were ultimately collated into two databases. The second stage of investigation involved using the results of these analyses to support identification of peptides/proteins using μLC coupled to a 9.4 T Fourier transform resonance ion cyclotron resonance (FTICR) MS without the limitations upon throughput imposed by tandem MS measurements. For the purpose of peptides/protein identification of BALF, the results from both approaches are combined and represent the largest collection of BALF proteins to date.

2. Experimental

2.1. Animals and animal care

Care and use of the animals were in conformity with the American Association for Laboratory Animal Science Policy on the Humane Care and Use of Laboratory Animals (AAALAC, 1991). The study protocols were approved by the Battelle IACUC prior to animal exposure. Young male C57BL/6 or ICR mice were obtained from Charles River Laboratories (Raleigh, NC) and acclimated for a week under an environmentally controlled housing condition.

2.2. Exposure to cigarette smoke in mice

Two animal studies were conducted: (1) single-dose study: a group of mice (4/strain/exposure) were exposed to 2-h nose-only inhalation to mainstream cigarette smoke (CS) of Kentucky Reference 1R4F cigarettes (0 or 600 μg wet total particulate matter per liter (WTPM/L)) for 7 consecutive days. BALF from this study was used primarily for the method development; (2) dose–response study: a separate group of mice (10/strain/exposure) were exposed 2-h nose-only inhalation to mainstream CS of Kentucky Reference 2R4F cigarettes (0, 75, 250 and 600 μg WTPM/L) for 7 consecutive days. BALF from both studies were used to construct the Potential Mass and Time tag (PMT) database, but only the μLC–FTICR-MS analysis of BALF from mice exposed for 7 days was used for final protein identification.

2.3. Bronchoalveolar lavage fluid

At 12 h following the last exposure, BAL was performed on the isolated lungs by cannulating the trachea and washing the lungs two times with phosphate-buffered saline (PBS, pH 7.2) using a volume of ∼1 mL/wash. Retrieved fluid was kept on ice.

The lavage washes were centrifuged at 400 RCF for 10 min at 4 °C. Ten microliters aliquots of cell-free supernatant (BALF) were used for protein concentration determination using a Bio-Rad Protein Assay (Bio-Rad, Hercules, CA). The BALF (∼1 mL per mouse for the single-dose study and ∼100 μL per mouse for the dose–response study) were flash frozen and stored at −80 °C prior to proteomic analysis.

2.4. Precipitation of proteins

The study BALF samples were pooled per treatment group (exposure and strain) to provide aliquots of 40 μg protein. A separate reference BALF sample was made by pooling the excess BALF available. Ice-cold trichloroacetic acid (TCA) was added to the BALF samples to a final concentration of 8%. All BALF samples were incubated on ice with regular shaking for 15 min, followed by micro-centrifugation at 16000 RCF and the supernatants discarded. The protein pellets were washed with acetone and lyophilized (dried under vacuum) and stored at −80 °C until used.

2.5. Trypsin digestion

The BALF protein pellets from study and reference samples were resuspsended in 30 μL denaturing buffer (filtered 50 mM Trizma pH 8.3, 6 M guanidine–HCl, 5 mM EDTA, overall pH 8.4). Each sample was then diluted into 500 μL of 100 mM ammonium bicarbonate, 5 mM EDTA, pH 8.4 and then exchanged into the same buffer using a 5 mL desalting column (Pierce, Rockville, MD). Fractions were taken at 200 μL intervals and manually measured for protein content by checking absorbance at 280 nM. The protein containing fractions 7–12 were pooled. The pooled fractions from the desalting column were mixed with acetonitrile to a final concentration of 10% by volume. The 10% acetonitrile was used to overcome reversed-phase column plugging in subsequent steps, which was found to be concurrent with early eluting species (likely non-protein). The proteins were digested with 2 μg trypsin (Promega, Madison, WI) for 4 h at 37 °C. Following trypsin digestion, the samples were lyophilized to dryness and stored at −80 °C until just prior to MS analysis.

2.6. Reversed-phase separation

Protein samples were first separated by ultra-high pressure micro-capillary liquid chromatography on-line before analysis in a multipole ion trap mass spectrometer (ThermoFinnigan, LCQ, Duo) and a μLC–Fourier Transform Ion Cyclotron Resonance MS (μLC–FTICR-MS) as reported previously [28]. This reversed phase μLC instrumentation was developed at PNNL (Richland, WA) and used a 150 μm i.d. × 360 μm o.d capillary (Polymicro Technologies Inc., Phoenix, AZ) fitted with a 2-μm retaining mesh and packed with 60 cm of 5 μm Jupiter C18 stationary phase (Phenomenex, Torrence, CA). Mobile phase A consisted of 0.05% trifluoroacetic acid (TFA), 0.2% acetic acid in water, and mobile phase B consisted of 0.1% TFA, 90% acetonitrile in water. The linear gradient mixing of mobile phase A with mobile phase B began while maintaining constant pressure (10,000 psi) 20 min following a 10 μL injection of the sample (∼1.0 μg/μL) in mobile phase A. The peptides were eluted with a linear gradient from 5 to 70% B over 80 min. The capillary flow rate was ∼1.8 μL/min. This flow rate was achieved by using a pair of 100-mL ISCO pumps (model 100DM) controlled by a ISCO series D controller and a microflow processor splitter (LC Packings, San Francisco, CA).

2.7. Ion trap MS/MS analyses

The reversed-phase chromatography flow was infused directly into the Thermo Finnigan (San Jose, CA) LCQ ion trap mass spectrometer. Each sample was analyzed three times on the ion trap mass spectrometer with precursor mass/charge (m/z) segments of 400–950, 900–1450, and 1400–2000 in different runs to increase overall proteome coverage. Following each MS scan, the three most intense m/z peaks were subject to MS/MS analysis (determined by mass spectrometer software in real time). The MS/MS spectra were performed by using an isolation width of 3 m/z, a normalized collision energy of 45, an activation time of 30 ms, and an exclusion duration of 3 min.

2.8. μLC–FTICR mass spectrometry

A Bruker Daltonics (Bremen, Germany) 9.4-T FTICR mass spectrometer was combined with the capillary LC system described above and modified for concurrent internal mass calibration, and auto-sampling was used [28,25]. Injected samples contained trypsin digested peptides equivalent to 10 μg protein. These analyses typically result in peptides analyzed with <5 part per million (ppm) mass measurement accuracy (MMA) depending on the dynamic range of the measurements [29]. When elution times are taken into account ∼50% of the SEQUEST identification peptides are useful for protein identifications in an organism like *Deinococcus radiodurans* [28]. This approach provides broad proteome coverage with a large dynamic range, and sensitivity for high throughput proteome measurements.

2.9. Mass and time tag database from SEQUEST results

The data used to establish the mass and time tag database were obtained from multiple sample sources. The raw LC-ion trap data from our previous multidimensional analysis were reanalyzed for this work along with individual ion trap analyses of each BALF sample. These combined peptide identifications were used to populate the database that was, subsequently, used for generating the AMT tag results and ultimate protein identifications. These SEQUEST data filters are: tryptic on at least the N- or C-terminus and a minimum DelCN value of 0.1 with +1, +2, and +3 charge states, using minimum XCorr values of 1.9, 2.2, and 3.75, respectively. The SEQUEST analysis was performed with the M_musculus_2005_12_08_NCBI.fasta version of the mouse database available at NCBI (www.ncbi.nlm.nih). The peptide

identifications were stored in database internally designated MT_Mouse_P250. The relatively high false-positive peptide identification rate was 22% for the MT database based on SEQUEST analysis against a reversed protein sequence database. This false-positive identification is consistent with similar studies using complex proteome and similar identification criteria. For example, recent analyses of human plasma and other samples provided the basis for estimating the false-positives rates for SEQUEST results, although the filter rules were similar, but not identical to those used here for populating the MT tag database [30]. The false-positive value was as high as 16–32%, depending on the calculation. As such, the false-positive identification rate of the BALF MT tag database could also fall in this range.

In our study, these MT identifications are provisional and the high mass measurement accuracy and inclusion of Net Elution Time afforded by the AMT tag approach is used to significantly improve the confidence in peptide identifications over SEQUEST analyses alone [31,32,25]. The estimated false-positive errors for these AMT identifications are estimated between 8 and 10% depending on the method used. First, the false-positive rate of the AMT tag process was estimated by peak matching against a combined forward and reverse database. This approach estimated the overall false-positive rate as 8% for this dataset (calculated with combined databases as performed by Peng et al., 2003) [33], a substantial reduction from the 22% estimated using the SEQUEST filters. A second method to evaluate false-positive error of the AMT tags was to use the mass error plot to estimate "false-positive background" (i.e., those measurements that contribute to a "noise" baseline). The number of identifications that fall below the baseline are false positives, and the "true positive identifications", are those that contribute to the peak centered at 0 ppm MMA. This value was calculated to be ~10% for this dataset (data not shown).

Any peptide identified by the AMT tag approach was mapped to all proteins containing that specific peptide to provide for the broadest possible description of the BALF proteome. It should be recognized that the identification confidence for peptides does not directly correlate to the identification confidence for proteins identified from these peptides. Most methodologies have attempted to limit false-positive peptide identifications and the false-positive incidences for proteins are often considered to be of similar magnitude as the peptides. However, protein misidentifications are actually greater because even correct peptide identifications can result in proteins with multiple identifications. Thus, the false-positive incidence for proteins identified by single peptide identifications should be viewed as having a higher false-positive rate than that for peptide identifications.

2.10. Gene ontology (GO)

All identified proteins were mapped to Gene Ontology functional labels using FatiGO [34]. The proteins were queried separately at the fourth level for Cellular Component, Molecular Function, and Biological Process.

3. Results and discussion

To generate the proteomic data discussed in this manuscript, over 412,000 tandem mass spectra from 123 capillary μLC–MS/MS runs were analyzed by SEQUEST to identify 14,332 MT tags. The pooled and digested BALF samples were analyzed in triplicate by μLC–FTICR-MS using a 5 ppm mass measurement accuracy filter. This analysis resulted in 6284 proteins identified by at least one AMT tag from the 27 μLC–FTICR-MS analyses. Of these 6284 proteins, 5461 have at least one tag that is mapped to a single protein or open reading frame (ORF). To increase confidence in protein identification, only proteins that were consistently detected by a tryptic peptide in all three replicate μLC–FTICR-MS analyses from one or more BALF samples were considered for further analysis. The Supplemental Data Table contains a listing of all of the peptides observed and associated with the resulting 959 protein identifications. Greater than 89% of the 959 BALF proteins identified were identified by tryptic peptides.

The BALF peptides identified by the AMT tag strategy employed herein include those that map to more than one ORF. Examples of MT tags that map to multiple ORFs include peptides from the various serine protease inhibitors that were found to be common in the BALF samples. Peptides may map to multiple ORFs or database accession numbers for several reasons. For example, the peptide may represent a highly conserved protein motif within a gene family. In addition, the databases used for peptide identification may contain multiple entries for the same gene product as polymorphisms or spliceoforms. The Supplemental Data Table includes each peptide found, the proteins to which it is mapped, and the NCBI Reference Identification.

Selected groups of identified proteins are summarized in Table 1 selected from the Supplemental Data Table. The proteins listed were selected to illustrate the breadth of mouse proteins identified by the analysis. The complete list of identified proteins and the peptides identified is provided in the Supplemental Data Table. The identification of peptides and the subsequent naming of a parent protein from those peptides is a statistical problem that involves analytical accuracy, bioinformatics, computation, and reproducibility of results. The protein identifications made from multiple, high abundance peptide measurements are thus more confident than those made from a single peptide measurement with a lower statistical but still significant score. A significant group of peptides/proteins identified from either control or smoke-exposed mice were classical plasma-based proteins (e.g., haptoglobin, hemopexin, and complement components). Therefore, in future studies immunodepletion of highly abundant protein could afford a greater dynamic range and larger peptide coverage of low abundant proteins which is a current limitation of this approach.

The list of proteins identified in this study (Supplemental Data Table) include many proteins identified by previous proteomic reports of BALF using 2D-SDS-PAGE-MS approaches [7–10]. For example, a list of human BALF proteins maintained by Dr. R. Wattiez (http://www.umh.ac.be/~biochim/list.htm, May 2000) includes approximately 60 different proteins, 36 of

Table 1
Summary of proteins identified in BALF

Protein function	Selected proteins identified in BALF
Protease and peptidase related	Cathepsins B, C, D, G, and S Serine protease inhitors-2; 2–2; 1–1; 1–2; 1–4; 1–5; and 10 SUMO-1 specific protease Kininogen Transmembrane specific protease: serine 3 Ubitiquin specific protease 15 Ubitiquination factor E4A Disintegrin and metalloprotease: ADAM and reprolysin types Membrane bound transcription factor protease (subtilisin/kexin isozymes-1) Membrane metallo endopeptidase Stromolysin 1 and 2 Murinoglobulin 1 and 2 Serine (or cysteine) proteinase inhibitor: clade A, members 3A, 3C, 3G, 3M, Serine (or cysteine) proteinase inhibitor: clade C (anti-thrombin) member 1 Serine (or cysteine) proteinase inhibitor: clade F, member 2
Receptor related	Adrenergic receptor alpha 2c Cholinergic receptor Colony stimulating factor 2 receptor beta 2 Endothelin receptor type B Epidermal growth factor receptor pathway substrate 8 Estrogen related receptor beta Fibroblast growth factor receptor 2 Glutamate receptor metabotropic 3 Growth hormone receptor G-protein coupled receptor 25 Mannose receptor, C type 1 Neurotinsin receptor 2 Glucocorticoid receptor 1 Olfactory receptors M04110-10, MOR178-1. 185-3, MOR204-8, MOR259-10 Protein tyrosine phosphatase receptors: types C, K, and N Toll-like receptor 9
Respiratory tract-related	Lung carbonyl reductase Palate: nasal: lung carcinoma associated protein Surfactant associated protein A Surfactant associated protein D Vomeronasal 1 receptor H4 Anti-oxidant protein 2 (peroxiredoxin 5)
Miscellaneous proteins	Polymeric immunoglobulin receptor, Class R Chemokine (C-X-C-motif) subfamily B: member 15 Clusterin (apolipoprotein J) Ubiquitin-protein isopeptide ligase (cullin 4A) Glutathione *S*-transferases omega 1; mu 1, 2 and 6; pi 2; alpha 3 and 4, Lipocalcin 2 Orosomucoids 1 and 2

which were identified in the BALF analysis described in this manuscript. These overlapping proteins include blood-derived proteins such as albumin, transferrin, ceruloplasmin, complement C3, hemoglobin and others. Lung cell-derived proteins identified in both databases include pulmonary surfactant protein A and alpha1-antichymotrypsin, lipocalcin 1, and cyclophilin. The Wattiez database contains proteins identified in BALF from individual humans with multiple malignant and non-malignant pulmonary diseases. Therefore, it is not surprising that the number of overlapping protein identifications are low compared to the proteins identified in BALF from healthy mice or from mice with short-term exposure to cigarette smoke.

An analogous study of mouse BALF using 1D-GE and LC coupled to low resolution MS/MS identified 297 proteins [23] and was compared to protein list found in the Supplemental Data of this manuscript. This task was challenging because both protein lists contained GenBank identifications. GenBank identification numbers are frequently retired which does not make them ideal for comparison of proteomic results. All GenBank identification numbers were therefore converted to Swiss Prot identifications prior to comparison. It should be noted that 65 of the 297 proteins identified by Guo et al. were isoforms and, therefore, no GenBank identification number was given. It was also found that 19 of the resulting 232 unique proteins had been retired and could not be converted. The Supplemental Data associated with this manuscript also contained 165 GenBank identification numbers that had been retired. Comparison of these results (213 compared to 794 proteins with Swiss Prot identification numbers) showed an overlap of 43.2%. This level of overlap in protein identifications is consistent with those reported by the HUPO Plasma Proteome Project for multiple laboratory analysis of plasma [35]. While the overlap in BALF protein identifications across studies is less than ideal, it does not discount the value of either dataset. Instead it exemplifies several of the challenges faced with the analysis of proteomic data including the comparison of results, the vast quantity of proteins found in biological fluids, and variation amongst sample treatment methodologies.

It is clear that the identification of proteins in a specific proteome is an analytical and bioinformatic challenge; the even more daunting challenge is the biological interpretation of the results. To assist in this interpretation, the GO Consortium orga-

Table 2
Cellular component percentage

Intracellular organelle	78.1	Membrane-bound vesicle	2.6
Intracellular membrane-bound organelle	55.8	Cytoplasmic vesicle	2.6
Cytoplasm	53.7	Mitochondrial envelope	2.4
Intracellular non-membrane-bound organelle	32.9	Endomembrane system	2.4
Nucleus	25.7	Organelle inner membrane	2.4
Cytoskeleton	25.2	Nuclear envelope	1.7
Intrinsic to membrane	12.5	Lamellipodium	1.7
Plasma membrane	11.5	Nuclear lumen	1.7
Chromosome	6.1	Spliceosome complex	1.7
Ribonucleoprotein complex	4.2	Synapse	1.7
Organelle envelope	4.2	Basement membrane	1.4
Membrane fraction	4.0	Ribosome	1.2
Organelle membrane	3.8		

Table 3
Molecular function percentage

Protein binding	34.6	Cofactor binding	2.3
Ion binding	22.6	Receptor binding	1.9
Nucleotide binding	18.8	Pattern binding	1.7
Hydrolase activity	16.5	Electron transporter activity	1.7
Nucleic acid binding	12.0	Lyase activity	1.6
Oxidoreductase activity	11.3	GTPase regulator activity	1.4
Transferase activity	8.0	Peroxidase activity	1.2
Enzyme inhibitor activity	5.9	Lipid transporter activity	1.2
Lipid binding	3.5	Tetrapyrrole binding	1.2
Receptor activity	3.3	Structural constituent of ribosome	1.0
Ion transporter activity	2.8	Vitamin binding	1.0
Carrier activity	2.8	Ligase activity	1.0
Isomerase activity	2.6	ATPase activity (movement of substances)	1.0
Carbohydrate binding	2.6	Protein transporter activity	1.0
Structural constituent of cytoskeleton	2.3		

Table 4
Biological process percentage

Physiological process	81.3	Death	4.6
Metabolism	62.2	Cell differentiation	4.4
Localization	31.1	Positive regulation of biological process	4.3
Regulation of physiological process	14.9	Locomotion	3.3
Regulation of cellular process	13.9	System development	3.1
Organismal physiological process	12.6	Homeostasis	2.5
Response to stress	12.2	Behavior	2.1
Cell communication	11.8	Coagulation	1.7
Organ development	8.9	Regulation of development	1.5
Response to biotic stimulus	7.9	Regulation of enzyme activity	1.5
Morphogenesis	7.3	Tissue development	1.5
Negative regulation of biological process	6.2	Response to endogenous stimulus	1.4
Cell adhesion	5.4	Tube development	1.2
Response to external stimulus	5.2	Embryonic development	1.0
Response to abiotic stimulus	5.2		

nizes genes into hierarchical categories based on biological process, molecular function, and subcellular localization [36]. The GO annotations for the BALF proteins identified in the supplemental material are summarized in Tables 2–4 where categories with greater than 1% of the proteins mapped are listed. After filtering redundant and unknown proteins 686 of the 959 proteins were used. FatiGO [34] recognized 503, 575, and 518 out of the 686 as proteins with annotations for the categories of Cellular Component (Table 2), Molecular Function (Table 3), and Biological Process (Table 4), respectively.

4. Conclusions

It is not surprising that there is a significant variety of function within the proteins elucidated by this research; nonetheless, the objective of this study was to extend the proteomic characterization of mouse BALF and thereby the knowledge base with respect to potential functional changes in the lung. High throughput monitoring profiles of this extensive collection of BALF proteins along with the incorporation of BALF proteins identified by complementary techniques (e.g., 2D-SDS-PAGE-MS) will facilitate the discovery and validation of biomarkers that would elucidate pathogenic or adaptive responses of the lungs upon toxic insults. Furthermore, the results of this study illustrate the extensive breadth and scope of protein profiles that can be extracted from the mammalian pulmonary epithelial lining fluid. A comprehensive analysis of the proteomics data as well as correlation with other biological endpoints measured in the same study [37] is currently underway.

Acknowledgments

The authors thank Bruce Westerberg and Mark Gritz (Battelle) and Philip Morris USA for supporting this research. The authors also would like to acknowledge the mass spectrometry assistance and expertise provided by PNNL (Ronald Moore, Harold Udseth, Jason McCann, and Gary Sedgwick) and animal studies performed at Battelle Toxicology NW (Richland, WA). The authors would like to thank Ken Auberry (PNNL) for his expert assistance, Rebecca Secrist for her assistance with the GenBank to Swiss Prot conversion, and Dr. George Patskan (Philip Morris USA) and Dr. Thomas Mueller (Philip Morris Research Laboratories Gmbh Cologne, Germany) for their critical review of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jchromb.2008.01.044.

References

[1] B.A. Merrick, Environ. Health Perspect. 111 (2003) 797.
[2] M. Waters, G. Boorman, P. Bushel, M. Cunningham, R. Irwin, A. Merrick, K. Olden, R. Paules, J. Selkirk, S. Stasiewicz, B. Weis, B. Van Houten, N. Walker, R. Tennant, Environ. Health Perspect. 111 (2003) 811.
[3] R. Aebersold, B.F. Cravatt, Trends Biotechnol. 20 (2002) 1.
[4] M. Tyers, M. Mann, Nature 422 (2003) 193.
[5] J. Rappsilber, U. Ryder, A.I. Lamond, M. Mann, Genome Res. 12 (2002) 1231.
[6] P.H. O'Farrell, J. Biol. Chem. 250 (1975) 4007.
[7] I. Noel-Georis, A. Bernard, P. Falmagne, R. Wattiez, Dis. Markers 17 (2001) 271.
[8] I. Noel-Georis, A. Bernard, P. Falmagne, R. Wattiez, J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 771 (2002) 221.
[9] M. Neumann, C. Bredow, F. Ratjen, M. Griese, Proteomics 2 (2002) 683.
[10] C. He, Proteomics 3 (2003) 87.
[11] R.B. Devlin, H.S. Koren, Am. J. Respir. Cell Mol. Biol. 2 (1990) 281.
[12] A.G. Lenz, B. Meyer, U. Costabel, K. Maier, Electrophoresis 14 (1993) 242.
[13] M. Lindahl, T. Ekstrom, J. Sorensen, C. Tagesson, Thorax 51 (1996) 1028.
[14] R. Wattiez, C. Hermans, C. Cruyt, A. Bernard, P. Falmagne, Electrophoresis 21 (2000) 2703.
[15] B. Ghafouri, B. Stahibom, C. Tagesson, M. Lindahl, Proteomics 2 (2002) 112.
[16] M. Griese, J. Noss, C. von Bredow, Proteomics 2 (2002) 690.

[17] J. Hirsch, K.C. Hansen, A.L. Burlingame, M.A. Matthay, Am. J. Physiol. Lung Cell Mol. Physiol. 287 (2004) L1.
[18] W. Song, J. Zhao, Z. Li, Chin. Med. J. (Engl.) 114 (2001) 1140.
[19] M. Tanino, T. Betsuyaku, K. Takeyabu, Y. Tanino, E. Yamaguchi, K. Miyamoto, M. Nishimura, Thorax 57 (2002) 405.
[20] C. de Torre, S. Ying, P.J. Munson, G.U. Meduri, A.F. Suffredini, Proteomics 6 (2006) 3949.
[21] Y. Lee, P. Chen, P. Tsai, S. Su, P. Liao, Proteomics 6 (2006) 2236.
[22] A. Plymoth, C.-D. Lofdahl, M. Dahlback, H. Lindberg, T.E. Fehniger, G. Marko-Varga, Proteomics 3 (2003) 962.
[23] Y. Guo, S. Ma, D. Grigoryev, J. Van Eyk, J.G.N. Garcia, Proteomics 5 (2005) 4608.
[24] R. Wattiez, P. Falmagne, J Chromatogr. B 815 (2005) 169.
[25] R.D. Smith, G.A. Anderson, M.S. Lipton, L. Pasa-Tolic, Y. Shen, T.P. Conrads, T.D. Veenstra, H.R. Udseth, Proteomics 2 (2002) 513.
[26] E. Cavarra, B. Bartalesi, M. Lucattelli, S. Fineschi, B. Lunghi, F. Gambelli, L.A. Ortiz, P.A. Martorana, G. Lungarella, Am. J. Respir. Crit. Care Med. 164 (2001) 886.
[27] K.C. Hodge-Bell, K.M. Lee, R.A. Renne, K.M. Gideon, S.J. Harbo, W.J. McKinney, Inhal. Tox. 19 (2007) 361.
[28] M.S. Lipton, L. Pasa-Tolic, G.A. Anderson, D.J. Anderson, D.L. Auberry, J.R. Battista, M.J. Daly, J. Fredrickson, K.K. Hixson, H. Kostandarithes, C. Masselon, L.M. Markillie, R.J. Moore, M.F. Romine, Y. Shen, E. Strittmatter, N. Tolic, H.R. Udseth, A. Venkateswaran, K.K. Wong, R. Zhao, R.D. Smith, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 11049.
[29] M.E. Belov, G.A. Anderson, M.A. Wingerd, H.R. Udseth, K. Tang, D.C. Prior, K.R. Swanson, M.A. Buschbach, E.F. Strittmatter, R.J. Moore, R.D. Smith, Am. J. Mass Spectrom. 15 (2004) 212.
[30] W.-J. Qian, T. Liu, W.E. Monroe, E.F. Strittmatter, J.M. Jacobs, L.J. Kangas, K. Petritis, D.G. Camp II, R.D. Smith, J. Proteome Res. 4 (2005) 53.
[31] K.K. Anderson, M.E. Monroe, D.S. Daly, Proc. Int. Conf. METMBS (2004) 151.
[32] M.E. Monroe, N. Tolić, N. Jaitly, J.L. Shaw, J.N. Adkins, R.D. Smith, Bioinformatics 15 (2007) 2021.
[33] J. Peng, J.E. Elias, C.C. Thoreen, L.J. Licklider, S.P. Gygi, J. Proteome Res. 2 (2003) 43.
[34] F. Al-Shahrour, R. Diaz-Uriarte, J. Dopazo, Bioinformatics 20 (2004) 578.
[35] G.S. Omenn, D.J. States, M. Adamski, T.W. Blackwell, R. Menon, H. Hermjakob, et al., Proteomics 5 (2005) 3226.
[36] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Nat. Genet. (2000) 25.
[37] C.J. Obot, K.M. Lee, A.F. Fuciarelli, R.A. Renne, W.J. McKinney, Inhal. Tox. 16 (2004) 701.