Proteomics: current techniques and potential applications to lung disease

Jan Hirsch,1 Kirk C. Hansen,2 Alma L. Burlingame,2 and Michael A. Matthay1
1Cardiovascular Research Institute and 2Mass Spectrometry Facility, Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

Hirsch, Jan, Kirk C. Hansen, Alma L. Burlingame, and Michael A. Matthay. Proteomics: current techniques and potential applications to lung disease. Am J Physiol Lung Cell Mol Physiol 287: L1–L23, 2004; 10.1152/ajplung.00301.2003.—Proteomics aims to study the whole protein content of a biological sample in one set of experiments. Such an approach has the potential value to acquire an understanding of the complex responses of an organism to a stimulus. The large vascular and air space surface area of the lung expose it to a multitude of stimuli that can trigger a variety of responses by many different cell types. This complexity makes the lung a promising, but also challenging, target for proteomics. Important steps made in the last decade have increased the potential value of the results of proteomics studies for the clinical scientist. Advances in protein separation and staining techniques have improved protein identification to include the least abundant proteins. The evolution in mass spectrometry has led to the identification of a large part of the proteins of interest rather than just describing changes in patterns of protein spots. Protein profiling techniques allow the rapid comparison of complex samples and the direct investigation of tissue specimens. In addition, proteomics has been complemented by the analysis of posttranslational modifications and techniques for the quantitative comparison of different proteomes. These methodologies have made the application of proteomics on the study of specific diseases or biological processes under clinically relevant conditions possible. The quantity of data that is acquired with these new techniques places new challenges on data processing and analysis. This article provides a brief review of the most promising proteomics methods and some of their applications to pulmonary research.

mass spectrometry; proteome; lung

PROTEOMICS IS THE INVESTIGATION of the protein content or the protein complement of the genome of a biological system, also termed the proteome (237, 255). The objective of proteome research is to identify and describe the complex responses of a biological system to different stimuli. A vast amount of information can be obtained from one set of experiments compared with the classic approach of observing concentration changes or modifications on the single protein level. The Nobel Prize for Chemistry in 2002 was shared among three scientists for the development of analytical methods for the study of biomolecules: Kurt Wüthrich for the nuclear magnetic resonance technique and John Fenn and Koichi Tanaka for development of the two ionization techniques that initiated the rapid evolution of biological mass spectrometry in the past decade, namely electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI).

In the first part of this review, the status of current proteomic techniques is outlined. The rapid evolution in mass spectrometry, which was initiated by the development of the ionization techniques MALDI and ESI, has led to significant improvements in the central step of a proteomics experiment, protein identification.

The subsequent application of these techniques to a large number of previously inaccessible categories of samples has in turn triggered progress in other crucial steps, namely protein separation techniques and the analysis of the resulting data. The classic proteomics approach of describing and comparing the protein content of a given sample has consequently been refined by the description of “posttranslational modifications” of the protein and widened by tools that allow quantitative comparison of two or more samples (“quantitative proteomics”). This article provides an outline of current techniques in both of these fields that will be used to investigate the lung proteome in the coming years.

The third part of this review focuses on the present status of the investigation of the lung proteome with specific examples from pulmonary studies that have evaluated bronchoalveolar lavage as well as other biological samples in a variety of acute and chronic lung diseases. Some future possibilities for lung research that may arise from the rapid progress occurring in proteome method development are also considered (3, 69, 103, 240).

IMPLICATIONS OF THE HUMAN GENOME PROJECT

After the completion of the Human Genome Project, it has become evident that the complexity of organisms is only to a small part the result of direct gene expression from the genome (2, 76, 203), and it is clear that the simple concept of one gene-one protein is incorrect. One important reason for this is...
that the product of one gene can be transformed to a whole family of gene products (106, 118, 214), i.e., one gene can produce multiple mature mRNAs via alternative splicing and other mechanisms (195). Furthermore, the correlation between mRNA and protein concentrations has been demonstrated to be insufficient to predict protein expression levels from quantitative mRNA data, since protein levels are regulated by degradation as well (40, 90). Although the correlation between mRNA levels and protein abundance was very good for a limited number of highly abundant proteins, it was poor for proteins with lower expression levels (90). In this group, 30-fold differences in protein abundance were found for proteins with the same mRNA levels. Proteins from genes with very low expression levels could not be detected at all in this study with a current two-dimensional (2D) electrophoresis-mass spectrometry approach. Posttranslational modifications such as glycosylation, phosphorylation, and ubiquitination produce further variations by increasing the number of components from the standard 20 amino acid to more than 140 possible amino acid forms (125). These modifications undergo rapid changes and usually are not mutually exclusive (152).

Therefore, the study of the genome or even mRNA levels (the transcriptome) will reveal only a small spectrum of the response to a particular stimulus. Even from the diseases known to be based on specific genetic defects, only a very small number are likely to be monogenic, since cellular systems include complex interactions with a high level of redundancy (234). Conversely, the function of a large number of the protein products that are encoded by these genes is still unclear (25).

Direct investigation of the proteome provides a more complete representation of changes in the status of an organism. However, there exist several impediments to such an approach, the sheer complexity of the proteome being the most important one (Fig. 1). Diversity is another issue, since there are at least 250 different types of human cells, each of which contains at least 2,000–6,000 different primary proteins (33, 59), and posttranslational modifications will multiply this number (152, 165, 257, 258). It has been estimated that the different types of human cells may differ from each other in ~400 unique proteins (32). Another important factor is the dynamic range of concentrations of proteins, since one cell can contain between one and more than 100,000 copies of a single protein (32).

Finally, the proteome of organisms is dynamic and changes with environment and with time (106).

Many of the detection and recognition methods currently used in protein chemistry, such as antibody assays or enzyme activity measurements, have the capacity to detect only one protein at a time. Consequently, many investigations measure the response of one gene, protein, or pathway in the context of normal physiology or a pathological condition. For example, we have studied how aquaporin deletions influence fluid transport in the lung by studying aquaporin knockout mice under normal physiological conditions as well as during clinically relevant stresses, such as at the time of birth or during experimentally induced lung injury (225). A large part of our current understanding of biological function is based on this type of investigation. This approach will continue to be useful for a detailed understanding of living organisms. However, to study the interactions between the proteins identified with the same methods, a large number of consecutive experiments are necessary. This is not only time consuming, but the interpretation of results may be hampered by additional factors that are introduced by variabilities in experimental parameters, differences in cell material, and the time of measurement. Emerging proteomics methods have the potential to overcome many of these limitations.

**WHAT CAN BE MEASURED USING A PROTEOMICS APPROACH?**

Proteomic investigation of a given cell or other biological system should ideally detect all proteins and their functional responses to a stimulus. Given this goal, it is not surprising that no approaches currently come close to achieving this (see below). However, despite the present limitations, a proteomics-based approach has the unique advantage to identify changes in protein patterns (clusters) between different states of the organism. Consequently, the screening for markers of disease has been one of the principal objectives in a large number of proteomics studies (18, 82, 83, 135, 138, 139, 173, 176, 228, 251, 252). This application uses a limited segment of the potential power of proteomics, which should be able to evaluate coherently the complex changes in the proteome (or a significant segment of it) in multifactorial diseases. This implies not only a gain in knowledge due to the massive increase in the data acquired from one set of experiments at one time point, but also provides additional information compared with conventional approaches by yielding insights into the complex interactions among different proteins and pathways (190). This type of discovery is difficult to accomplish with reductionist methods and should improve our understanding of complex pathologies, like sepsis or acute lung injury, which involve
multiple and constantly interacting components of the immune system and signaling pathways (175).

There has been an expansion of proteomics into “functional proteomics,” the correlation of changes in the proteome with different states of the organism. This field is currently expanding in several different dimensions. “Protein profiling techniques” take a global view at complex protein samples, such as plasma. Given the complexity of these samples, these techniques need to be streamlined to achieve high throughput. The resulting protein patterns have diagnostic value as biomarkers on their own and indicate directions for more specific investigations. The application of protein profiling to tissue samples provides a combination of spatial information and protein profiles. The current results clearly indicate that these techniques are a valuable complement to histology (38, 265). The continuing improvement in protein identification will provide further insights into pathological processes and will most likely be especially valuable in cancer research. The application of mass spectrometry technology to the evaluation of “protein modifications” further extends the scope of proteomic analysis in depth. The physiological responses of an organism are only to a small part represented by changes in protein concentrations; especially, rapid responses to stimuli are transmitted by the modification of existing proteins. In spite of this complexity, this emerging field has, therefore, a large potential for clinically relevant research. The development of quantitative proteomics has widened the applicability of these techniques beyond a purely descriptive study design. Novel techniques in this field, namely differential gel electrophoresis (DIGE) and isotope-coded affinity tagging (ICAT), allow the direct comparison of samples, e.g., of different disease states.

CURRENT CHALLENGES

The main general issues that have impeded proteomics research in recent years include 1) difficulties in the detection of low-abundance proteins due to limitations in dynamic range, 2) identification of individual proteins within a complex biological sample, and 3) problems associated with the evaluation of all potentially useful information from the raw data. Typical samples in medical science are body fluids, such as plasma, urine, pleural fluid, bronchoalveolar lavage (BAL), pulmonary edema fluid, and cell lysates. These types of samples are complex mixtures of proteins with a dynamic range of protein concentrations of up to 10 orders of magnitude (2, 7, 214). The expression and modification changes in less abundant proteins (“low copy number proteins,” 10–1,000 copies per cell (25)) may be the most interesting ones. Their visualization is frequently obscured by highly expressed proteins [housekeeping proteins, >10,000 copies per cell (25)].

For example, plasma and pulmonary edema fluid contain large amounts of albumin (30–50 mg/ml in plasma and 20–25 mg/ml in pulmonary edema fluid) but comparatively small quantities of cytokines such as TNF-α or IL-1β (ng/ml to pg/ml range). Therefore, protein separation and purification techniques are key elements of proteome research that represent one of the major challenges (7, 15, 33).

Although the size of the proteome is unknown, the number of expressed proteins can be estimated from the open reading frames in a sequenced genome. It has been reported that 20% (1,484 proteins) from *Saccharomyces cerevisiae* (249) and >61% of the predicted proteome of *Deinococcus radiodurans* (145) could be identified by a current multidimensional chromatography-tandem mass spectrometric approach. These results indicate that identification of a significant part of the proteome of a cell is feasible.

Other common obstacles to proteomics are more dependent on the individual sample and the specific techniques. The validity of the results of a proteomic experiment is dependent on the initial sample, the purity of cell and protein isolation, and the subsequent sample fractionation steps. Salts, mucus, and other contaminants may require purification procedures that lead to loss of proteins of interest. The presence of proteases in samples can cause additional cleavages of the investigated proteins, complicating protein identification and quantitation. Ongoing cellular protein synthesis and posttranslational processing, by phosphatases and kinases, for example, can influence the results as well.

ANALYSIS METHODS

Initial approaches to investigate the proteome of cell lysates and body fluids were performed using 2D polyacrylamide gel electrophoresis (2D-PAGE) (34, 46, 61, 120). 2D-PAGE has been used in many studies to identify protein patterns in body fluids such as serum (7) or BAL fluid (55, 135, 138, 176). This technique is steadily improving and remains an essential part of many approaches to proteome analysis (201).

The rapid progress in mass spectrometry in the last decade has made it a key technique for the investigation of the proteome (2, 3, 29, 69, 103, 150). Mass spectrometry can be used to identify proteins by providing the molecular mass to electric charge (m/z ratio) of molecular species in a sample. Due to the high accuracy of this method, which under some circumstances can detect peptides in the femtomole to attomole range with an accuracy of <10 parts per million (ppm) (45, 70), it is now possible to identify proteins by using search algorithms that interrogate public “protein databases,” such as the nonredundant National Center for Biotechnology Information (NCBI) database, which can be accessed over the Internet.

Because the human genome is virtually known (243), every protein sequence can be predicted and included in these databases. Mass spectrometry is most often used as the identification technique after 2D-PAGE (46) or other separation techniques such as liquid chromatography (LC) (92, 93, 249).

Sample Preparation

Because many components of biological samples interfere with analysis, it is necessary to remove them before study. Insoluble substances can be removed by centrifugation. For 2D-PAGE and mass spectrometry, it is necessary to remove salts before analysis. This can be achieved by dialysis, size-exclusion filtering, protein precipitation, or reverse-phase chromatography (12, 54, 108). Frequently, abundant proteins such as albumin or immunoglobulins need to be removed first (7, 214). Complex samples need to be fractionated before analysis to obtain simpler subfractions and to decrease the dynamic range of components, if possible. For example, the dynamic range of concentrations in a plasma sample exceeds 10 orders of magnitude (7), whereas a current one-dimensional chromatography-mass spectrometry approach can only detect proteins in a dynamic range of approximately 4 orders of magnitude (7).
Electrophoresis methods in proteomics

<table>
<thead>
<tr>
<th>Analysis Method</th>
<th>Most Common Proteomic Applications</th>
<th>Advantages</th>
<th>Limitations</th>
<th>Examples, Additions, and Alternatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDS-PAGE</td>
<td>Sample separation before detection and quantitation techniques (Western blotting, MS)</td>
<td>Powerful separation technique if used after affinity purification</td>
<td>Limited separation power for LMW proteins due to loss of peptides at gel bottom (215)</td>
<td>Identification of proteins interacting with known nucleoporins in the nuclear pore complex of <em>Saccharomyces cerevisiae</em> (102). Proteins interacting with recombinant nucleoporins as bait were purified from yeast extracts, separated by 1D-SDS-PAGE, and identified by MS. Characterization of NMDA-receptor multiprotein complexes (NRC) isolated from mouse brain (105). Novel components of NRCs identified using a Western blotting screen for candidate proteins known to be involved with NMDA-receptor signaling and scaffolding: additional proteins identified by MS of the gel bands.</td>
</tr>
<tr>
<td>2D-PAGE</td>
<td>Mature technique with well-known strengths and limitations</td>
<td>2D-PAGE procedure and interface with MS difficult to standardize (65, 76)</td>
<td>Limited sensitivity for HMW and LMW proteins due to peptide loss at bottom and reactivity with carrier ampholites (215)</td>
<td>Most often used technique for the first dimension are gel strips that contain an immobilized pH gradient (IPG strips) (24, 202) The use of narrow-range IPG strips zooming in to a specific pH range in the sample has been shown to increase the discriminative power of 2D-PAGE cell lysates (88) and the recovery of low-abundance proteins in BAL (208), but requires more sample volume and time Paper bridge loading (208, 209), cup loading (100, 133, 209), combinations of denaturing with non-denaturing 2D-PAGE, and agarose electrophoresis (147, 148) have been proposed for high protein concentration samples Capillary isoelectric focusing (147, 148, 215) is an alternative for high-throughput analysis, since it can be combined online with chromatography columns (107, 176) and MS (148, 216). It can be fully automated, which should improve reproducibility (65, 107). The protein content of the sample can be visualized by UV absorbance (148). The resolution that can be achieved using this technique is currently inferior to 2D-PAGE (148). Capillary electrophoresis has recently been applied to lung airway surface fluid (50, 81) Chromatofocusing (235, 262), fractionation according to pI using analytical columns, has been employed as first step of a multidimensional chromatography-2D-PAGE-liquid chromatography-mass spectrometry approach to compare drug-treated and untreated human colon adenocarcinoma cells (263). The authors reported that over 1,000 protein bands could be detected in 0.2 pH fractions over a pH range of 4–7 Free-flow electrophoresis (126, 276) may be an option for protein separation; particles are deflected differently in an electric field according to their surface charge and are separated through a buffer flowing perpendicular to the electric field</td>
</tr>
</tbody>
</table>

Overview of current techniques of gel electrophoresis used for proteomics. One-dimensional SDS-PAGE (130) provides only one-dimensional protein separation according to molecular weight. It is most often complemented with other separation, quantification, or detection methodologies. Two-dimensional (2D)-PAGE (121, 178) provides both high-resolution sample separation and quantitative information if used as single technique. It has long been the principal method for the investigation of the proteome. In modern proteomics, it is most often combined with methods for protein quantification or identification, such as different stains and mass spectrometry. The most widely employed staining methods today are capable of detecting proteins in the $10^{-10}$ g range, and new technologies such as DIGE (Table 2) allow the comparison of up to 3 different samples in the same gel (75, 240, 241). LMW, low-molecular-weight; HMW, high-molecular-weight; pI, isoelectric point.
Table 2. Different staining methods for gel electrophoresis

<table>
<thead>
<tr>
<th>Stain</th>
<th>Sensitivity (grams of protein)</th>
<th>Dynamic Range</th>
<th>Endpoint Stain</th>
<th>Visualization</th>
<th>Price</th>
<th>Compatibilty with MS</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silver staining</td>
<td>$10^{-9}$</td>
<td>10-fold</td>
<td>No</td>
<td>Visual</td>
<td>$$5$</td>
<td>By replacement of glutaraldehyde with formaldehyde (decreases sensitivity)</td>
<td>Sensitivity</td>
<td>Certain proteins stain poorly, negatively, or not at all</td>
<td>(20, 78, 186, 201, 202, 215, 222, 238)</td>
</tr>
<tr>
<td>Coomassie brilliant blue</td>
<td>$10^{-8}$</td>
<td>10-fold</td>
<td>No</td>
<td>Visual</td>
<td>$$5$</td>
<td>Protein modification and ion suppression by trichloroacetic acid can be avoided by using phosphoric acid and methanol</td>
<td>Fast</td>
<td>Sensitivity</td>
<td>(20, 186, 187, 222)</td>
</tr>
</tbody>
</table>
| Colloidal coomassie staining  | $10^{-8}$                      | 10-fold       | Yes            | Visual       | $\$5$ | Radiation damage is a powerful approach to reduce the complexity of a sample by specifically isolating individual proteins “protein complexes” (15). These preparation steps are often more time consuming than the subsequent analysis steps and influence the sensitivity and discriminatory power of mass spectrometry-based protein identification (108, 191).

**Electrophoresis**

Gel electrophoresis, especially 2D-PAGE (121, 178), has long been the major method for the investigation of the proteome. An overview on the most frequently used electrophoresis techniques is provided in Table 1. For visualization, proteins in the gel are stained using a variety of different methods. A synopsis of the most widely employed staining methods is given in Table 2. With the use of this method, gel maps of body fluids, such as human plasma (6, 149, 194) or BAL fluid (176, 252), have been published (see Fig. 2). The large number of spots in a 2D gel is partly due to posttranslational and proteolytic modifications of proteins; one protein may, therefore, be present in several locations in the gel (25).
Although this phenomenon is potentially useful for the further analysis of these modifications, the increased number of spots for analysis can lead to additional effort, since >25% of the spots on a gel may be due to modified proteins (34) found elsewhere on the gel. The number of protein spots in complex samples makes computer-assisted image analysis necessary. Digital image analysis is also needed for quantitative information. There are several software suites for this purpose that are commercially available.

In modern proteomics, 2D-PAGE is most often used as a step before other protein detection techniques, especially mass spectrometry. However, although it has been shown that mass spectrometry can detect serially diluted, gel-embedded proteins down to the very low femtomole range (49, 70), generally 5–50 ng (corresponding to 100–1,000 fmol for a 50-kDa protein, an amount visible by silver staining) are considered necessary for successful mass spectrometry identification of proteins. Important reasons for this problem are the dynamic range of the current staining procedures (Table 2) and poor recovery of the peptides from the gels. It has also been shown that in 2D-PAGE, several classes of proteins are systematically underrepresented (Table 1); this limitation is relevant for many of the potential proteins of interest in pulmonary research (208). These shortcomings are constantly motivating efforts to improve 2D-PAGE and to find alternative methods to supplement or replace it (69).

Chromatography

Chromatography, especially LC, can be carried out as a purification step before or after 2D-PAGE (12, 163, 194). The progress in separation science has made this method a competitive alternative to electrophoresis. LC-LC-MS-MS (tandem mass spectrometry)-based techniques such as multidimensional protein identification technology (MudPIT) may have advantages over gel-based techniques in speed, sensitivity, reproducibility, and applicability to different samples and conditions (84, 144, 248, 249, 259, 260). The purification process of all LC techniques can be automated to a large extent (107, 137). The main shortcoming of this technique is the lack of quantitative information. The development of protein labeling techniques such as ICAT can overcome this disadvantage (see below).

Fig. 2. Silver-stained two-dimensional (2D)-PAGE image of human bronchoalveolar lavage (BAL) fluid samples. Isoelectric focusing was performed using pH 3–10 nonlinear immobilized pH gradient (IPG) strips; after equilibration, the second dimension was run on an Excel Gel XL 12–14%. A: gel pattern of a healthy subject. B: gel pattern of a patient with idiopathic pulmonary fibrosis (IPF). C: gel pattern of a patient with sarcoidosis. Gel spots were identified by matching with the human plasma reference maps and other published gel maps by NH2-terminal sequencing or by mass spectrometry. Bars and arrows indicate plasma proteins increased in the BAL fluid of patients with IPF (B) and sarcoidosis (C). Surfactant protein A (SP-A) is not present in the gel of the patient with IPF (B). Several small acidic proteins are upregulated in IPF and the matching spots in healthy controls are labeled by circles (108: cathepsin D, heavy chain; 172: epidermal fatty acid-binding protein (FABP-E); 174: cathepsin D, light chain; 179: intestinal trefoil factor; 183: FABP-E; 194: cathepsin D, light chain; 201, 202, and 206: calgranulin A; 207: saposin, D chain; 210: ubiquitin-like protein; 212: calcyclin; 216: calvasculin). MW, molecular weight. [From Noel-Georis et al. (176). Reprinted with permission from Elsevier.]
Another interesting set of approaches to visualize changes in the proteome content of a sample are the protein profiling techniques (37, 164) (see below).

**MASS SPECTROMETRY**

Traditionally, protein patterns in 2D-PAGE were identified by matching with a master 2D-PAGE pattern (e.g., SWISS-2DPAGE), with reference proteins (139, 141, 176, 252) or with Western immunoblots (83). Important progress in the identification of gel spots was made by the developed of automated NH$_2$-terminal (Edman) sequencing used in a large number of studies (4, 141, 251, 256). Mass spectrometry has rapidly replaced Edman sequencing for protein identification due to faster analysis times and much higher sensitivity (45, 46, 150). Mass spectrometry provides highly accurate measurements of the molecular weight and charge of the proteins or peptides in a sample. With the use of enzymatic digestion and peptide mass fingerprinting (see below), proteins can be identified even if they are truncated or posttranslationally modified. By adding a second mass analyzer (tandem mass spectrometry or MS-MS), the amino acid sequence of peptides can also be determined directly due to the fact that peptides fragment in a predictable fashion (22). After acquisition, the data are interrogated against protein sequence databases in an automated fashion (64, 217, 267) or interpreted manually (21, 46, 161).

**Types of Mass Spectrometers**

An overview of mass spectrometers currently being used for protein identification is provided in Table 3. The relatively soft ionization techniques of MALDI (117) and ESI (63) have made it possible to generate ions from large, nonvolatile analytes such as proteins without significant fragmentation. Both methods can be used to analyze proteins $\geq$100 kDa (2, 29). Their introduction in the late 1980s revolutionized the applicability of mass spectrometry to biomolecules and initiated an era of rapid progress that persists today (3).

There are several reasons for the popularity of MALDI mass spectrometers (Fig. 3) since their introduction in 1988 (117), which are summarized in Table 3. Recently introduced MALDI instruments include the MALDI-Qq-TOF (Q stands for quadrupole, TOF is time-of-flight mass analyzer) (218) and the MALDI-TOF-TOF (162). Both of these instruments are capable of analyzing the sequence of peptides by using two mass analyzers. Between the two TOF mass analyzers is a collision cell; peptide ions selected from the first mass analyzer are subject to collision with gas molecules resulting in vibronic activation, which induces dissociation processes. The second mass analyzer is used to measure the m/z ratio of the resulting fragment ions.

After its introduction (63), ESI (Fig. 4) soon established itself as an alternative to MALDI. To improve accuracy and deviate scanning of the second mass analysis step, a TOF analyzer has recently been used instead of the third quadrupole (Qq-TOF, Fig. 4) (150, 266). Other promising techniques are the protein profiling methods. Protein profiling is the rapid screening of samples by mass spectrometry with limited or no sample preparation. The resulting profile of m/z ratio peaks of different samples (which can be body fluids, cell lysates, or even tissue samples) can then be compared, and differences in the relative abundance of proteins can be identified. The samples are then further purified by chromatography and identified by techniques such as peptide fingerprinting or MS-MS. These techniques provide a complementary method to 2D-PAGE for protein visualization. For protein profiling, surface-enhanced laser desorption-ionization (SELDI) and imaging mass spectrometry (IMS) (30, 37–39) are currently being evaluated (Table 3).

**Protein Identification By Mass Spectrometry**

Mass measurements of the intact proteins provide a mass balance and rapid and valuable information on the protein profile of a sample. It is, however, not practical to attempt to identify a protein solely on the basis of its m/z ratio. This is mainly due to splice and sequence variation from database entries combined with a heterogeneous set of posttranslational modifications, which lead to variable differences in the molecular weight of a protein compared with the theoretical mass derived from the database. Therefore, additional strategies have been developed for protein identification, and these can be used separately or in combination.

“Peptide mass fingerprinting” is based on mass measurements of peptide fragments derived from a single protein. Before mass spectrometry, proteins are cleaved into peptides at specific, reproducible points in their amino acid sequence using chemical agents or enzymes. A protein covalent modification will only be reflected in one or a few of the peptide mass values, whereas the rest will remain unchanged. Because of its highly reproducible cleavage on the COOH-terminal side of arginine and lysine residues, trypsin is the proteolytic enzyme used most often. With the use of this specificity, the anticipated mass values of all peptides in virtual digests of all proteins in the database are calculated. The protein identity is determined by comparing the measured peptide mass values with those calculated (45, 98, 110, 151, 208, 268). The reliability of peptide mass fingerprinting is dependent on: 1) the mass accuracy of the peptide measurements (45); 2) the number of matched vs. unmatched peaks in the spectrum; 3) the number of peptides that could be matched to a single protein; and 4) the number of proteins that are present in the digested sample, since random matches can occur at a level of confidence similar to real matches in complex mixtures. The decreased reliability of results using peptide fingerprinting with complex mixtures of proteins has been exacerbated by the massive increase in the size of the databases. Other potentially critical factors are the increased rate of false-positive matches and bias toward high-molecular-weight proteins, which yield a larger number of peptides and are, therefore, more likely to be matched by this technique than smaller proteins. Scoring systems included in the analysis software packages (see below) aim at compensating for these potential problems.

With the use of the two sequential mass analyzers (tandem mass spectrometry or MS-MS), primary structural analysis of the amino acid sequence can be obtained (3, 22, 150, 161) by fragmenting one or more of the peptides (Fig. 5). Peptide fragmentation is achieved by preferential cleavage of the backbone bond of polypeptides upon collisional activation with a gas [collision-induced dissociation (CID)] (21, 161). Tandem mass spectrometry can be carried out using both ESI (e.g., ESI-triple-quadrupole or ESI-Qq-TOF) (42) and MALDI ionization (MALDI-TOF-TOF) (102, 162). Often, fragmentation
## Table 3. Overview of biological ionization techniques in mass spectrometry

<table>
<thead>
<tr>
<th>Technique</th>
<th>Description</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrospray ionization (63)</td>
<td>Proteins are solubilized and ionized at atmospheric pressure by pumping the solute through a capillary at high voltage. As the micromolar-sized droplets enter the mass spectrometer, solvent is removed by heat or energetic collisions with a gas and impart charge onto analyte molecules (Fig. 4). Proteins and peptides acquire multiple charge states. For proteomics, ESI is usually used with ion traps, FT-ICR, or Qq-TOF instruments.</td>
<td>Can be interfaced online with chromatographic separation strategies (e.g., an LC unit), since it performs continuous measurements of peptides from the liquid phase. ESI on a FT-ICR or Qq-TOF is more accurate than the MALDI instruments.</td>
<td>More susceptible to contamination. Capillary can become blocked. Requires more maintenance.</td>
</tr>
<tr>
<td>Matrix-assisted laser desorption/ionization (117)</td>
<td>The sample is cocrystallized with the matrix on a metal slide (target). The target is positioned in front of the ion source. A laser pulse irradiates each spot, causing a rapid excitation of the matrix and ejection of matrix and analyte ions into the gas phase. The ions are accelerated by an electric field that directs them toward a mass detection unit (Fig. 3). Proteins and peptides usually acquire single charge. MALDI is usually used with TOF, Qq-TOF, TOF-TOF, or ion trap instruments.</td>
<td>Sensitivity in combination with a TOF analyzer: current MALDI-TOF mass spectrometers require less than a femtomole of peptide sample for molecular weight analysis (150). A mass accuracy of a few parts per million can be obtained (150) with internal calibration. MALDI is most often coupled with TOF analyzers, which are highly sensitive and have a wide mass range (2). Robustness to low levels of salts, buffers, and other compounds often used in protein chemistry (2). Spectra are relatively easy to interpret, since they consist almost exclusively of singly charged ions (2, 16).</td>
<td>Samples need to be spotted on metal targets and dried; therefore, online measurements with HPLC possible (LC separation can be performed offline). Fragmentation of singly charged peptides can be less informative.</td>
</tr>
<tr>
<td>Surface-enhanced laser desorption/ionization</td>
<td>Modification of MALDI, in which the target is a protein chip array composed of various chromatographic, immunologic, or enzymatic surfaces (109, 164, 189, 193). Several different probe surfaces are available (reverse phase, ion exchange, IMAC, and others) (164). Using different chip arrays, SELDI can be modified for use with proteins of cationic, anionic, hydrophobic, hydrophilic, and other properties (164).</td>
<td>For surface-enhanced affinity capture, a 100-fold dynamic range was reported (164). High throughput; methods may be useful in the discovery of new drug targets (271) and biomarkers (109, 164, 189, 193).</td>
<td>Difficult to identify the proteins that contribute to the mass profile (56), since peptide fingerprinting and MS-MS requires purification and enzymatic digestion steps that are difficult to integrate with this approach (56). This may possibly improve in the future, since the targets can now be used with a Qq-TOF mass spectrometer (164).</td>
</tr>
<tr>
<td>Imaging mass spectrometry</td>
<td>IMS utilizes MALDI-MS for the direct analysis of tissue samples (37). A 1-μM slice of frozen tissue is coated with crystallization matrix or the tissue is blotted on a target coated with C18 beads (30, 37–39).</td>
<td>Successfully used to identify disease markers in human glioblastoma (235) and nonsmall cell lung cancer (265). IMS generates ion images of samples and maps them to the original sample, combining spatial information of peptide/protein distributions (Fig. 6).</td>
<td>Still difficult to identify proteins in the profile. Potential overlay of different cell layers in tissue specimens.</td>
</tr>
</tbody>
</table>

All mass spectrometers consist of three main parts: 1) an ion source, which converts molecules into gas-phase ions; 2) a mass analyzer that separates these ions according to their mass-to-charge (m/z) ratio using a physical property such as deflection in electric or magnetic field or the time of flight; and 3) an ion detector, which produces a current whose magnitude is proportional to the number of ions reaching the detector at a given time (62, 96). The relatively soft ionization techniques of electrospray ionization (ESI) (63) and matrix-assisted laser desorption/ionization (MALDI) (117) are used in the vast majority of mass spectrometry studies of biological samples. The mass analyzer measures the m/z ratio of the analytes. For biomedical research, analytes can be proteins, oligosaccharides, lipids, nucleotides, peptides, or small molecules such as metabolites or hormones. There are three main types of mass analyzers in widespread use for biomolecules: separation of the ions by time-of-flight (TOF), by oscillating electric fields between conduction rods (quadrupole, Q), or by selective ejection of ions from a 3D ion trapping field (ion trap) (23, 114, 150, 155, 160, 187, 218). A simple nomenclature is generally used to specify mass spectrometers according to these components (e.g., MALDI-TOF, ESI-Qq-TOF). For sequence analysis, 2D mass analyzers can be assembled sequentially. In recent years, several hybrid optical designs have complemented the common triple quadrupole tandem mass spectrometer, such as the Qq-TOF (162), TOF-TOF (134), Q-TRAP (48, 156, 157). The ions are detected by an ion multiplier in all but the last types of mass spectrometers (76, 254). Recently, the protein profiling techniques surface-enhanced laser desorption/ionization (SELDI) and imaging mass spectrometry (IMS) have been introduced. Their technology is based on MALDI ionization. Protein profiling refers to the screening of samples by mass spectrometry with limited or no sample preparation. The resulting profiles of m/z ratio peaks of different samples can be compared, and differences in the relative abundance of proteins can be identified.
spectra of only a few peptides are sufficient for unambiguous protein identification (45, 150).

Although sequence information can also be obtained with relatively inexpensive instruments using the metastable decay of some ions after desorption by MALDI (postsource decay), this time-consuming technique is rapidly being replaced by the faster and more sensitive tandem time-of-flight mass spectrometry (102, 150, 162, 266).

**Protein Profiling Techniques**

Protein profiling is the rapid screening of samples by mass spectrometry with limited or no sample preparation. The resulting profile of m/z ratio peaks of different samples (that can be body fluids, cell lysates, or even tissue samples) can then be compared, and differences in the relative abundance of proteins can be identified. The samples can then be further purified by chromatography and identified by techniques such as peptide fingerprinting or MS-MS. These techniques provide a complementary method to 2D-PAGE for protein visualization.

In SELDI (Table 3), proteins are retained on a protein chip array composed of various chromatographic, immunologic, or enzymatic surfaces and subsequently detected directly by time-of-flight mass spectrometry. In contrast to the metal sample target employed in MALDI mass spectrometry, in SELDI the probe surfaces play an active role in the extraction, structural modification, and presentation of the protein of interest from the sample. There are several different probe surfaces available, thus SELDI can be modified for use with proteins of different properties (164). Of the different SELDI applications in development today, surface-enhanced affinity capture is considered the most promising, with a reported 100-fold dynamic range (164). The special advantage of this technique is the possibility of high-throughput analysis. Protein chips may be useful in the discovery of new drug targets (271) and biomarkers (109, 164, 189, 193).

IMS utilizes MALDI-MS for the direct analysis of tissue samples (37) (Table 3). This is carried out by coating a slice of frozen tissue with crystallization matrix or by blotting the tissue on a target coated with C18 beads (30, 37–39). Mass spectrometry generates ion images of samples providing the capability of mapping specific molecules to 2D coordinates on the original sample, thus giving spatial information on peptide/mass peptide/protein identification.
One-letter nomenclature. Amyloid A, is given. Amino acids are labeled according to the standard sequence, which could be attributed to a peptide fragment of serum MS-TAG from the ProteinProspector suite of programs (45, 102). The obtained sequence was automatically interrogated against the NCBI database using MS spectrum.

There are only a few components detected in this MS spectrum. The resulting ions can be separated into y and b ions according to the retention of the charge on the COOH-terminal or the NH2-terminal fragment (111, 206). The data were automatically interrogated against the NCBI database using MS/MS from the ProteinProspector suite of programs (45, 102). The obtained sequence, which could be attributed to a peptide fragment of serum amyloid A, is given. Amino acids are labeled according to the standard one-letter nomenclature.

Protein distributions (Fig. 6). This technique has been successfully applied to brain tumors (233) and non-small cell lung cancer (265); the latter study is described in more detail later in this article. This methodology will certainly continue to be increasingly utilized.

Analysis of Protein Modifications

Posttranslational modifications play a crucial role in cell signaling and protein function (77, 152, 190). More than 200 different protein modifications have been described (125, 257, 258). Important posttranslational modifications include phosphorylation, acetylation, glycosylation, ubiquitination, and nitration (125, 152, 242). The analysis of posttranslational modifications on a proteome scale is still considered an analytical challenge (66, 69, 152, 159, 177, 229, 274); reasons for this are the fragility of the chemical bonds of many protein modifications upon sequencing by CID, signal suppression of negatively charged (phosphate-, sulfate-containing) molecules in the commonly used positive detection mode, and difficulty of obtaining full-sequence coverage (123). Moreover, most modifications are substoichiometric; therefore, modified peptides are frequently present at much lower levels than unmodified peptides (124, 269).

Phosphorylation is an important regulation mechanism of protein activity and signaling networks. It is crucial in protein kinase activation, cell-cycle progression, cellular differentiation, transformation, response, and adaptation of peptide hormones (47, 77, 154, 165). Approximately 30% of all mammalian proteins are phosphorylated at any given time (153). The more than 500 protein kinases and ~100 phosphatases have relatively wide substrate specificities and work in different combinations to achieve a variety of biological responses, which can make analysis of these complex networks challenging (47, 153, 154). Phosphopeptides are generally difficult to analyze by mass spectrometry. One reason for this is their negative charge, which reduces ion intensity (electrospray is generally performed in the positive mode). Other impediments include their presence at substoichiometric levels, their hydrophilicity, which interferes with reverse-phase chromatography, and other factors (8, 124, 153, 221, 269). Currently, phosphorylation is evaluated most often by labeling a previously defined protein with 32P-inorganic phosphate followed by 2D-PAGE and/or reverse-phase chromatography, which is a relatively complex, time-consuming procedure (124, 152, 153, 269). For example, in a recent comprehensive study (182), the regulatory mechanisms controlling the activity of 3-phosphoinositide-dependent protein kinase-1 (PDK1), which plays a central role in signal transduction pathways that activate phosphoinositide 3-kinase, were evaluated. With the use of site-directed mutants, phosphorylation on Tyr373/Tyr376 was shown to be important for PDK1 activity, whereas phosphorylation on Tyr9 had no effect. Other novel approaches to investigate phosphorylation include the 14N:15N labeling of immunoprecipitated phosphoprotein-isotope-coded affinity tag method (79, 80), the use of immobilized metal ion affinity chromatography to affinity capture phosphopeptides (95, 196, 238, 261), and the chemical transformation of phosphoserine and phosphothreonine residues into lysine analogs that are then cleaved with a lysine-specific protease to map sites of phosphorylation (123).

In response to various inflammatory stimuli, lung endothelial cells, alveolar and airway epithelial cells, and activated alveolar macrophages produce nitric oxide and superoxide, products that may react to form peroxynitrite. Peroxynitrite can nitrate and oxidize amino acids in various lung proteins, such as surfactant protein A (SP-A), and inhibit their function. It has been shown that the nitration and oxidation of a variety of alveolar proteins is associated with diminished function in vitro; in addition, both modifications have been identified in proteins sampled from patients with acute lung injury using immunoassays (132, 275). The selective nitration of tyrosine residues in different cytoplasmatic high-molecular-weight proteins and histone proteins in murine tumor cells by neutrophil...
has been demonstrated by Western blotting and mass spectrometry in vivo and in vitro (94). The authors found that histone nitration was relatively stable, making it a potentially useful marker for extended exposure of cells or tissues to nitric oxide-derived reactive species.

Novel methodologies for the evaluation of other protein modifications are available as well. N- and O-linked glycosylation occurs throughout the entire phylogenetic spectrum and plays key roles in reactions in the endoplasmic reticulum, Golgi apparatus, cytosol, and nucleus (53, 227). Glycosylation is present especially on proteins destined for extracellular environments (207); consequently, many therapeutic targets and clinical biomarkers are glycoproteins. For example, CFTR is an integral membrane glycoprotein that normally functions as a chloride channel in epithelial cells (210). The most common mutation in cystic fibrosis, ΔF508, results in mislocalization and altered glycosylation of CFTR. Moreover, altered fucosylation and sialylation of both membrane and secreted glycoproteins occur in cystic fibrosis, and the two major bacterial pathogens causing chronic infection in the cystic fibrosis lung, *Pseudomonas aeruginosa* and *Haemophilus influenzae*, have binding proteins that recognize these altered sites. For the investigation of protein glycosylation, mass spectrometry has been widely used (28, 53, 129) in the last years, especially the Qq-TOF instrument (Fig. 4) (35, 53, 232). In a recent study (270), glycoproteins were conjugated to a solid support by hydrazide chemistry, and glycopeptides were labeled with stable isotopes. Subsequently, the formerly N-
linked glycosylated peptides were specifically released using peptide-N-glycosidase F and identified and quantified by MS-MS. The methodology has been used to investigate plasma membrane and serum proteins.

A rapidly evolving part of functional proteomics is the investigation of specific protein complexes (67, 68, 264). Protein complexes can be isolated from complex mixtures by affinity extraction techniques such as direct antibody coprecipitation (5) or indirect tagging of the bait protein with an epitope that is then recognized by an antibody using tandem affinity purification tags. (72, 205). Chemical cross-linking can be used to prevent the loss of components from the protein complex during precipitation (213). Affinity purification techniques for the analysis of protein complexes have been reviewed (15, 264). The resulting isolated complexes are subsequently analyzed by mass spectrometry. A more general approach is the comprehensive identification of proteins in macromolecular complexes after separation by liquid chromatography (144).

**Quantitative Proteomics**

With the use of tandem mass spectrometry, the sequence of one peptide can be sufficient to identify an entire protein. This simplification of protein identification has triggered the development of methods that aim at increasing throughput by performing protein separation and identification in one suite of experiments (87). Because cutting out individual gel spots from a 2D gel is a very time-consuming procedure, many recently introduced approaches use chromatography for sample separation. These techniques either couple LC directly to ESI-MS-MS or robotically spot the chromatographically separated fractions to a MALDI target. However, 2D-PAGE provides quantitative information that has only been obtained to a very limited extent from mass spectrometry-based methods. The lack of quantitative results is obviously a serious shortcoming that would limit a LC-mass spectrometry approach to a purely descriptive study design. The use of isotope ratio mass spectrometry (IRMS) is one method being used to close this gap.

Currently, the most advanced IRMS technique is the ICAT technology (89). In an ICAT experiment, the reduced cysteine residues of proteins are labeled differentially. The two different tags consist of an iodoacetamide group that reacts with the free cysteine, a biotin tag that can be used for affinity purification of labeled peptides, and a linker region containing the different isotopic labels. The light version and the heavy version differ in eight protons within the linker region of the ICAT reagent that have been substituted with eight deuterons in the heavy version. The two samples can be discriminated by mass spectrometry according to this mass difference of 8.0 Da (89). After being labeled, the two samples are pooled and digested with trypsin. The tagged peptides are then extracted with an avidin-containing column. Because only cysteine-containing peptides are evaluated, the complexity of the sample is reduced by more than one order of magnitude (89). The frequency of cysteine residues in proteins varies slightly from species to species and averages \(\sim 1\%\) (27). In yeast, \(~9\%\) of all theoretically possible peptides after tryptic digestion contain cysteine (89).

A disadvantage of ICAT is that no absolute concentrations of proteins are measured and that comparisons of the expression of two different proteins are not possible. Another shortcoming is the low-sequence coverage, since only cysteine-containing peptides are labeled. The applicability of ICAT to the analysis of posttranslational modifications or protein isoforms is therefore limited (186). This restriction of ICAT to cysteine-containing peptides can be partially overcome by separate analysis of the unlabeled peptides that are not captured in the affinity chromatography step. However, quantitative information will not be available in this case unless a corresponding ICAT-labeled peptide is identified for the same protein (144). Another potential problem is that the differentially labeled peptides can separate from each other during the chromatography process because deuterium affects the retention time in reverse-phase chromatography. Consequently, they may be ionized at separate time points and eventually in different fractions, which can lead to different quantitation intensities (272). In addition, the ICAT tag is relatively large, which may interfere with the detection of large peptides (186). Furthermore, the dynamic range for the quantification of different expression levels of one protein is relatively small (~10-fold) (9, 89), which is inferior compared with fluorescent dyes (186). Some of these limitations can be overcome by using a newly introduced cleavable ICAT reagent. The new reagent utilizes \(^{13}\)C with a mass difference of 9 Da between the heavy and the light marker. The advantages are a smaller tag (227 Da compared with the 442 Da of the original ICAT), which interferes less with the analysis of larger peptides, a mass difference that can easily discriminate a peptide with two ICAT labels (2 cysteine residues) from the common oxidation of methionine, and a reduction of CID fragmentation.

**Fig. 7. Comparison of 2 different samples with the isotope-coded affinity tagging (ICAT) method.** These mass spectra from the pooled membrane fractions of pulmonary alveolar type II cells from rats with and without ventilator-induced lung injury (low- and high-molecular-weight isotope envelope, m/z at 755.34 and 759.85, respectively) were identified as lysozyme 1 precursor by sequence analysis using tandem mass spectrometry as shown in Fig. 5B. The mass difference between the 2 differentially labeled forms of the peptide is the result of labeling with the cleavable ICAT reagent (see text). Because the ion is double charged, the actual difference between the low and the high peaks (circled numbers) is only 4.5 m/z units (amu). The 2 monoisotopic peaks have different intensities, which is correlated to the expression of the corresponding protein in the 2 differentially ventilated animals. The difference can be quantified by calculating the ratio of the areas under the 2 peaks.
byproducts, which improves the quality of the resulting mass spectra (93).

The ICAT technique has successfully been employed for the labeling of membrane protein extracts in prostate and breast tumor cell lines (10). Another recent study (220) compared differences in the expression of protein patterns between rat cells that did or did not contain the myc oncogene. These authors reported expression differences among functionally related proteins in myc-positive cells, such as induction of protein synthesis pathways, upregulation of anabolic enzymes, and reduction of proteases, and changes in the levels of adhesion molecules, of actin network proteins, and Rho pathway proteins that correlated with the known qualities of myc-positive cells. Another interesting application of ICAT was a comparison of the microsomal fraction of cells from the human myeloid cell line HL-60 with and without the induction of differentiation by phorbol 12-myristate 13-acetate; the authors identified and quantified 491 proteins. One example of quantitative analysis of alveolar type II cells using the cleavable ICAT technique from our research is given in Fig. 7 (93). The method is an active area of research and development (86, 92, 223, 224).

A comparison of the coverage of the known 80 ribosomal proteins from the 80S mammalian ribosome by ICAT and 2D-PAGE showed that 35 could be found by ICAT (92, 186), whereas a highly elaborate 2D-PAGE system specifically tailored to the detection of ribosomal proteins was able to detect 55 proteins. A standard 2D-PAGE approach found only two ribosomal proteins (71, 186). ICAT and 2D-PAGE are different methodologies that have different biases and that frequently detect different segments of the proteome of the same sample (186). The only study that has used a combination of the two methods made use of the observation that proteins labeled with light and heavy forms of the ICAT reagent comigrate during 2D gel electrophoresis. Therefore, two or more labeled samples can be analyzed concurrently in the same gel (223), which may be useful for the quantitative and qualitative analysis of differentially expressed or posttranslationally modified proteins. For protein quantification, a larger number of gels might be necessary compared with a 2D-PAGE approach with DIGE (75). Protein modifications can lead to the presence of one protein in several different spots on the gel; to compare samples, it is, therefore, either necessary to run three gels (1 with each sample separate and 1 with the samples combined) or to quantify all spots on the gel containing the combined samples using ICAT (223).

DATA ANALYSIS AND INTERPRETATION

Proteomics and Bioinformatics

Given the complexity of the proteome, an adequate proteomics approach requires the identification of thousands rather than several or a few proteins at a time (76). Therefore, bioinformatics plays a key role in proteomic studies and is often the rate-limiting step (183, 246). The data obtained from mass spectrometry must be interpreted by interrogation against protein databases, the quality of which is crucial for protein identification. Both peptide masses and peptide sequence information can be used for protein identification. There are several protein databases readily available over the Internet that differ in the frequency with which they are updated and the amount of redundancy. Currently, the most complete and most frequently updated database is provided by NCBI, which is a combination of several databases, including Swiss-Prot and Owl. Consequently, this database also contains the most redundancy of protein entries.

Several software packages are available for the analysis of mass spectrometry data. They interrogate the obtained peptide or sequence data against the protein databases and rank the results according to a scoring system (often-used scoring algorithm, Molecular Weight Search (MOWSE) (181)). Software packages include Mascot from Matrix Science (London, UK; http://www.matrixscience.com) (192), ProFound from Rockefeller University (http://prowl.rockefeller.edu) (273), ProteinProspector, a software suite developed at the University of California, San Francisco (http://prospector.ucsf.edu) (45), the SEQUEST algorithm developed at the University of Washington (http://thompson.mbt.washington.edu/sequest) (60), and others (2). Each of these programs provides additional utilities; for example, ProteinProspector includes additional tools for the interpretation of mass spectrometry, MS-MS, and ICAT data (at present not included in the public Internet version) as well as a batch mode for repetitive tasks and other analysis tools.

Another bioinformatics challenge is the analysis and description of the large amount of information into a comprehensive model. This includes the development of methods for data comparison between different research groups (183) and the integration of gene ontologies (10).

INVESTIGATING THE LUNG PROTEOME

Proteomics research often focuses on the investigation of either body fluids or specific cell types. Because the lung is the site of several different biological processes, the interpretation of proteome experimental results must take into account potential contamination from pathogens as well as the contributions of the different cell types in the lung.

During the development of proteomics over the last two decades, there have been numerous attempts to apply proteomic methodologies to pulmonary medicine. These shall be briefly reviewed in this section.

EXPERIMENTAL DESIGNS

Classic, reductionist studies, e.g., an ELISA or a Western blot analysis, will most often provide relatively simple answers to the initial scientific question, e.g., the presence of a specific protein or a concentration change. Moreover, since the researcher has to decide beforehand on which antibodies to use, there is a need for a specific hypothesis of the potential reactions. On the other hand, the investigator will most likely only find answers to questions conceived of beforehand. This is not the case in proteomics experiments. These studies are likely to provide answers even if these questions have not been thought of in the initial scientific question or hypothesis. Thus proteomics experiments have the advantage that the results are less biased by the theories or beliefs of the investigator and only limited by the sensitivity of the method. For this reason, proteomics results have a high potential to give rise to new discoveries and generate new hypotheses. Moreover, proteomics experiments are less likely than reductionist methods to mask eventual weaknesses in the initial experimental design under the cover of an apparently simple, clear-cut result. To
Invited Review

THE PROTEOME OF BAL FLUID

Evaluation of BAL fluid has been useful in diagnosis and research of several inflammatory lung diseases, including emphysema, pulmonary fibrosis, cystic fibrosis, pulmonary transplantation, and acute lung injury. Early proteome investigations of BAL fluid done to investigate alveolar proteinosis resulted in a 2D-PAGE database of normal BAL fluid published in 1979 (18). In this study, as well as a subsequent study of BAL fluid from smokers and nonsmokers (17, 18), by pattern matching, most of the proteins found in BAL fluid could be identified as serum proteins. The authors found 23 serum derived-proteins, which accounted for 97% of the protein content of normal BAL fluid. The study identified significant differences in the BAL proteome of smokers, who had increased levels of IgG, C4, and C3 and decreased α2-thioglycoprotein, α1-acid glycoprotein, and Gc-globulin. In 1990, Lenz and colleagues (136) published a method for 2D-PAGE of BAL fluid from dogs and then compared protein patterns in BAL fluid proteins from patients with idiopathic pulmonary fibrosis, sarcoidosis, and asbestosis with normal controls (135). In idiopathic pulmonary fibrosis, the spot intensity of one surfactant-associated protein, SP-A, was decreased, whereas in sarcoidosis, the immunoglobulins (IgG, IgA) were increased. Another group of protein spots with a molecular weight of 55 kDa and one spot with a molecular weight of 12 kDa were identified. Compared with normal samples, the number and intensity of low-molecular-weight proteins were significantly increased in patients with asbestosis and, in some cases, in patients with idiopathic pulmonary fibrosis and with sarcoidosis.

At the time of this early proteomics research, many of the characterized spots could not be identified. Although the results of these studies provided the first information for a basic understanding of the protein composition of BAL fluid, the value of these results for clinical medicine was limited. Since then, gradual progress in staining and imaging techniques and improvements in standardization have made it possible to identify the most abundant proteins and refine the information on proteomic changes in different disease states. In 1995, Lindahl and coworkers (142) evaluated the BAL fluid proteome in patients after occupational exposure to irritating chemicals. They defined >1,000 protein spots. Plasma proteins were identified by pattern matching. After occupational exposure, 14 protein spots were increased, and one spot decreased by a factor of more than 3 compared with the levels before exposure and in healthy individuals. Subsequently, the same group found higher levels of basic proteins in smokers than in nonsmokers, whereas subjects exposed to asbestos had increased amounts of several high-molecular-weight and basic proteins (138). The results of protein identification showed lower levels of albumin and higher levels of immunoglobulins in smokers than in nonsmokers, whereas the levels of transferrin were higher in asbestos-exposed subjects. Further progress in the proteomic analysis of BAL fluid was boosted by the development of the SWISS-2D-PAGE database containing compiled maps of human BAL fluid (139, 251, 252). The current master gel of BAL proteins encompasses >1,200 spots visualized by silver staining (Fig. 2) (176). Information is available on changes in 2D-PAGE protein patterns of BAL for smoking (17, 135, 138, 139, 141, 143, 176, 252), sarcoidosis (135, 138, 139, 176, 251, 252), idiopathic pulmonary fibrosis (135, 138, 139, 176, 251, 252), lupos erythematos (251), Wegener’s granulomatosis (251), hypersensitivity pneumonitis (135, 138, 139, 176, 252), lipid pneumonia (251), chronic eosinophilic pneumonia (251, alveolar proteinosis (18), bacterial pneumonia (251), other infections, malignancies and immunosuppression (82, 173), cystic fibrosis before and after α1-antitrypsin treatment (83), and asbestosis (251).

The application of narrow-range immobilized pH gradient (IPG) strips can further increase the resolution of 2D-PAGE (208). Interestingly, the improvement in protein spot detection has been shown to be more significant for the protein spots present exclusively in BAL (55%) than for the spots present in both BAL and serum. This finding suggests that many of the BAL fluid-specific proteins, which are likely to be of pulmonary origin, are low-abundance proteins.

Improvements in protein identification increased the clinical relevance of 2D-PAGE studies. Three years after their initial studies, Lindahl and coworkers (141) published a more detailed report on the changes in BAL and “nasal lavage fluid” 2D patterns. Using Edman sequencing and pattern matching with the Swissprote database, Lindahl and coworkers (141) found five previously unidentified protein spots. The proteinase inhibitor lipocalin was significantly reduced in the nasal lavage fluid of asthmatic patients, and two isoforms of the cysteine proteinase inhibitor cystatin S were significantly reduced in the nasal lavage fluid of smokers. Other proteins that were identified were transthyretin, immunoglobulin binding factor, and a previously undescribed 11-kDa fragment of albumin. All of these studies, although promising, demonstrated the limitations of the time-consuming NH2-terminal (Edman) sequencing technique for protein identification from gels with a large number of spots. This long-standing problem has only been overcome recently with the applicability of mass spectrometry to real biological samples. Especially in combination with the narrow-range IPG strip technique, the sensitivity of the 2D-PAGE-mass spectrometry approach has been substantially increased (208).
Other factors that have complicated the analysis of the proteome of the epithelial lining fluid are the highly variable dilution factor (104, 112, 204), the wide dynamic range of protein concentrations, and the high salt concentration of BAL fluid, which further increases with sample concentration. In this field, progress has been made with approaches that specifically address the known problems of BAL fluid analysis, such as prefractionating and desalting protein samples before mass spectrometry by HPLC, which can be done before or after the 2D-PAGE step (12, 163), or capillary electrophoresis (50, 81). It is now also possible to skip the 2D-PAGE step altogether, especially for the analysis of BAL fluid (176), in favor of HPLC-based techniques. New methods for protein quantification, such as ICAT, provide alternatives to 2D-PAGE for the quantification of differences between samples.

In conclusion, long-standing difficulties in the identification and quantification of proteins of lower abundance have been responsible for results with limited clinical applicability in the proteomic analysis of BAL fluid in the past. The cited studies reflect the considerable progress that has been made in the last two decades. The identification of concentration changes of prognostic markers such as SP-A (41) or inflammatory mediators such as calgranulin A (119) in the current 2D-PAGE master gel (176) are clinically relevant results that give rise to further insights in underlying pathological processes. The current state of the art in proteomic research is clearly still far away from the goal of a coherent representation of the protein content of BAL fluid, and even more effort will be necessary for accurate quantification and the evaluation of protein modifications. However, recent developments in several important sample preparation steps give reason to believe that, after a long method development period, there may soon be further breakthroughs in the proteomic analysis of BAL fluid.

ALTERNATIVES TO BAL

Bronchoscopy is an invasive procedure that cannot be performed in all patients. The advantage of “plasma measurement of specific lung proteins” is that blood samples are readily accessible. Lung-specific proteins, such as surfactant proteins A, B, and D, are elevated in plasma in several disease states (41, 58, 128). In theory, plasma should contain a large part of, if not all, human proteins (214) and should, therefore, be an ideal target for a proteomics approach. However, the dynamic range of concentrations in plasma is even greater than in BAL fluid \[\sim 10^{10} – 10^{12} \text{ (214)}\], and the concentration of pulmonary proteins in plasma is usually relatively low (58, 128, 148). Consequently, this approach may only be useful to evaluate changes in a small subset of the lung proteome.

The “induction of sputum with hypertonic saline” avoids bronchoscopy and has been applied to children (113) and to adult patients with asthma (226), cystic fibrosis (97), tuberculosis (6), and interstitial lung disease (179). Up to now, there have been no proteomic studies on induced sputum samples, partly because the high concentrations of salt and contaminants make purification and 2D-PAGE of these samples difficult.

“Direct aspiration of pulmonary edema fluid” is possible in patients who suffer from acute respiratory failure due to cardiogenic pulmonary edema or acute lung injury. The use of pulmonary edema fluid can avoid some of the dilution and concentration problems that are associated with the use of BAL fluid. Pulmonary edema fluid has been used to characterize hydrostatic pulmonary edema and acute lung injury in a large number of studies (1, 11, 43, 57, 73, 115, 127, 158, 166, 174, 198, 222, 244, 275). Comparison of the proteomic profile of lung injury vs. hydrostatic edema fluid provides new biological protein markers with diagnostic or potential therapeutic value (101). Figure 8 provides an overview of the different summarized proteins in human pulmonary edema fluid (101).

The study of nasal lavage fluid offers an alternative approach to investigate lung diseases. The proteome of nasal lavage fluid has been characterized by Lindahl and coworkers (139–143) by 2D-PAGE, identifying proteins by matching with reference proteins, Western immunoblots, and Edman sequencing. Many of the protein spots that could be identified in nasal lavage fluid could be assigned to proteins that are also in BAL fluid and plasma. Nasal lavage fluid expression changes have been demonstrated in levels of lipocalin-1, cystation S, transthyretin, and IgBF in individuals that smoke or suffer from upper airway irritation or asthma (141). In a more recent study from the same group, the gel protein pattern of human nasal lavage fluid was further characterized using MALDI-TOF mass spectrometry and sequence analysis by postsource decay after 2D-PAGE. Decreased levels of Clara cell secretory protein, a truncated variant of lipocortin-1, three acidic forms of \(\alpha_1\)-proteinase inhibitor, and one phosphorylated form of cystatin S were found in smokers (74). A new marker of airway irritation in asthmatic patients, nasal epithelial clone protein, was discovered by this group using the same method (140).

An interesting source for lung proteome studies may be “frozen condensates of exhaled breath” (211). A variety of measurements could be obtained from frozen breath condensates in previous studies: carbon monoxide and nitric oxide metabolites were analyzed in smokers and chronic obstructive pulmonary disease (COPD) (13, 169, 170), inflammatory cy-

Fig. 8. Overview of the different summarized proteins in human pulmonary edema fluid after depletion of albumin and IgG (101). Other immunoglobulins and highly abundant plasma proteins, such as transferrin, account for a substantial percentage of the proteome of pulmonary edema fluid. Other important constituents are proteases and protease inhibitors, innate immunity proteins, and complement factors and secretory proteins and cytokines. Cellular constituents, such as ion channel proteins, transcription factors, and ribosomal proteins are most likely a sign of cellular turnover.
tokines in patients with different pulmonary diseases vs. healthy controls (211), isoprostan in asthmatic patients (169), smokers (168), patients with COPD (168), and patients with acute lung injury (31), as well as hydrogen peroxide in COPD patients (52). Due to the noninvasiveness of the procedure, it can be applied to a wider spectrum of patients than BAL. 2D-PAGE maps of the exhaled proteins have been published (211). Proteomic investigations of other specimens, such as pleural fluid, have yet to be done.

The protein content of body fluids is influenced by a large number of different factors, such as influx of plasma proteins, dilution, and protein turnover by degrading enzymes and oxidants. “Lung cell analysis” evaluates the proteome in a closed compartment, which can be more readily interpreted. The additional information on cell function makes it easier to attribute changes in the proteome to specific stimuli. Although the range of protein expression in cell lysates is wide (from 1–10⁶ copies per cell), it is considerably smaller than the range of concentrations in BAL fluid or serum (25). Furthermore, the protein concentration can be titrated by increasing or decreasing the number of cells, and the complexity of the sample can be reproducibly reduced by cell fractionation procedures (26, 171, 250). In 1990, Devlin and Koren (55) demonstrated changes in the proteome of alveolar macrophages isolated from BAL fluid after acute exposure of humans to 0.4 ppm ozone using 2D-PAGE. Changes in protein expression after air or ozone exposure were analyzed by 2D-PAGE and computerized densitometry. Of the nearly 900 proteins analyzed, 45 (5.1%) were expressed at a significantly increased rate after ozone exposure, whereas 78 (8.8%) were expressed at a significantly reduced rate (55). The possibilities for analysis with 2D-PAGE and mass spectrometry were demonstrated by Witzmann et al. (256) in a mouse model of jet fuel exposure. By digital comparison of gel patterns, the protein expression of the cytosolic fraction of cell lysates from exposed and unexposed mice was quantified. Identification of relevant protein spots was carried out using MALDI-mass spectrometry after tryptic digestion of the proteins. In cases where MALDI-TOF was not sufficient for protein identification, sequence tags were obtained using electrospray MS-MS. With this large-scale approach, significant differences in 44 gel spots were found, and 18 of these spots were identified. Toxic effects of jet fuel on protein synthesis and lung ultrastructure, the resulting increase in the activity of cellular detoxification systems, signs of metabolic stress, and carbonic anhydrase activity (probably as a functional response to an increase in CO₂ and acidosis) could be defined.

Westergren-Thorsson et al. (253) correlated protein expression to the physiological status of cell cultures derived from asthmatic patients and healthy volunteers. More than 1,000 proteins could be evaluated in a single experiment. They concluded that the expression of actin and tropomyosin had increased due to transforming growth factor-β (TGF-β) stimulation. These proteins were correlated to the transformation of normal fibroblasts to myofibroblasts, an important step in the remodeling processes observed in asthma (253). The same group investigated cultured fibrotic cells originating from 12 lung biopsies taken from different central pulmonary locations in three patients with asthmatic-like disorders (146). Viable cells could be isolated from 10 out of 12 biopsies. Using 1D- and 2D-PAGE with protein identification by MALDI-TOF, the authors found a proteoglycan expression pattern that was different from previous findings of the same group in normal patients and could be linked to the pathophysiology of asthma. Another recent evaluation studied the changes in the proteome of the mink lung epithelial cell line Mv1Lu in response to TGF-β1 treatment by 2D-PAGE and peptide mass fingerprinting (116). Thirty-eight proteins with altered protein synthesis could be detected by 2D-PAGE and identified by MALDI-TOF mass spectrometry. Twenty-eight of these 38 proteins had not been previously described as targets of TGF-β. Among these were proteins involved in DNA repair, the synthesis of ATP and the regulation of transcription, RNA stability, and other intracellular mechanisms. Another study (197) investigated changes in protein synthesis following the stimulation of human lung fibroblasts with endothelin-1 using pulsed [³⁵S]methionine labeling for the identification of newly synthesized proteins. Approximately 70 proteins with altered protein synthesis could be detected in 2D-PAGE, and the 35 proteins showing the largest changes were identified by MALDI-TOF mass spectrometry. Groups of functionally linked proteins were differentiated based on their kinetic behavior. The authors claim that the combination of techniques made the detection of newly identified proteins down to 10 copies per cell possible. A recent evaluation (36) compared a normal and a malignant lung epithelial cell line by peptide mass fingerprinting. An increase in the expression of aldehyde dehydrogenase, peroxiredoxin I, fatty acid-binding protein, aldotereductase, and dextrin, and a decrease in the expression of galectin-1, transgelin, and stathmin, were found (36). Because the human lung is continuously exposed to oxidative stress, the finding of the increase in the antioxidant enzyme peroxiredoxin I might be useful as a potential biomarker for lung cancer and eventually even as a possible therapeutic option. Ostrowski et al. (180) undertook a comprehensive proteomic analysis of ciliary axonemes isolated from cultured human epithelial cells; these were obtained from excess surgical tissue from transplant donors and cystic fibrosis patients. Analysis by 2D-PAGE resulted in a reproducible 2D map consisting of ≥240 individual protein spots. Digestion with trypsin and sequencing by LC-MS-MS resulted in peptide matches to 38 proteins. To identify ciliary components not resolved by 2D-PAGE, proteins were separated by 1D-PAGE and analyzed by LC-MS-MS, which resulted in peptide matches to an additional 110 proteins. In a third approach, preparations of isolated axonemes were digested with a different enzyme, the endoprotease Lys-C, and the resulting peptides were analyzed directly by LC-MS-MS or by multidimensional LC-MS-MS, leading to the identification of a further 66 proteins. In total, 214 potential axonemal proteins were identified.

A novel approach for the evaluation of changes in the lung proteome is the evaluation of “frozen tissue slices” by imaging mass spectrometry (Fig. 6 and Table 3). In a study of nonsmall cell lung cancer (265), expression profiles of several hundred cells from single frozen sections of surgically resected lung tumors were evaluated using MALDI-TOF. Twelve-micrometer sections were cut from frozen tissue samples on a cryostat and positioned on a MALDI sample plate and a glass slide. The section on the glass slide was stained with hematoxylin and
Proteomics will provide further access to the investigation of posttranslational modifications of proteins in different conditions, a considerable analytical challenge of enormous physiological relevance.

Proteomics has broadened our view of protein complexes and machines and is complementary with other investigative approaches. The typical characteristics of a proteomics approach, as explained above, imply a large potential for novel discoveries and new testable hypotheses that should be maximized by choosing an appropriate study design (3, 103, 240). The absence of constraints that limit the results of reductionist methods is, in our view, a special advantage of proteomics. The broad spectrum of the results, however, can make the interpretation and presentation of the results more challenging. Therefore, special attention should be paid to well-focused scientific questions in the study design.

The ongoing rapid evolution in separation science, mass spectrometry, and bioinformatics will continue to stimulate the investigation of the lung proteome and will lead to new insights in the near future.

ACKNOWLEDGMENTS

The authors thank Drs. Robert J. Chalkley and Rachel L. Zemans for suggestions and critical reading of the manuscript.

REFERENCES


66. Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, Fey SJ and Larsen PM. *Fiegeys D.*


