Advances in Urinary Proteome Analysis and Biomarker Discovery

Danilo Fliser,* Jan Novak,† Visith Thongboonkerd,‡ Angel Argilés,§ Vera Jankowski,‖ Mark A. Girolami,¶ Joachim Jankowski,** and Harald Mischak***

*Medical School Hannover, Hannover, Germany; †University of Alabama at Birmingham, Birmingham, Alabama; ‡Faculty of Medicine at Siriraj Hospital, Mahidol University, Bangkok, Thailand; §Department of Nephrology, University Hospital, Néphrologie et Dialyse St Guilhem and SAS RD, Néphrologie, Montpellier, France; ‖Nephrology Department, Charité Hospital, Berlin, Germany; ¶Department of Computing Science, University of Glasgow, Glasgow, Scotland; and **Mosaiques-Diagnostics and Therapeutics AG, Hannover, Germany

Noninvasive diagnosis of kidney diseases and assessment of the prognosis are still challenges in clinical nephrology. Definition of biomarkers on the basis of proteome analysis, especially of the urine, has advanced recently and may provide new tools to solve those challenges. This article highlights the most promising technological approaches toward deciphering the human proteome and applications of the knowledge in clinical nephrology, with emphasis on the urinary proteome. The data in the current literature indicate that although a thorough investigation of the entire urinary proteome is still a distant goal, clinical applications are already available. Progress in the analysis of human proteome in health and disease will depend more on the standardization of data and availability of suitable bioinformatics and software solutions than on new technological advances. It is predicted that proteomics will play an important role in clinical nephrology in the very near future and that this progress will require interactive dialogue and collaboration between clinicians and analytical specialists.

What Do Proteome and Proteomics Entail?

Proteins and peptides within a particular body compartment comprise a proteome. In contrast to the genome, which is unique and relatively stable (most cells of an organism possess the total genomic capital during life), proteomes are cell and tissue specific and change over time in response to different situations. Proteomics (i.e., the assessment of these proteomes) may contribute to the elucidation of the proteome of a given cell type (e.g., epithelial, mesangial, endothelial), tissue (e.g., renal cortex), or even specific parts of the tissue (e.g., glomerulus). Furthermore, proteomics can reveal proteome changes in biologic fluids (e.g., urinary or plasma proteome). The term “peptidome” has been used for a “subset” of the proteome, the lower molecular weight peptides in a sample. Because of these somewhat ill-defined differences and because the invention of additional terms does not improve comprehension of an already complicated matter, in this review we use the term “protein” for all naturally occurring polypeptides (poly-amino acids).
Two basic sources of material are available for proteomic studies: Body fluids (e.g., urine, blood) and tissue. Proteome research related to nephrology has generally focused on the examination of urine because it is easily accessible in a large quantity without the use of invasive procedures. Furthermore, as a rule, pathophysiologic changes in the genitourinary tract and the kidney are reflected by changes in the urinary proteome. Although many studies have shown that proteins in biologic fluids may degrade rapidly when handled inappropriately, urinary proteins have been shown to remain stable long enough to perform reliable proteome analysis. In two independent sets of experiments, Schaub et al. (7) and Theodorescu et al. (8) showed that the urinary proteome did not undergo significant changes when urine was stored for 3 d at 4°C or 6 h at room temperature, respectively. In addition, urine can be stored for several years, even at −20°C, without significant alterations in its proteome. Although these reports suggest a much greater stability of the urinary proteome compared with the blood proteome, it must be clearly noted that other issues will influence data quality and comparability. Given the complexity of the urinary proteome that we can currently only estimate, there are likely to be important changes associated with differences among samples as a result of variations in procedures for collection, storage, and, of course, processing. As outlined in more detail elsewhere (3), these issues must be taken into account, and standardized protocols for urine sampling and for handling of the samples should be adopted.

In contrast to urine, collection of blood is invasive and requires meticulous preanalytical handling. Its proteomic analysis is prone to analytical artifacts. A detailed comparison of serum and plasma proteomes revealed that an array of proteases are activated immediately upon clotting, resulting in the generation of many degradation products (9). As a consequence, the human proteome consortium has recommended that blood be examined as plasma rather than as serum and established a standardized sample collection protocol (10). Similar and absolutely essential efforts for standardization of collection and processing are under way for urine and would certainly be desirable for various types of tissue.

**Technical Aspects of Proteomics**

Because of the inherent complexity of a proteome, all approaches for its examination generally rely on a separation step, followed by ionization and subsequent mass spectrometry (MS) analysis. This complexity is further increased by posttranslational modifications (PTM), which may well serve as biomarkers for disease (e.g., advanced glycation in diabetes [11]), hence must be considered. Because of their ability to analyze mass, in general all MS-based proteomics technologies will identify PTM, because these result in a change in mass. Furthermore, PTM frequently result in changes in the migration in any of the pre-MS separation approaches described in more detail herein. It is beyond the scope of this review to outline the differences in the ionization processes and the modern mass spectrometers; these topics have been summarized elsewhere (9). In general, quadrupole (Q), ion-trap, time-of-flight (TOF), and Fourier transform-ion cyclotron resonance (FT-ICR) instruments or their combinations (e.g., hybrid instruments such as Q-TOF, combining quadrupole and time-of-flight detectors) are currently used for proteome analysis. To obtain sequence information (as well as information on PTM), sequential use of these techniques, termed tandem mass spectrometry (MS/MS), is used. In general, the first MS instrument serves as a mass filter, “collecting” only the ions with the mass of interest (“parent ions”), and the second MS instrument is used to analyze fragmentation products (“daughter ions”) that may be generated by collision with other molecules (collision-induced dissociation) or transfer of electrons (electron transfer dissociation). Frequently, individual advantages of the different mass detectors are combined (e.g., the precision of quadrupoles and the high accuracy of the measurements of mass with TOF in a Q-TOF instrument). This approach cannot be easily applied to sequencing large proteins (top-down approach); instead, these proteins are usually digested by proteases, such as trypsin. The resulting smaller fragments are then sequenced with good success. Recent technical advances enabled sequencing of large native peptides and proteins up to 66 kD (12–14). In general, every pre-MS separation technology (with the exception of surface-enhanced laser desorption/ionization [SELDI]) can be combined with any MS approach. The highest resolution and the best mass accuracy (deviation of the measured mass from the theoretical mass) can be obtained by using FT-ICR instruments (<1 ppm). Unfortunately, the high cost of these instruments has limited their widespread use. Alternatively, any other mass spectrometer that is capable of delivering high-quality data, with mass deviation <50 ppm and resolution >5000 would be acceptable for proteomic applications. However, these values should probably be regarded as the lowest standards for accurate proteome analysis.

All MS instruments analyze a mass/charge ratio (m/z) and
consequently require ions for analysis. Ionization of the compound of interest can be achieved using matrix-assisted laser desorption ionization (MALDI) or electro-spray ionization (ESI). For MALDI, the sample is mixed with a matrix, spotted onto a target, and ionization is achieved with a pulsed laser. Matrix absorbs the energy of the laser, transfers it to the analyte, and thereby assists in its ionization. This method is generally applied off-line (separation is not coupled to a mass spectrometer). It is technically less demanding, but it is susceptible to “signal suppression,” a phenomenon whereby certain analytes are preferentially ionized and more easily detected at the expense of other compounds that may even become undetectable (15,16). ESI generates charged droplets in a high-voltage field; solvent from these charged droplets evaporates, giving rise to multiply charged ions of the analyte. This approach is generally used on-line, is less stable, but also is less susceptible to signal suppression.

Proteomics that aim toward biomarker discovery can be considered a sophisticated comparative analysis. Consequently, quantification of proteins is an essential consideration. For gels, protein stains are generally used for quantification. The newer fluorescence stains seem to provide superior results, because they have large linear dynamic ranges and similar or better sensitivity compared with the older Coomassie blue and silver stains (17). There are also good MS methods for relative quantification on the basis of ion counting (18), but care in instrument and technical consistency is essential. Absolute quantification usually requires previous identification of the biomarker sequence and/or chemical derivation that consequently may become restrictive (19). Relative quantification of biomarker abundances with reference to constant peaks generally seems sufficient, especially when considering biologic variation.

Four different approaches that frequently are used for pre-MS fractionation are briefly reviewed because they provide very different types and quality of data. For a comparative evaluation of the potential of these MS-based approaches for clinical applications, the most prominent advantages and disadvantages are listed in Table 1.

Table 1. Advantages and disadvantages of various mass spectrometry–based proteomic techniques for clinical applications

<table>
<thead>
<tr>
<th>Technology</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2DE-MS</td>
<td>Applicable to large molecules</td>
<td>Not applicable to peptides &lt;10 kD, no automation, time-consuming, quantification difficult, expensive</td>
</tr>
<tr>
<td>LC-MS</td>
<td>Automation, multidimensional, high sensitivity</td>
<td>Time-consuming, sensitive toward interfering compounds, restricted mass range</td>
</tr>
<tr>
<td>SELDI-MS</td>
<td>Easy-to-use system, high throughput, automation, low sample volume</td>
<td>Restricted to selected proteins, low-resolution MS, low information content, lack of comparability</td>
</tr>
<tr>
<td>CE-MS</td>
<td>Automation, high sensitivity, fast, low sample volumes, multidimensional, low cost</td>
<td>Not suited for larger proteins (&gt;20 kD)</td>
</tr>
</tbody>
</table>

*a2DE-MS, two-dimensional gel electrophoresis followed by mass spectrometry; LC-MS, liquid chromatography coupled to mass spectrometry; SELDI-MS, surface-enhanced laser desorption/ionization coupled to mass spectrometry; CE-MS, capillary electrophoresis coupled to mass spectrometry.*
duction of an internal standard labeled with a third dye (Cy2), thereby allowing quantitative analysis. Although the comparison of two samples with 2D-DIGE has been satisfactory (24), the comparison of several different experiments remains chal-

Figure 2. Two-dimensional gel electrophoresis coupled to mass spectrometry (2DE-MS). (A) Proteins in samples from different individuals are separated in two dimensions according to the isoelectric point and molecular weight. The resulting 2DE gels are stained and compared. (B) Protein spots that seem to differ between the two gels are excised and digested with trypsin, and the resulting peptides are analyzed by mass spectrometry. (C) Derivation of the samples before analysis using different fluorescence dyes allows analysis of two different samples in the same gel.
but, in part because of its demand in skills and time, has not been adapted for clinical applications.

**Liquid Chromatography Coupled to Mass Spectrometry**

Liquid chromatography (LC) provides a powerful fractionation method that is compatible with virtually any mass spectrometer. This method separates large amounts of analytes on an LC column (15,25,26) and offers high sensitivity. A sequential separation using different media in two independent steps provides a multidimensional fractionation that can generate vast amounts of information. For example, the multidimensional protein identification technology (MudPIT) (27,28) and a 2D liquid-phase fractionation approach (29) are well suited for an in-depth analysis of body fluids and tissues (Figure 3). Limitations of LC-MS include difficulties with comparative analysis, in part because of the variability in multidimensional separations and the substantial time required for the analysis of a single sample (generally, in days). Larger protein (above approximately 10 kD) cannot generally be analyzed by this technique; instead, they have to be cleaved by a protease (e.g., trypsin), and the resulting fragments can be subsequently analyzed. Furthermore, the method suffers from its sensitivity to interfering compounds (e.g., lipids, detergents). When data from tryptic digests that were analyzed by LC-MS were compared with 2DE-MS analysis, mostly different proteins were detected by each technique (30). Therefore, both techniques either individually or in combination have the potential to identify valid biomarkers that may not be detected with any other technology that is available. However, neither seems to be suitable for the comparative analysis of hundreds of samples or application in clinical diagnostic processes. Therefore, it cannot be overemphasized that it is imperative to have a strategy for subsequent validation and application of an assay in the clinical laboratory based on another technology (see Mischak et al. [3]).

Two alternative MS-based approaches have been used in “clinical proteomics.” These techniques have been found applicable not only to the discovery phase but also to the validation and subsequent application phases.

**SELDI-MS**

SELDI technology (31–34) reduces the complexity of a sample by selective adsorption of proteins to different active surfaces. Proteins bind to a specific surface (hydrophilic matrix, reverse-phase material, or affinity reagents, e.g., lectins, antibodies) with varying degrees of selectivity while the unbound sample is washed away. A matrix that absorbs energy and allows vaporization and ionization of the sample by laser is added. The sample is subsequently analyzed by MS that provides a low-resolution mass “fingerprint” (Figure 4). Advantages of SELDI include its capacity to analyze multiple samples in a short time. SELDI has been used in numerous attempts to define biomarkers for a variety of diseases (32,35,36). Although the technology is easy to use, it is also prone to generating artifacts (37,38). This may be due, in part, to difficulties with calibration and lack of precision of the determined molecular masses of the analytes. Furthermore, only a very small fraction of all proteins in a sample binds to the chip surface, and the binding varies depending on sample concentration, pH, salt content, presence of interfering compounds such as lipids, etc. Therefore, because most of the information contained in a biologic sample is eliminated during the sample preparation, comparability of data sets is limited.

**Capillary Electrophoresis Coupled to MS**

This approach is based on capillary electrophoresis (CE) at the front end coupled to a mass spectrometer (Figure 5). CE separates proteins in a single step with high resolution based on their migration through a gel in the electrical field (300 to
CE-MS offers several advantages: (1) It provides fast separation and high resolution (39); (2) it is robust and uses inexpensive capillaries instead of expensive LC columns (9); (3) it is compatible with most buffers and analytes (40); and (4) it provides a stable constant flow, thus avoiding elution gradients that may otherwise interfere with MS detection (41). As shown in Figure 5B, CE-MS enables the generation of comparable high-resolution data sets. The data sets from individual analyses can be compiled to generate a typical proteome pattern that can be based on >100 individual analyses. A disadvantage of CE (although to a lesser degree than LC) is that it cannot be easily used for the analysis of high molecular weight proteins because the large proteins tend to precipitate at the low pH that generally is used in the running buffer. However, such proteins can be digested and the resulting fragments can be analyzed (as described for LC-MS). Furthermore, they can be effectively removed by ultrafiltration (42). In addition, certain proteomes, such as the urinary proteome of healthy individuals, contain mostly low molecular weight proteins; in such cases, the restricted ability to analyze large native proteins does not constitute a severe drawback. Another limitation of CE-MS is the relatively small sample volume that can be loaded onto the capillary, leading to a potentially lower sensitivity of detection. This obstacle has been resolved by improved methods of ionization and by better delivery of the separated protein from the end of a capillary to the MS instrument in a small stream of liquid by nano-ion spray. Also, improvements in the detection limits of mass spectrometers render the issue of sensitivity less important (43–45). Sequencing of potential protein biomarkers that are defined by CE-MS analysis can be achieved by directly interfacing CE with MS/MS instruments (46) or by subsequent targeted sequencing using LC-MS/MS (12,15). Consequently, CE-MS has become a viable alternative to the commonly used proteomic technologies and recently was successfully applied in several clinical studies (8,47,48).

Protein Arrays

As a non–MS-based approach, protein arrays can be used to detect specific proteins (“targeted proteomics”). For more detailed information, we refer the reader to a recent review by Kozarova et al. (49). In general, this technique can be considered the modern version of immune detection of multiple proteins. Specific antibodies or antigens are printed on a surface (generally a slide or membrane). A single sample is hybridized to the array that may contain several hundred targets; the captured antigens or antibodies are then detected. As an example, such arrays have been successfully used for detection of antibodies that are directed against specific glomerular antigens (50). The targeted proteome analysis has uncovered novel associations of autoantibodies with disease. Investigators have been able to differentiate between patients with lupus nephritis and healthy and disease controls and to categorize further the severity of the lupus nephritis on the basis of differential IgG and IgM autoreactivity. Similar to other approaches, protein arrays show several limitations (49): The need for a specific probe for every protein to be analyzed (in contrast to nucleic acid arrays that can use the antisense sequence), the generally low density that allows detection of only a few proteins, and posttranslational modifications are usually not detected.

Applications of Proteomics in Nephrology

The main focus of proteome analysis in nephrology is on detection and identification of (urinary) proteins that significantly change (in abundance, distribution, etc.) during (patho)physiologic changes of the kidney structure and/or function.
To allow specific and early assessment of disease, at least some of these proteins should be biomarkers that are independent from proteinuria. These biomarkers may be directly related to the disease (e.g., IgA immune deposits in IgA nephropathy) or may result from secondary events (e.g., generation of specific cleavage products by metalloproteases that are upregulated during inflammatory processes in the kidney). After validation, some of these changes may potentially be new therapeutic targets or novel biomarkers for disease detection and/or prognosis.

Although this review is focused on the urine, kidney tissue certainly contains relevant proteomic information. However, analysis of its proteome encompasses several disadvantages: The kidney is composed of different cell types (all of which express different and specialized proteomes) and tissue samples must be obtained invasively, rendering proteome analysis especially of the normal human kidney (which is required as the control) ethically difficult or even impossible. Therefore, most research has focused on tissue that is obtained from experimental animals. The proteome of the rat kidney was described recently by Arthur et al. (51). The authors showed differential expression of proteins in the renal cortex and medulla. 2DE resolved 1095 spots from the cortex and 885 spots from the medulla. By MALDI-TOF MS, 54 unique proteins were identified. Nine of them were differentially expressed in the cortex and medulla, and four were expressed in only one region. Xu et al. (52) examined glomeruli that were obtained by laser capture dissection and subsequently analyzed the proteome of tissue in the five-sixths nephrectomy rat model of FSGS. They identified thymosin β4 as a marker of glomerulosclerosis. Recently, the first report of such “protein maps” from murine tissue has also been published (53).

Urine Proteomics

As already outlined, the urine seems to be an ideal source of potential biomarkers. Urinary proteins can be analyzed directly or separated by centrifugation into distinct fractions. For example, supernatants from low-speed centrifugation contain proteins that are derived from filtered plasma proteins and secreted by tubular epithelial cells. This supernatant can be further centrifuged at high speed (ultracentrifugation), yielding pellet-containing exosomes, small vesicles (with diameter <80 nm) with cell membrane, and cytosolic proteins. These exosomes are derived from epithelial cells that line the urinary tract with a contribution from filtered exosomes from blood cells (54,55).

Before the “proteomics” era, many investigators have sought to define better the urinary proteome in a variety of clinical situations. In this respect, one of the first attempts to define proteins in the urine was published by Spahr and co-workers (56,57). Using LC-MS, they analyzed pooled urine samples after trypic digestion and identified 124 proteins. Although this study did not attempt to define any urinary biomarkers for a disease, it clearly highlighted the plethora of information in the urinary proteome and also a possible approach toward its mining. This conclusion was underscored by Fang et al. (58), who used not only 2DE but also 1D- and 2D-LC to define potential biomarkers for inflammation. Using acetone-precipitated urine samples from healthy volunteers, Thongboonkerd et al. (59) defined the first human urinary proteome map, consisting of 67 proteins and their isoforms, that could be used as a reference. In a subsequent study by Oh et al. (60), pooled urine samples from 20 healthy volunteers were used to annotate 113 proteins on a 2DE by peptide mass fingerprinting. Additional experiments that further expanded the knowledge of the normal urinary proteome were reported by Pieper et al. (61), Sun et al. (62), and Castagna et al. (63). Taken together, these approaches have identified approximately 800 proteins and laid the foundation for the subsequent discovery of biomarkers in the urinary proteome. In a recent study on urine that was obtained from healthy individuals, Adachi et al. (64) identified more than 1500 proteins (or fragments) in the urine of healthy individuals, further underlining the complexity of the human urine proteome. A large proportion of the proteins that were identified in this study were represented by membrane proteins. This may be due to the presence of exosomes (55). Recently, exosomal fetuin-A was proposed as biomarker of acute kidney injury, based on data from a rat model (65), which were further supported by Western blots on three patients. Although these data and the concept of exosomes are very promising, these preliminary observations need to be verified and further explored.

Identification of “Biomarkers” for Kidney Diseases

The definition of disease-specific biomarkers in the urine is complicated by significant changes in the urinary proteome during the day, most likely as a result of exercise, variations in the diet, circadian rhythms, etc. (66). As a consequence, the reproducibility of the assay is reduced as a result of these physiologic changes, even if the analytical method shows high reproducibility. In addition, clear differences between first-void and midstream samples can be noted (Mischak et al., unpublished data), further highlighting the importance of standardized protocols for urine sample collection. In one of the first reports on specific urinary proteomic biomarkers in 1979, the identification of several urinary proteins was reported by Anderson et al. (67). 2DE analyses of urine from living kidney donors identified several potential proteins related to compensatory kidney growth (68). This work was followed by studies that showed a significant effect of retinoids on renal cells (69,70). Furthermore, 2DE of urine from patients with various biopsy-proven (primary) kidney diseases displayed distinct differences. For example, α1-antitrypsin was increased in a group of patients with FSGS (71); these results are in line with more recent findings that were obtained with modern proteomic technologies.

On the basis of three patients with diabetic nephropathy, Sharma et al. (72) used 2D-DIGE and identified urinary proteins that were differentially present in disease. α1-Antitrypsin was identified as a potentially upregulated biomarker; this finding was confirmed by Western blotting. More recently, Park et al. (73) examined pooled urine from 13 patients with IgA nephropathy and compared the data with those from 12 normal control
subjects. The authors found an array of differentially present proteins and used the data to initiate the establishment of a human urinary proteome map of IgA nephropathy. This study also outlined the limitations of 2DE: The method is tedious and time-consuming; therefore, only a few samples and controls can be analyzed with reasonable effort. However, it is evident that such approaches will enable the definition of several potential biomarkers, and it will be essential in the future to develop a strategy for their evaluation.

A procedure that increases throughput by reducing complexity is SELDI technology. As reviewed already, this technique was recently used by several researchers (e.g., Clarke et al. [32] and Schaub et al. [74]) to detect potential biomarkers for allograft rejection in kidney transplant patients. Clusters of five and three urinary proteins, respectively, were sufficient for correct classification of 34 and 50 patients with high sensitivity and specificity. It is interesting that these two groups defined completely different biomarkers for the same disorder, and neither found differences between patients with transplanted and native kidneys, certainly an unexpected observation. Urine from patients with acute rejection of a renal allograft has also been obtained from control subjects or patients with different renal diseases permits the definition of an array of biomarkers that differentiate healthy subjects from patients and additional markers that define the specific (kidney) disease or clinical condition. The latter type of biomarkers is useful for differential diagnosis.

It is of interest to note that many of the identified biomarkers in the urine of patients with renal diseases are proteolytic fragments of larger proteins. Apparently, specific proteases in the urine may cleave these excreted proteins, as suggested by a recent study of patients with nephrotic syndrome as a result of several glomerular diseases. The authors demonstrated specific urinary proteases that cleaved albumin and α1-antitrypsin and generated almost 100 proteolytic fragments; these polypeptides appeared as distinct spots on 2D gels (83).

In a recent study, Decramer et al. (47) used CE-MS–based urinary proteome analysis to define specific biomarker patterns for different grades of ureteropelvic junction obstruction, a frequently encountered pathology in newborns (of note, these patients do not have significant proteinuria). In a blinded prospective study, the biomarker patterns predicted with 94% accuracy the clinical outcome of these newborns 9 mo in ad-

Figure 6. Protein patterns of healthy individuals (NK) and patients with IgA nephropathy (IgA-N) or vasculitis (Vasc). (Top) Patterns that consist of 20 to 100 single measurements, molecular mass (0.8 to 20 kD, on a logarithmic scale) against normalized migration time (18 to 45 min), peak height, and color encode the signal intensity. (Bottom) Zoom of the top patterns (1.5 to 5 kD, 19 to 30 min). As evident, an array of general biomarkers for kidney disease that are present both in IgA-N and vasculitis can be defined. In addition, biomarkers that are specific only for IgA-N or vasculitis can be detected.
vance (Figure 7). These results indicated the potential of urinary proteomics not only to diagnose the renal disease accurately but also to predict its prognosis correctly.

Urinary proteome analysis may also be an excellent tool for fast, noninvasive, and unbiased monitoring of disease progression or response to therapy. In a randomized, double-blinded study, Rossing et al. (84) evaluated the treatment of macroalbuminuric patients with daily doses of 8, 16, and 32 mg of candesartan or placebo for 2 mo. Candesartan treatment resulted in a significant change in 15 of 113 proteins that are characteristic for diabetic renal damage. Similar data have been obtained in patients with vasculitis (Haubitz et al., manuscript in preparation) (85), where therapy improved/changed the vasculitis-specific protein pattern toward a normal urinary proteome.

Most current proteomic approaches were used to define new biomarkers for disease. The (patho)physiologic relevance of these biomarkers, although initially unknown, can be clarified by sophisticated bioinformatic approaches. Recently, an alternative approach was used for the molecular phenotyping of human samples (86,87). First, samples are fractionated by chromatographic methods. Specific physiologic effects that are caused by the resultant fractions are then analyzed by an appropriate bioassay after each chromatographic step. The identity of the underlying substance is identified by MS methods such as TOF-TOF MS or FT-ICR MS. After identification of the compound, the (patho)physiologic effect(s) of the substance can be validated by analysis of the effects of the substance that was obtained by chemical synthesis. Finally, the concentration of the substance in samples of individuals is determined. Using this approach, not only the biomarker but also its (patho)physiologic relevance can be identified, if an appropriate in vitro model is available.

Urinary Proteomics for Nonrenal Diseases

Proteome analysis of the urine has revealed biomarkers for several nonrenal diseases. As in the case of ureteropelvic junction obstruction, these diseases generally do not result in significant proteinuria. Not surprising, biomarkers for urothelial cancer have been found in urine. Whereas the first studies that were based on SELDI technology analyzed only a few samples and reported different biomarkers for the same disease (88,89), Theodorescu et al. (8) recently used CE-MS to assay more than 600 samples, including 180 samples that were examined in a blinded manner, as a validation set. The discovered biomarkers correctly classified all blinded urothelial cancer samples and normal controls; however, nine of 138 patients with various chronic kidney diseases or nephrolithiasis were incorrectly classified as having urothelial cancer.

Kaiser et al. (48) defined biomarkers for graft-versus-host disease after bone marrow transplantation using CE-MS–based urine proteomics. This preliminary observation was validated in a recent prospective multicenter study with more than 600 urine samples from more than 100 patients (Weissinger et al., manuscript submitted). In recent studies, we were able to define several biomarkers that are indicative of cardiovascular disease (von zur Muhlen et al. and Zimmerli et al., manuscripts submitted). Although these observations at first sight seem intriguing, they may be explained by the microvascular architecture of the kidney. Graft-versus-host disease and cardiovascular disease cause or are the consequence of endothelial dysfunction that may also alter kidney structure and/or function. This complication may, in turn, influence glomerular filtration and/or tubular function and subsequently add disease-specific proteins to the urine.

Identification of Uremic Toxins Using Proteomics

Another application of proteomics that has gained considerable interest is the examination and definition of potential uremic toxins. Spent dialysate and hemofiltrate fluid is an excellent source for proteomic analysis because it contains little albumin and other interfering large proteins (90). In 1994, Forssmann and colleagues (90,91) used an advanced LC-MS approach to identify proteins from hemofiltrate fluid using a “peptide bank” with up to 300 different chromatographic fractions that were prepared from 10,000 L of human hemofiltration fluid. Several additional peptides with various biochemical functions were isolated (e.g., human peptide hormone guanylin, endostatin and resistin as angiogenesis inhibitors, a proopiomelanocortin-derived peptide with lipolytic activity) (92,93).

Ward and Brinkley (94) recently used a proteomic approach that was based on 2DE and MALDI-TOF-MS to identify uremic
toxins from an ultrafiltrate. Six proteins that harbored several posttranslational modifications (thus presenting as multiple spots for the same protein) were identified. The proteins included α2-microglobulin (95), as well as α1-antitrypsin, albumin (mature and complexed), complement factor D, cystatin C, and retinol-binding protein. Molina et al. (96) performed a proteome analysis of human hemodialysis fluid applying 1D gel electrophoresis in combination with LC-MS/MS. With this approach, 292 proteins were identified; 205 had not been previously found in serum or plasma. Additional Western blot analysis of a subset of these proteins confirmed their presence in normal serum. This observation indicates that a low sensitivity of detection may explain why most of these proteins had not been previously identified in serum or plasma. The authors concluded that this discrepancy may have resulted from enrichment of the low molecular weight proteins in the hemodialysis fluid.

CE-MS represents a supplement to these proteomic techniques, enabling the analysis of molecules in the low molecular mass range, from 1 up to 10 kD ("middle molecules"). In an initial approach, the effect of different dialysis membranes (low-flux versus high-flux) on the number of polypeptides in the dialysate was investigated (77). More than 600 polypeptides have been analyzed in a single sample. Larger polypeptides (>10 kD) were present mostly in dialysates that were obtained with high-flux membranes, whereas most of the polypeptides in dialysates that were obtained with low-flux membranes were smaller than 10 kD. Another study assessed the potential of CE-MS and CE-MS/MS to identify uremic retention molecules in dialysis fluids that were obtained with low-flux and high-flux membranes (97). For obtaining further insight into the uremic toxins within a mass range of 800 to 15,000 Da, the same CE-MS setup was used in combination with a different sample preparation procedure. CE-MS analysis detected 1394 polypeptides in the spent dialysate samples that were obtained with high-flux membranes, whereas 1046 polypeptides were recovered in the dialysate from the same patients that was obtained after hemodialysis with low-flux membranes. In an unrelated study, the same technology was used to identify polypeptides in the plasma of dialysis patients that are generally absent from normal control subjects (98). A combination of data from the study of human plasma and hemodialysate fluid should identify multiple previously unknown uremic toxins.

Bioinformatic Approaches in Proteomics

The results of most, if not all, proteomic studies indicated that a single biomarker does not allow reliable diagnosis, staging, or prognosis of a kidney disease. This finding immediately raised the question how to combine several biomarkers to provide a precise diagnostic or predictive pattern. Although a definitive answer is probably still on the horizon, a number of approaches emerged, which we discuss only in brief.

Hierarchical decision tree–based classification methods, such as classification and regression trees (99), were among the first algorithms to analyze the available information on multiple biomarkers. However, these approaches were not too successful because the number of incorrect predictions increased with the number of biomarkers (and, consequently, the complexity of the decision tree). Support Vector Machines (SVM; for an example, see Burges [100]) seemed to be an excellent way to overcome this problem. Indeed, reliable results have been obtained when the number of variables (biomarkers) was less than 20 and substantial differences between the data sets (biomarker panel) existed. However, when the differences were more subtle, the precision decreased considerably, particularly when blinded data sets were analyzed (H.M. et al., unpublished data), indicating once more the importance of the blinded samples in any clinical proteomics study.

An important caveat for the use of biomarker patterns for a predicted diagnosis with a classification algorithm is the level of confidence in the prediction. In other words, a classification such as “this urine sample is from an individual with type 2 diabetes” should have a numeric score indicating how likely it is to be a correct classification (e.g., “with 90% confidence, this urine sample is from an individual with type 2 diabetes”). Unfortunately, SVM are generally unable to provide levels of confidence to any classification. Therefore, the clinician is left with no information on the statistical significance of such a prediction. A promising classification method that shares many of the positive characteristics of the SVM but in addition provides the levels of confidence with each classification prediction is based on the Gaussian process (see Rasmussen and Williams [101]). An efficient Gaussian process–based classification method was recently developed (102) and successfully applied to the problem of correct prediction of BRCA1 and BRCA2 heterozygous genotypes (103). No matter which of these mathematical approaches is used, two basic considerations apply: (1) The number of independent variables should be kept to a minimum and should certainly be below the number of samples investigated, and (2) an approach is valid only when it is tested with a blinded validation set; it should be imperative to include such a blinded data set in any report on potential biomarkers.

Limitations of Proteome Analysis

Major limitations of proteomics are related to the type of biologic material to be analyzed and the sensitivity of the methods that are available for the analyses. For example, the role of some proteins in (patho)physiologic processes is not necessarily proportional to the concentration of that protein in the biologic compartment. Therefore, one of the main challenges is to identify scarce compounds and determine their changes between samples. Another limitation of proteome analysis—even more important than the analytical limits—is the stability of the proteome from the time of collecting/processing until completion of the preparation for analysis. This is especially evident when examining blood, as outlined previously.

The evident lack of standards and, subsequently, of comparability of results is another major limitation. The vast majority of the existing reports cannot be compared, thereby greatly reducing their relevance. The situation can be improved by using standard protocols for sample collection, storage, and preparation as well as by using standard analytical perfor-
mance (e.g., mass resolution and accuracy). The establishment of reliable 2DE-, LC-, and CE-MS databases using such standardized protocols would benefit this field. In addition, stricter rules for publication (e.g., mandatory blinded data sets) in peer-reviewed journals may improve the situation. Most of these issues have been outlined recently elsewhere (3), and adherence to the proposed guidelines will hopefully culminate in commonly accepted standards for clinical proteomics.

Lack of appropriate and user-friendly bioinformatics software for data evaluation also hinders development of clinical applications. So far, no standard has been developed for data evaluation, resulting in a set of different solutions that may work well only for a particular problem. However, because the different groups use highly divergent approaches, the data are generally not comparable. A data repository using a specific format, together with certain software solutions that would be universally available, may be an excellent first step toward establishing databases that can be directly compared.

Conclusion

Since the very first clinical observations of kidney diseases, it has been apparent that urinary proteins imply pathologic changes in the kidney. In the past, personal skills (simple observation, smelling, or even tasting of urine) were required and were skillfully performed by our predecessors. Presently, advanced technologies are available to improve the analytical description of the protein content of urine. During what we call the modern era, the contribution of proteomics to the understanding of the pathogenesis, diagnosis, and treatment of kidney disease has already been significant. However, its impact is modest in comparison with the expectations that have been generated by the more than 25 yr of technological progress.

Essentially all studies indicate that the “perfect biomarker” (i.e., a single molecule that clearly defines one disease) does not exist. A panel of distinct biomarkers may be better suited for disease detection (diagnosis) and also for assessment of disease progression and response to therapy. However, this panel of biomarkers must consist of individual biomarkers that are clearly defined and subsequently sequenced (in clear contrast to an ill-defined “pattern”). Furthermore, it is imperative to validate the clinical utility of such biomarkers using a blinded set of samples. Adherence to these simple guidelines should greatly improve the value of future proteome-based studies. The reports based on different technologies, albeit promising, clearly indicate an urgent need for standardization and show that a “common platform” that allows comparison of data sets from different laboratories is needed. Otherwise, these bits of information will never paint a “big picture” that is essential for the full expansion/realization of the potential of proteomics. Given the complexity of the task, it is essential that thousands of comparable data sets be available for data evaluation and validation. Because this cannot be accomplished separately by each laboratory, the establishment of comparability and standards for quality control (e.g., minimal requirements for mass accuracy and resolution of the choice of the mass spectrometers) is essential. First steps in that direction are the definition of guidelines for clinical proteomics (3) and the establishment of the Human Kidney and Urine Proteome Project” (HKUPP; http://hkupp.kir.jp).

Proteome analysis is still far from displaying its full potential as a routine tool in clinical examination, assessment of disease progression, etc. However, first studies with several hundred patients clearly reveal its utility for accurate noninvasive clinical diagnosis (8,47). It may take years or even decades until the entire urinary (or any other) proteome is explored. The question is whether this should be our primary goal or we should take full advantage of a subset of the proteome that contains highly valuable information for clinical use today.

Acknowledgments

J.N. was supported in part by grants DK61525 and DK71802 from the National Institutes of Health and by the General Clinical Research Center of the University of Alabama at Birmingham (M01 RR00032). A.A., V.J., J.J., and H.M. are members of the European Uraemic Toxin Group of the ESAO (EuTox).

We are grateful to Bruce A. Julian, Eric Schiffer, Jochen Ehrich, Tadashi Yamamoto, and Joost Schanstra for critically reviewing the manuscript.

Disclosures

None.

References

10. Omenn GS, States DJ, Adamski M, Blackwell TW, Menon


cortin-derived peptide from hemofiltrate of chronic renal failure patients. *Endocrinology* 146: 2060–2068, 2005


