Proteomic Strategies and Their Application in Studies of Renal Function

Pedro Cutillas,1,2 Alma Burlingame,1 and Robert Unwin2

1Ludwig Institute for Cancer Research and Department of Biochemistry, University College London, W1W 7BS, and 2Centre for Nephrology and Departments of Medicine and Physiology, Royal Free and University College Medical School, University College London, NW3 2PF, United Kingdom

Proteomics is a promising new tool for functional genomics. In addition to two-dimensional gel electrophoresis, other methods that are based on liquid chromatography and mass spectrometry are now available to study proteins. In this brief article, we review the strengths and limitations of the proteomic approaches currently available to the researcher, and we provide examples of how proteomics has been, and can in the future be, used to study the kidney.

The publication of genome sequences from several organisms is an important recent advance; however, the benefits derived from this information can only be obtained when the spatial and temporal expression, function, and interactions of the gene products are known. With very few exceptions, these gene products are proteins, and although RNA and DNA microarrays can be used to detect gene expression in cells and tissues, these techniques cannot be used for the analysis of biological fluids. Furthermore, the intracellular location, post-translational modifications, and protein-protein interactions can only be analyzed at the protein level.

Thus there is considerable interest in the rapidly expanding field of proteomics, which can be defined as the analysis of the spatial and temporal expression of a subset (and ultimately a full set) of proteins in a defined biological system. Significant recent contributions have focused on use of iterative glutathione-S-transferase fusion probes (1) and tandem affinity purification strategies for discovery of protein interactions. Additional major efforts involve studies of protein machines and organelles (3). Analysis of protein-protein interactions and posttranslational modification in a “multiplex” or global manner is also a part of proteomics. The hallmarks of proteomics reside in its emphasis on taking a global and comprehensive view, involving in many cases some notion of “high throughput”; but in contrast to genomics, there is not a single biochemical method that can be used for the analysis of all proteins. This is due to the obvious fact that the chemistry of proteins involves physicochemical properties that can affect the composition of all individual macromolecules comprising a living system, whereas that of nucleic acids is uniform. Also, no equivalent for the polymerase chain reaction is available for proteins. Therefore, the choice of analytical methods used to visualize the proteome(s) under study does, in fact, influence the subset of proteins that will be identified.

Therefore, our aim is to summarize the most important of these methods that are currently available to the researcher and to highlight their strengths and weaknesses, particularly in the context of investigating kidney function. Advances in technological and mass spectrometric methods, such as electrophoresis, electrospray ionization (ESI), and matrix-assisted laser desorption/ionization (MALDI), which are the subject of intense current research and development, will not be discussed.

Targeted approaches for the detection and quantitation of proteins

Traditionally, proteins have been detected and identified by Edman degradation and quantified by immunochemical methods. This immunology-based approach is, in principle, very powerful as long as the antibodies for the proteins under study are specific and readily available. Due to the specificity of validated antibodies, virtually any protein or peptide can be detected in a background of other proteins without the need for sample pretreatment. In addition to their use in Western blots and immunoassays, antibodies can be used for the immunocytochemical and immunohistochemical localization of proteins in cells and tissues, respectively, making it possible to establish (although not always unambiguously) the intracellular localization of the protein in question. Several investigators have used immunochemical methods in a multiplex fashion for the parallel identification of several proteins or peptides. For example, the group of Knepper (14) has produced a set of antibodies for each of the solute transporters involved in sodium reabsorption from the glomerular filtrate along the renal tubule. These antibodies have been used to follow the changes in expression of each transporter as a result of changes in the physiological status of the animal model being investigated (see Ref. 14 for a comprehensive review). Norden et al. (18) also used immunochemical methods for the detection and quantitation of particular proteins in patients with the renal Fanconi syndrome (FS) and in normal human urine, which gave important insights into the nature of low-molecular-weight (so-called tubular) proteinuria. Although this strategy is both powerful and useful to follow the changes in expression levels of specific proteins, it is not suitable for the discovery and identification of new candidates involved in physiological processes, because there is a limit to the number of antibodies that can be used at any one time (and the ultimate goal of comprehensive protein arrays has yet to be achieved). Furthermore, immunochemical methods are biased in that it is the investigator who chooses in advance the set of
antibodies to be used. Problems of specificity, cost, and the availability of antibodies must also be considered, and antibodies cannot be used to follow the expression of genes whose products have not yet been characterized.

Analysis of proteomes by two-dimensional gel electrophoresis and mass spectrometry

Two-dimensional gel electrophoresis (2DE) separates proteins according to their isoelectric point on the first dimension and to molecular weight on the second. This separation technique was first introduced for proteins in 1975 (13, 19) and has been extensively used to compare the proteomes of related samples (e.g., treated vs. control). The reproducibility of protein separation by 2DE improved with the development and introduction of immobilized pH gradient strips (10). An important issue is the choice of detection method employed to detect gel-separated proteins. Ideal protein stains should be sensitive to permit the detection of low abundant proteins, but they should also have a broad linear dynamic range of quantitation to provide accurate estimations of abundance (21). Traditionally, proteins have been detected by Coomassie Brilliant Blue (which has a linear response of intensity as a function of protein abundance but low sensitivity) or with silver nitrate-based staining methods (which offer high sensitivity but low dynamic range). In addition, novel methods based on fluorescent dyes (e.g., Sypro Ruby staining) are now available that are as sensitive as silver staining and show a greater dynamic range. Detection of proteins can also be accomplished by labeling them with fluorescent compounds before their separation by 2DE (see below).

In a typical 2DE experiment, the protein samples to be compared are separated in parallel gels. After gels are stained and scanned, gel spots that show an altered level of expression are excised and “in-gel” digested with a suitable protease (usually trypsin). In-gel digestion protocols exist that are compatible with downstream analysis by mass spectrometry (MS) (17, 22). The peptides produced in this way can then be analyzed by mass spectrometry (MS) or mass spectrometry-mass spectrometry (MS-MS) for identification.
MS, either MALDI time-of-flight (TOF) or ESI tandem MS (MS/MS). The latter mode of MS is often used “online” with liquid chromatography (LC) operating at low flow rates (nanoliter-per-minute flow rates are often used; hence the term nanoflow LC or nanoflow). Technical advances in biological MS and bioinformatics during the 1990s meant that mass spectrometers now have enough sensitivity for the identification of proteins showing only a faint signal in silver-stained gel spots (see Refs. 4 and 16 for more detailed information on biological MS). Thus the limit on sensitivity of the proteomics “workflow” is no longer MS.

An approach widely used for the identification of proteins with in-gel digestion and MS data is that of peptide mass fingerprinting (PMF), in which the masses [or, more properly, the mass-to-charge ratios (m/z)] of the peptide ions observed in the spectrum are used to search a protein database. Search algorithms have been developed that use those m/z values and compare them with those derived from the theoretical digestion of all proteins in the database (e.g., see Ref. 6). A hit is defined as when the theoretical and the observed m/z values coincide (within a defined mass error). Algorithms that assess the statistical probability that the hit is correct have also been developed (6), although operator intervention is needed to identify false positives. An example of the proteomics workflow based on 2DE, in-gel digestion, MALDI-TOF MS, PMF, and database searching is shown in Fig. 1 and Table 1.

Although the methodology for protein identification described above has been extensively used in the past to identify gel-separated proteins, protein identification by this approach is never completely unambiguous, because a set of peptides often matches more than one protein entry, a problem that is being accentuated as a result of the inclusion of translated genome sequences into protein databases. Therefore, alternatives to MALDI-TOF MS and PMF for the identification of gel-separated proteins have been developed. One of these approaches is LC-ESI-MS/MS, followed by sequence-tag searches, in which the peptides generated by in-gel digestion of the protein spot are separated by LC and detected and sequenced by MS/MS. Sequencing by MS is carried out using tandem mass spectrometers, which are instruments capable of isolating and fragmenting ions. With this instrument, the isolated ion (i.e., the peptide precursor or “parent ion”) is collided with neutral gas molecules that increase the internal energy of the peptide, which is then fragmented in a process termed collision-induced dissociation (CID). The “daughter” or fragment ions produced by CID are analyzed on a second mass analyzer in the mass spectrometer, which produces a MS/MS spectrum that is used for primary structure (i.e., sequence) determination. Figure 2 shows an example of peptide sequence determination using a CID spectrum obtained by LC-ESI-MS/MS.

Proteomics studies using the approach of two-dimensional SDS PAGE in combination with MS have shown the usefulness of these methods in comparing gene expression profiles of related proteomes. For example, 2DE has been used to identify proteins differentially expressed in kidney medulla and cortex (2) and to analyze comparatively rat urinary proteomes before and after sodium loading (23). However, it has been shown that 2DE-based methods only detect abundant and water-soluble proteins (11), and therefore this strategy cannot

<table>
<thead>
<tr>
<th>TABLE 1. Protein identification by peptide mass fingerprinting and database searching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Submitted m/z</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>879.4441</td>
</tr>
<tr>
<td>879.4441</td>
</tr>
<tr>
<td>1161.5222</td>
</tr>
<tr>
<td>1165.5688</td>
</tr>
<tr>
<td>1177.5195</td>
</tr>
<tr>
<td>1198.6143</td>
</tr>
<tr>
<td>1285.6205</td>
</tr>
<tr>
<td>1302.6229</td>
</tr>
<tr>
<td>1303.594</td>
</tr>
<tr>
<td>2064.9266</td>
</tr>
<tr>
<td>2612.1564</td>
</tr>
<tr>
<td>2628.1766</td>
</tr>
<tr>
<td>2693.1139</td>
</tr>
<tr>
<td>3156.6600</td>
</tr>
</tbody>
</table>

The mass/charge ratio (m/z) values obtained from the spectrum shown in Fig. 1 were used to search the NCBI protein database with the search engine MS-Fit, which is part of Protein Prospector set of software (these programs can be accessed at http://prospector.ucsf.edu/ and http://128.40.158.151/mshome3.4.htm). Thirteen of the ions observed in the spectrum matched to theoretical tryptic peptides from retinol binding protein within a mass error of 50 ppm [13/41 matches (31%), 22868.1 Da, p1 = 5.48, GenBank accession no. X00129, Homo sapiens precursor RBPI]. The matched peptides covered 76% of the protein. MH+, theoretical mass of protonated peptide; ppm, parts per million; PTM, posttranslational modification.
resolve complex proteomes or hydrophobic proteins, such as membrane proteins, which in most cases fail to enter the gel used for isoelectric focusing (the first dimension). Furthermore, identification by 2DE of small or large proteins and peptides, and those with extremes in pI, remains a problem. An additional caveat is that gel spots may contain more than one protein, thus making quantitation unreliable. Finally, reproducibility and dynamic range of quantitation are still important issues, although significant advances are being made with a method termed difference gel electrophoresis (DIGE) (24). In this method, the proteomes to be compared are labeled with fluorescent dyes that emit at different wavelengths and have different colors, such as red and blue. The labeled samples are pooled, and the proteins are separated in a single gel. After scanning, the relative intensities of each color in each protein spot are used for relative quantitation. Due to the cost of the reagents and scanning apparatus, not many groups have implemented DIGE technology, which also suffers from many of the problems associated with the 2DE approach.

Analysis of proteomes by LC-based methods

As an alternative to 2DE-based methods, in which proteins are separated and then analyzed one at a time by MS, strategies based on LC-MS/MS are being used for the “shotgun” (reminiscent of the approach used by some workers in genome analysis) identification of all of the proteins in a sample (see Ref. 15 for a recent review). In this method, proteins are first digested with a suitable protease and then the peptides generated (there could be several hundred thousand if a whole cell lysate is to be analyzed) are separated by 2D-LC, the first dimension being strong cation exchange (SCX) and the second reversed-phase LC. Peptides are sequenced essentially as described for the identification of gel-separated proteins in Fig. 2, with the exception that longer LC runs are used to increase the peak capacity of the system. A 90-min reversed-phase LC run can, in combination with ESI-MS/MS detection, sequence ~900 peptides/run. Therefore, a 2D-LC-MS/MS experiment in which 20 SCX fractions are analyzed could generate 18,000 MS/MS spectra (peptide sequences, although not all CID spectra can define a sequence). The drawback of this method, when compared with 2DE-based approaches, is that quantitative estimates are not easy to make and labeling with stable isotopes is required for proper relative quantitation.

A method for labeling peptides before 2D-LC-MS/MS is isotope-coded affinity tags (ICAT), first introduced by Aebersold and colleagues (12) and later commercialized by Applied Biosystems. The ICAT reagents are alkylation compounds that label cysteine amino acid residues. One of the samples to be analyzed is labeled with an ICAT reagent containing nine $^{13}$C, whereas the other is labeled with the same compound but containing the common $^{12}$C. After being labeled, the samples are mixed and analyzed by 2D-LC-MS/MS as outlined earlier and illustrated in Fig. 3. A problem associated with the ICAT strategy is that these reagents label cysteine residues, and proteins that do not contain this amino acid in their sequence cannot be detected or quantified using this method. Recently, other investigators have reported the development of compounds based on the principle of ICAT. Alternative strategies have also been described, which involve metabolic labeling by growing cell cultures in media containing amino acids that incorporate heavy stable isotopes (20) or in media in which $^{15}$N replaces $^{14}$N as a nitrogen source (25).
Direct comparison of proteomic approaches to the study of kidney function

We have assessed the relative merits of four different proteomic approaches to study the qualitative differences in the proteome of Dent’s disease patients (7, 8). This is an inherited form of FS characterized by a dysfunction of the proximal renal tubule and manifests as a generalized loss of several important solutes in urine, especially small peptides and low-molecular-weight proteins. Patients often develop kidney stones and eventually renal failure. The genetic cause of some examples of FS is now known: mutations in a chloride channel, ClC-5, have been linked to the form known as Dent’s disease. This channel is believed to be intracellular and highly expressed within the renal proximal tubule, where it is critically involved in the megalin-cubilin receptor-mediated endocytic pathway that reabsorbs filtered polypeptides from the glomerular filtrate. ClC-5 is believed to regulate acidification of the intracellular endosomes containing the megalin-cubilin-polypeptide complex and to facilitate its dissociation and thus recycling (and reuse) of the receptor to the brush border of the proximal tubule (5). Since the reuptake of proteins and peptides from the tubular fluid is disturbed in patients with FS, it is believed that the protein composition of their urine is close to that of the glomerular filtrate.

Analysis of Dent’s disease urine by 2DE demonstrated that several vitamin carrier proteins comprise a larger proportion of the Dent’s urinary proteome when compared with that of normal subjects (an example of such a finding is shown in Fig. 1). These results were later confirmed by experiments using LC of ICAT-labeled, as well as unlabeled, peptides, as illustrated in Fig. 3. Furthermore, other peptides and proteins that were not detected by 2DE were found to be present in Dent’s urine at higher levels than in normal urine when the analysis was standardized to protein concentration. As an example, several bioactive peptides were present at different levels with respect to total protein when Dent’s and normal urine were compared. In this respect, Fig. 4 shows that the EGF precursor occupied a larger proportion of the normal urinary proteome, whereas IGF-II was present at higher levels in Dent’s urine compared with normal urine, in respect of total protein. These findings were consistent within sample groups and demonstrated the usefulness of 2D-LC-MS/MS of ICAT-labeled peptides for the identification and quantitation of polypeptides. However, although LC-MS/MS of unlabeled polypeptides detected the presence of several lipoproteins at high levels in the urine of Dent’s patients, the fold difference in these proteins between the Dent’s disease patients and normal subjects could not be determined using the ICAT method, because most lipoproteins
do not have cysteine residues in their coding sequence.

Concluding remarks

Several approaches now exist to characterize and compare complex proteomes. These methods are complementary in nature, because proteins are heterogeneous entities from a physicochemical standpoint. The functions of proteins range from structural to catalytic, and their properties differ as a result. We predict that 2DE will continue to be a useful method for the separation and visualization of proteins. An important advantage of 2DE is that it can separate protein isoforms originating from posttranslational modifications, which change the pl of the protein (e.g., phosphorylation). On the other hand, LC or one-dimensional SDS-PAGE-based methods, followed by LC-MS/MS, are powerful for identifying proteins in a high-throughput and unbiased manner and may give better results than 2DE approaches when the aim is to study small and hydrophobic proteins. When relative quantitation is required, stable isotope techniques can be used in conjunction with 2D-LC-MS/MS. Thus the choice of method to be used will depend on the particular application; when several approaches are used in parallel, they should provide a means to detect or exclude methodological artifacts as the source of any observed differences.

Our studies on the proteome of Dent’s disease urine indicated that there are qualitative differences between the urinary proteome of this form of FS and that of normal subjects, in addition to the quantitative differences reported previously (9). This may reflect specificity of the reabsorption of polyopeptides from the glomerular filtrate, such that vitamin-binding proteins are reabsorbed more readily than other proteins (e.g., albumin) from the glomerular filtrate. Our studies also indicate that the composition of bioactive peptides in Dent’s and normal urine differs significantly and that these abnormalities are also likely to be present in the renal tubular fluid. Since receptors for some of these peptides have been located on the luminal side of renal tubular cells, alterations in the bioactivity of tubular fluid in Dent’s disease may contribute to tubular cell dysfunction and ultimately to progressive renal disease. This is one example of how proteomic methods, especially when combined, can provide novel and complementary information, which can then be used to investigate normal and abnormal renal function.

We are grateful to Prof. M. Waterfield for his help and encouragement. P. Cutillas and R. Unwin thank the National Institutes of Health for National Center for Research Resources Grant RR-01614.

References